

Ming Ronnier Luo
Editor

Encyclopedia of Color Science and Technology



SpringerReference

Encyclopedia of Color Science and Technology

Ming Ronnier Luo
Editor

Encyclopedia of Color Science and Technology

With 831 Figures and 52 Tables

 **Springer** Reference

Editor

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

School of Design
University of Leeds
Leeds, UK

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

ISBN 978-1-4419-8070-0 ISBN 978-1-4419-8071-7 (eBook)
ISBN 978-1-4419-8072-4 (print and electronic bundle)
DOI 10.1007/978-1-4419-8071-7

Library of Congress Control Number: 2016939599

© Springer Science+Business Media New York 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC New York

I like to dedicate this work to my family, my wife, Huimin, my son, Alan, and my daughter, Tina. They have given their unlimited supports over the years. Also, to my colleagues and students, for whom we have been able to continue to challenge the boundary of color science and technology.

Preface

Color is everywhere in daily life. It makes the world vibrant and exciting. However, we frequently make difficult color decisions and struggle to describe them precisely. Color on its own spans physics, chemistry, computing, psychology, and perception, and contributes fundamentally to art, design, and media. With this in mind, the primary goal of this publication was to provide a comprehensive and multidisciplinary reference to the main fields in which color is relevant. It can be used to promote the fields of color science and technology, to cover a wide range of disciplines, to provide good introductory material for beginners, and to attract a new generation of color researchers and engineers. I am very proud to say we have successfully achieved this goal. Readers will be able to survey information on color related subjects from many different aspects and influence the way people work with color. It has been a great honor and pleasure to coordinate this ambitious project. As an educator in the field of color technology, I have been fortunate to lead the project that includes 14 sections divided into 254 chapters. In total, 153 experts were invited to contribute to the work. This is the first reference work to refer to color from the different points of view of color vision, color appearance, optical phenomena, colorants, metrology, color spaces, and many other color related topics. It also includes domain specific entries such as color and visualization, color in graphics, image processing, and color management.

Readers can access each topic as a printed reference and also as part of Springer online reference material. In addition, each author will continue to update their material for years to come. Although a substantial amount of work has been accomplished, it is planned to extend the encyclopaedia to include new content as it becomes relevant and available.

Ming Ronnier Luo

Acknowledgments

I would like to thank all the contributors, including the section leaders, who each assembled a strong group of authors in their specific expert areas, and all the authors who prepared high quality input in a timely manner. I would like to particularly thank Prof. van Bommout who almost single handedly delivered the section on lighting technology based on his vast knowledge of the subject. I would also like to thank the late Prof. Janos Schanda for leading the section on CIE standards. It is with sadness that we learnt he had passed away during a busy period of preparation: he will be greatly missed by all color people.

About the Editor



Over the past 30 years, Ming Ronnier Luo has been involved in the field of color science and color technology. His major contributions to color science include advances in color difference evaluation, color appearance modeling, color emotion and harmony understanding and prediction, and lighting quality measurement. His mission has always been one of international dissemination of color expertise. He has created color and appearance postgraduate departments at universities in China, Taiwan, and UK, and currently he is a Global Expertise Professor at Zhejiang University (China), the Professor of Color and Imaging Science at the University of Leeds (UK), and a Chair Professor at the National Taiwan University of Science and Technology (R.O.C.). He and his colleagues have copublished over 500 refereed journal and conference papers. He has successfully supervised 45 PhD students and has taught color science courses to over 500 master students. These are now spread across all continents in academic institutions and industry. He has been heavily involved with the International Commission on Illumination (CIE), the world authority for the standardization and specification of color and lighting, presently holding the post of Vice President Publishing (2015–2019). He served the Director of CIE Division One on Vision and Color (2007–2015), and a chair or an active member of various technical and standardization committees. He was also Chairman of the Color Measurement Committee of the Society of Dyers and Colorists (1997–2002).

Ronnier has always worked closely with industry and strongly believes in the transfer of technology. Many of his new technologies have become successful industrial products, collaborating with international companies

covering the surface color, imaging, and illumination and supply chain industries.

In recognition of his contributions to color science and technology, he has gained a number of awards. These include the Dyers' Company Silver Medal (1986–1987), The Bartleson Research Award (1994), and the Davis Medal of the Royal Photographic Society (2003). He was awarded the title of Fellow of the Society of Dyers and Colorists in 2000, their Centenary Medal in 2004, and their Gold Medal in 2009. He became a Fellow of the Society of Image Science and Technology (2002) and the Dyers' Company Research Medallist in (2006).

Section Editors

1. Section: Language and Cognition



Kimberly A. Jameson Institute for Mathematical Behavioral Sciences,
University of California, Irvine, Irvine, CA, USA

2. Section: Standards: CIE



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

School of Design, University of Leeds, Leeds, UK

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



János Schanda Veszprém, Hungary

3. Section: Color Engineering: Computations



Philipp Urban Fraunhofer Institute for Computer Graphics Research IGD,
Technische Universität Darmstadt, Darmstadt, Germany

4. Section: Metrology: Concepts and Devices



Tzeng-Yow Lin National Measurement Laboratory R.O.C., Hsinchu,
Taiwan, R.O.C.

5. Section: Vision: Concepts



Sophie Wuerger Department of Psychological Sciences, University of Liverpool, Liverpool, UK

6. Section: Lighting



Wout van Bommel Nuenen, The Netherlands

7. Section: Materials: General



Stephen Burkinshaw School of Design, University of Leeds, Leeds, West Yorkshire, UK

8. Section: Computer Graphics



Erik Reinhard Technicolor Research and Innovation, Cesson-Sévigné, France

9. Section: Physics



Joanne Zwinkels National Research Council Canada, Ottawa, ON, Canada

10. Section: Vision: Concepts-2



Miyoshi Ayama Graduate School of Engineering/Center for Optical Research and Education, Utsunomiya University, Tochigi, Utsunomiya, Japan

11. Section: Art and Design



José Luis Caivano Secretaria de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

12. Section: People



Renzo Shamey Color Science and Imaging Laboratory, North Carolina State University, Raleigh, NC, USA

13. Section: Vision: Phenomena

14. Section: Vision: Physiology



Caterina Ripamonti Cambridge Research Systems Ltd, Rochester, Kent, UK

15. Section: Color Science: Computations



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

List of Contributors

Jonathan Aboshiha The UCL Institute of Ophthalmology and Moorfields Eye Hospital, London, UK

Nancy Alvarado Department of Psychology and Sociology, California State Polytechnic University, Pomona, Pomona, CA, USA

Seyed Hossein Amirshahi AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran

Malvina Arrarte-Grau Architecture, Landscape and Color Design, Arquitectura Paisajismo Color, Lima, Peru

Ulrich Bachmann Institute of Colour and Light, Zurich, Switzerland

Thomas Bechtold Research Institute for Textile Chemistry and Textile Physics, Leopold Franzens University of Innsbruck, Dornbirn, Vorarlberg, Austria

Tony Belpaeme School of Computing and Mathematics, University of Plymouth, Plymouth, UK

Nathan A. Benjamin Calit2, Computer Science, University of California, Irvine, Irvine, CA, USA

Martin Bide Department of Textiles, University of Rhode Island, Fashion Merchandising and Design, Kingston, RI, USA

David L. Bimler School of Psychology, Massey University, Palmerston North, New Zealand

Valerie Bonnardel Psychology, University of Winchester, Winchester, Hampshire, UK

Alex Bourque Department of Chemistry, Dalhousie University, Halifax, NS, Canada

David Briggs Public Programs, National Art School, Sydney, NSW, Australia

Michael H. Brill Datacolor, Lawrenceville, NJ, USA

Berit Brogaard Department of Philosophy and the Brogaard Lab for Multi-sensory Research, University of Miami and University of Oslo, Miami, FL, USA

Angela M. Brown College of Optometry, Department of Optometry, Ohio State University, Columbus, OH, USA

Robert R. Buckley National Archives of the UAE, Abu Dhabi, UAE

Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

John D. Bullough Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY, USA

José Luis Caivano Secretaria de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

Ellen C. Carter Color Research and Application and, Konica Minolta, Pennsville, NJ, USA

J. N. Chakraborty Department of Textile Technology, National Institute of Technology, Jalandhar, Punjab, India

Stephanie M. Chang Calit2, Computer Science, University of California, Irvine, Irvine, CA, USA

Mark Changizi Institute for the Study of Human and Machine Cognition, Boise, ID, USA

Cheng-Hsien Chen Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Vladimir G. Chigrinov Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong

Jae Hong Choi Department of Textile System Engineering, Kyungpook National University, Daegu, South Korea

Yi-Chen Chuang Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Filip Ciesielczyk Institute of Chemical Technology and Engineering, Poznan University of Technology, Poznan, Poland

Richard S. Cook Department of Linguistics, University of California, Berkeley, CA, USA

Oswaldo Da Pos Department of Psychology, University of Padua, Padua, Italy

Olga Daneyko Department of Neuroscience, University of Parma, Parma, Italy

Don Dedrick Department of Philosophy and Department of Psychology, University of Guelph, Guelph, ON, Canada

Peter Dehoff Zumtobel Lighting, Dornbirn, Austria

Pruthi S. Deshpande Cognitive Sciences, University of California, Irvine, CA, USA

Oskar Elek Computer Graphics Department, Max-Planck-Institut Informatik, Saarbrücken, Germany

Mark D. Fairchild College of Science, Rochester Institute of Technology, Rochester, NY, USA

Mariana Figueiro Lighting Research Center, Troy, NY, USA

Anna Franklin Psychology, University of Surrey, Surrey, UK

Harold Freeman Fiber and Polymer Science Program, North Carolina State University, Raleigh, NC, USA

Sergio Gago Calit2, School of Engineering, University of California, Irvine, Irvine, CA, USA

Karl Gegenfurtner Department of Psychology, Giessen University, Giessen, Germany

Alan Gilchrist Department of Psychology, Rutgers University, Newark, NJ, USA

Edward J. Giorgianni Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

Adam Glaz Institute of English Studies, Maria Curie-Skłodowska University, Lublin, Poland

Simone Gori Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy

Developmental Neuropsychology Unit, Scientific Institute “E. Medea”, Bosisio Parini, Lecco, Italy

Parikshit Goswami School of Design, University of Leeds, Leeds, West Yorkshire, UK

Paul Green-Armytage School of Design and Art, Curtin University, Perth, Australia

Thomas L. Griffiths Department of Psychology, University of California at Berkeley, Berkeley, CA, USA

J. Richard Hanley Department of Psychology, University of Essex, Colchester, UK

C. L. Hardin Department of Philosophy, Syracuse University, Syracuse, NY, USA

Amanda Hardman School of Psychology, University of Aberdeen, Aberdeen, UK

Ian G. Harris Computer Science, University of California, Irvine, CA, USA

Mathieu Hébert ERIS Group, CNRS, UMR 5516, Laboratoire Hubert Curien, Université de Lyon, Université Jean Monnet de Saint-Etienne, Saint-Etienne, France

Roger D. Hersch Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

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

Donald Hoffman Department of Cognitive Sciences, University of California at Irvine, Irvine, CA, USA

Stephanie Huette Department of Psychology, University of Memphis, Memphis, TN, USA

Shao-Tang Hung Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Robert W. G. Hunt Department of Colour Science, University of Leeds, Leeds, UK

Kimberly A. Jameson Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA

Teofil Jesionowski Institute of Chemical Technology and Engineering, Poznan University of Technology, Poznan, Poland

Yang Jiao Computer Science, University of California, Irvine, CA, USA

Gabriele Jordan Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK

Paul Kay Department of Linguistics, University of California, Berkeley, CA, USA

Algy Kazlaucius Department of Color Science, University of Leeds, Leeds, West Yorkshire, UK

Sung-Hoon Kim Department of Textile System Engineering, Kyungpook National University, Daegu, Republic of Korea

Daniel C. Kiper Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

Eric Kirchner Color Research, AkzoNobel Performance Coatings, Sassenheim, The Netherlands

Akiyoshi Kitaoka Department of Psychology, Ritsumeikan University, Kyoto, Japan

Martine Knoop Berlin University of Technology, Berlin, Germany

Balázs Kránicz University of Pannonia, Veszprém, Hungary

John H. Krantz Natural Sciences, Department of Psychology, Hanover College, Hanover, IN, USA

Rolf G. Kuehni Charlotte, NC, USA

Eun-Mi Lee Department of Textile System Engineering, Kyungpook National University, Daegu, Republic of Korea

David J. Lockwood Measurement Science and Standards, National Research Council Canada, Ottawa, ON, Canada

Jorge Lopez-Moreno GMRV Group, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain

Robert Lucas Faculty of Life Sciences, University of Manchester, Manchester, UK

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

School of Design, University of Leeds, Leeds, UK

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

Colin M. MacLeod Department of Psychology, University of Waterloo, Waterloo, ON, Canada

Kevin Mansfield UCL Institute for Environmental Design and Engineering, London, UK

John Mardaljevic School of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, UK

Maria Manuela Marques Raposo Department of Chemistry, University of Minho, Braga, Portugal

Jasna Martinovic School of Psychology, University of Aberdeen, Aberdeen, UK

Barry Maund Department of Philosophy, The University of Western Australia, Crawley, WA, Australia

John McCann McCann Imaging, Arlington, MA, USA

Terry McGowan Lighting Ideas, Inc, Cleveland Hts., OH, USA

Manuel Melgosa Optics Department, University of Granada, Granada, Spain

Ralf Michel Institute Integrative Design | Master Studio Design, University of Applied Sciences FHNW Basel, Basel, Switzerland

Galen Minah College of Built Environments, Department of Architecture, University of Washington, Seattle, WA, USA

Yoko Mizokami Graduate School of Advanced Integration Science, Chiba University, Chiba, Japan

John D. Mollon Department of Experimental Psychology, Cambridge University, Cambridge, UK

Guillermo Monros Departamento de Química Inorgánica y Orgánica, Universitat Jaume I, Edifici Científico-Tècnic, Castelló, Spain

Nathan Moroney Hewlett-Packard Laboratories, Palo Alto, CA, USA

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

Maria Luisa Musso Architecture, Design and Urbanism, Color Research Program, University of Buenos Aires, Buenos Aires, Argentina

Antal Nemcsics Department of Architecture, Budapest University of Technology and Economics, Budapest, Hungary

Larissa Noury Colour-Space-Culture, Paris, France

Leonhard Oberascher FH Joanneum University of Applied Sciences, Graz, Salzburg, Austria

Zena O'Connor Architecture, Design and Planning, University of Sydney, Sydney, Australia

Claudio Oleari Department of Physics, University of Parma, Parma, Italy

Nilgün Olguntürk Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Ankara, Turkey

Gertrud Olsson School of Architecture, KTH, Royal Institute of Technology, Stockholm, Sweden

Roy Osborne Color Research and Application, The Colour Group of Great Britain, London, UK

Gilles Page Philips Lighting, Miribel, France

Stephen E. Palmer Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

Galina V. Paramei Department of Psychology, Liverpool Hope University, Liverpool, UK

Neil Parry Vision Science Centre, Manchester Royal Eye Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Centre for Ophthalmology and Vision Sciences, Institute of Human Development, University of Manchester, UK

Baingio Pinna Department of Humanities and Social Sciences, Università degli Studi di Sassari, Sassari, Italy

Renata Pompas Milan, Italy

Bao-Jen Pong Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Tania Pouli Technicolor, Cesson-Sévigné, France

Peter Raynham University College of London, London, UK

Tobias Ritschel Department of Computer Graphics, MPI Informatik, Saarbrücken, Germany

- Herzog Robert** Max-Planck Institute for Informatics, Saarbrücken, Germany
- Marisa Rodriguez-Carmona** Optometry and Visual Science, City University London, London, UK
- Georges Roque** Centre National de la Recherche Scientifique (CNRS), Paris, France
- Miho Saito** Department of Human Sciences, Waseda University, Mikajima, Tokorozawa, Saitama, Japan
- Masato Sakurai** Kanazawa Institute of Technology, Nonoichi, Ishikawa, Japan
- Abhijit Sarkar** Surface, Microsoft Corporation, Redmond, WA, USA
- János Schanda** Veszprém, Hungary
- Karen B. Schloss** Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA
- Eriko Miyahara Self** California State University, Fullerton, CA, USA
- Renzo Shamey** Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA
- David R. Simmons** School of Psychology, University of Glasgow, Glasgow, Scotland, UK
- Alessandro Soranzo** Sheffield Hallam University, Sheffield, UK
- Andrew Stockman** Department of Visual Neuroscience, UCL Institute of Ophthalmology, London, UK
- Kartic Subr** Department of Computer Science, University College London, London, UK
- Suchitra Sueeprasan** Department of Imaging and Printing Technology, Chulalongkorn University Intellectual Repository, Chulalongkorn University, Pathumwan, Bangkok, Thailand
- Lois Swirnoff** The Cooper Union for the Advancement of Science and Art Cooper Square, New York, NY, USA
- Ken Tapping** D.R.A.O., National Research Council Canada, Penticton, BC, Canada
- Muneto Tatsumoto** Department of Neurology, Dokkyo Medical University, Tochigi, Japan
- Sean Tauber** Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA
- Li-Lin Tay** Measurement Science and Standards, National Research Council Canada, Ottawa, ON, Canada
- Krzysztof Templin** Computer Graphics, Max Planck Institute for Informatics, Saarbrücken, Germany

Peter Thorns Strategic Lighting Applications, Thorn Lighting Ltd, Spennymoor, Durham, UK

Richard J. D. Tilley Queen's Buildings, Cardiff University, Cardiff, UK

Andrew Towns Vivimed Labs Europe Ltd., Huddersfield, West Yorkshire, UK

H. Joel Trussell Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA

Philipp Urban Fraunhofer Institute for Computer Graphics Research IGD, Technische Universität Darmstadt, Darmstadt, Germany

Arne Valberg Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway

Wout van Bommel Nuenen, The Netherlands

Peter van der Burgt Philips Lighting, Valkenswaard, The Netherlands

Michal Vik Faculty of Textile Engineering, Laboratory Color and Appearance Measurement, Technical University of Liberec, Liberec, Czech Republic

Stephan Volker University of Technology Berlin, Berlin, Germany

Poorvi L. Vora Department of Computer Science, The George Washington University, Washington, DC, USA

Mark Wainwright Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, UK

David Alan Warburton Topoi, Berlin, Germany

Romann M. Weber Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA

Mike Webster Department of Psychology, University of Nevada, Reno, NV, USA

Bor-Jiunn Wen Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

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

Mary Anne White Department of Chemistry, Dalhousie University, Halifax, NS, Canada

Jonathan Winawer Department of Psychology, Stanford University, Stanford, CA, USA

Nathan Witthoft Department of Psychology, Stanford University, Stanford, CA, USA

Christoph Witzel Department of Psychology, Giessen University, Giessen, Germany

Wayne Wright Department of Philosophy, California State University, Long Beach, Long Beach, CA, USA

Sophie Wuerger Department of Psychological Sciences, University of Liverpool, Liverpool, UK

Kaida Xiao Department of Psychological Sciences, Institute of Psychology, Health and Society, University of Liverpool, Liverpool, UK

Bei Xiao Department of Computer Science, American University, Washington, DC, USA

Daniele Zavagno Department of Psychology, University of Milano-Bicocca, Milan, Italy

Sunčica Zdravković Department of Psychology, University of Novi Sad, NoviSad, Serbia

Laboratory for Experimental Psychology, University of Belgrade, Belgrade, Serbia

Jiyou Zhong National Engineering Research Center for Rare Earth Materials, General Research Institute for Nonferrous Metals, Beijing, China

Weidong Zhuang National Engineering Research Center for Rare Earth Materials, General Research Institute for Nonferrous Metals, Beijing, China

Joanne Zwinkels National Research Council Canada, Ottawa, ON, Canada

Ming Ronnier Luo
Editor

Encyclopedia of Color Science and Technology



SpringerReference

A

Achromatic Color

- [Anchoring Theory of Lightness](#)

Achromatopsia

- [Motion and Color Cognition](#)

Acquired Color Vision Impairment

- [Color Perception and Environmentally Based Impairments](#)

Acquired Color Vision Loss

- [Color Perception and Environmentally Based Impairments](#)

Adaptation

Claudio Oleari
Department of Physics, University of Parma,
Parma, Italy

Definition

Consider the ILV: International Lighting Vocabulary of the “Commission Internationale de l’Éclairage” [1] (CIE ILV) to identify what areas of vision science regard the adaptation phenomena [2–5].

CIE ILV defines

Adaptation Process by which the state of the visual system is modified by previous and present exposure to stimuli that may have various luminance values, spectral distributions, and angular subtenses.

NOTE Adaptation to specific spatial frequencies, orientations, sizes, etc. is recognized as being included in this definition.

Overview

Adaptation is a response modification of the visual system to light stimulation. The human visual system (HVS) changes its sensitivity as a function of the evolution over time of the observed scene and therefore has an evolution over time itself. Once the observed scene has stopped changing, the adaptation process continues until it becomes complete and then stops. The notation “adaptation” is used both for the process of adjustment as well as for the end state of complete adaptation. CIE ILV defines:

State of Adaptation	State of the visual system after an adaptation process has been completed.
------------------------	--

NOTE The terms *light adaptation* and *dark adaptation* are also used, the former when the luminances of the stimuli are of at least 10 cd/m^2 and the latter when the luminances are of less than some hundredths of a cd/m^2 .

The HVS, after being adapted to a bright light, may need a time in the order of more than one half hour to become completely dark adapted, while, after being adapted to darkness, it may need only a few minutes to become completely daylight adapted. These two processes are not symmetrical and are considered separately.

Dark adaptation is at the basis of the *duplicity theory* that states that two transduction mechanisms exist, which are related to two different kinds of photosensitive cells: the rods and the cones. In each retina, there are approximately between 75 and 150 million rods and six to seven million cone cells [6]. The rods are extremely sensitive to light; they contain rhodopsin as the light-absorbing pigment responsible for the transduction and provide achromatic vision. The cones are of three different classes (L cones, M cones, S cones) containing three different light-absorbing pigments responsible for the transduction. They are less sensitive than the rods and provide color vision. The distributions of cones and rods on the retina are very different and very nonuniform.

According to which type of photosensitive cell is responding to light, it is possible to distinguish between three different modes of vision. As CIE ILV defines:

Photopic Vision	Vision by the normal eye in which cones are the principle active photoreceptors.
--------------------	--

NOTE 1 Photopic vision normally occurs when the eye is adapted to levels of luminance of at least 10 cd/m^2 .

NOTE 2 *Color perception* is typical of photopic vision.

(Cone activity exists in the luminance range from 0.01 to 108 cd/m^2 .)

Scotopic Vision	Vision by the normal eye in which rods are the principle active photoreceptors.
--------------------	---

NOTE 1 Scotopic vision normally occurs when the eye is adapted to levels of luminance of less than $\sim 10^{-3} \text{ cd/m}^2$.

NOTE 2 In comparison to photopic vision, scotopic vision is characterized by the lack of color perception and by a shift of the visual sensitivity toward shorter wavelengths.

(Rod activity exists in the luminance range from 10^{-6} to 10 cd/m^2 .)

Mesopic Vision	Vision by the normal eye intermediate between photopic and scotopic vision.
-------------------	---

NOTE In mesopic vision, both the cones and the rods are active.

(The mesopic vision is between two luminance levels from 0.01 to 10 cd/m^2 .)

The state of adaptations enters the CIE ILV definition of:

Perceived Color	Characteristic of visual perception that can be described by attributes of hue, brightness (or lightness), and colorfulness (or saturation or chroma).
--------------------	--

NOTE 2 Perceived color depends on the spectral distribution of the color stimulus, on the size, shape, structure and surround of the stimulus area, on the *state of adaptation* of the observer’s visual system, and on the observer’s experience of the prevailing and similar situations of observation.

NOTE 4 Perceived color may appear in several modes of color appearance. The names for various modes of appearance are intended to distinguish among qualitative and geometric differences of color perceptions. Some of the more important terms of the modes of color appearance are given in “object color,” “surface color,” and “aperture color.” Other modes of color appearance include film color, volume color, illuminant color, body color, and ganzfeld color. Each of these modes of color appearance may be further qualified by adjectives to describe combinations of color or their spatial and temporal relationships. Other terms that relate to qualitative differences among colors perceived in various modes of color appearance are given in “luminous color,” “nonluminous color,” “related color,” and “unrelated color.”

(Different adaptations correspond to different color-appearance modes.)

The adaptation phenomenon depends on the adapting light, both on the luminous level and on the chromatic quality. CIE ILV defines the following other phenomena:

Chromatic Adaptation	Visual process whereby approximate compensation is made for changes in the colors of stimuli, especially in the case of changes in illuminants.
Adaptive Color Shift	Change in the perceived color of an object caused solely by change of <i>chromatic adaptation</i> .
Illuminant Color Shift	Change in the perceived color of an object caused solely by change of illuminant in the absence of any change in the observer’s state of <i>chromatic adaptation</i> .

Resultant Color Shift	Combined <i>illuminant color shift</i> and <i>adaptive color shift</i> .
-----------------------	--

CIE ILV defines the following quantities, useful to describe the adaptation phenomena:

Contrast Sensitivity [<i>Sc</i>]	Reciprocal of the least perceptible (physical) contrast, usually expressed as $L/\Delta L$, where L is the average luminance and ΔL is the luminance difference threshold.
------------------------------------	---

Unit: 1

NOTE The value of Sc depends on a number of factors including the luminance, the viewing conditions, and the *state of adaptation*.

Luminance Difference Threshold [ΔL]	Smallest perceptible difference in luminance of two adjacent fields.
---	--

Unit: $\text{cd/m}^2 = \text{lm}/(\text{m}^2 \cdot \text{sr})$

NOTE The value depends on the methodology, luminance, and on the viewing conditions, including the *state of adaptation*.

[Often, instead of luminance, the retinal illuminance I measured in troland (td) is considered, i.e., the luminance of the observed scene (cd/m^2) times the observer pupil area (mm^2).]

Luminance Threshold	Lowest luminance of a stimulus which enables it to be perceived.
---------------------	--

NOTE The value depends on field size, surround, *state of adaptation*, methodology, and other viewing conditions.

Moreover, there exists another phenomenon involving different photopic adaptation levels, known as *Bezold-Brücke phenomenon*, and regarding the shifts of hues produced by changing the luminance of a color stimulus while keeping its chromaticity constant.

The mechanisms behind the different adaptation processes are still not completely understood; therefore, here the main phenomenological

aspects of the adaptation are described. First, general aspects of the adaptation are considered and then the adaptation phenomena in the following order: the time-course of adaptation, subdivided into dark adaptation and light adaptation, and then brightness adaptation, chromatic adaptation, and color adaptation. Adaptation to specific spatial frequencies, orientations, sizes, etc. are not considered although included in the definition of adaptation given by CIE ILV.

The effect of the stimulation on the photoreceptors is a modification of its own effectiveness. Whenever there is a change in retinal illuminance and/or spectral power distribution of the stimulus, the visual mechanism starts readapting to the changing stimulus. Thus, in the real situation, when the light flux crossing the retina is continuously changing, the visual sensitivity at any particular time and place on the retina is the resultant of the effects of the stimulations due to the light flux in the time, previous and actual, and on the considered place and contiguous places. The term *local adaptation* refers to the effect of a stimulus which has been confined to a specific region of the retina.

The time change in sensitivity depends on the duration and also on the degree of stimulation. If the adapting light level changes by a relatively small amount, the visual system compensates for the change almost immediately, but, if the light level changes by a large amount, it takes a long time to reach complete adaptation. If a new stimulus remains the same for long enough time, the adaptation level reaches an equilibrium and the adaptation is *complete*. The time required for a complete adaptation depends on the starting level and on the new light level. The sensitivity change is very sudden in the initial phase, but its duration is a small fraction of the time required for a complete adaptation.

The term *adaptation level* defines the kind and degree of steady stimulation that would produce the same state of sensitivity as exists at any moment and place on the retina.

The adaptation level determines the range of responsiveness. The fully dark-adapted visual system cannot discriminate any luminances below an *absolute threshold* nor can the fully

light-adapted visual system discriminate any luminances above an upper *terminal threshold*. The overall range within which the optimally adapted visual system is effective is approximately from 10^{-5} to 10^5 cd/m². The retina is bound to be adapted to some level at any time and only a small part of the full range is available at any time. The ratio of maximum luminance over minimum luminance detectable for the full range is 10 billion to 1, while the momentary range for ordinary levels of luminance is in the order 1,000 to 1.

The adaptation mechanisms produce changes in threshold visibility, color appearance, visual acuity, and sensitivity over time.

Brightness Adaptation

The adaptation process operates over a luminance range of nearly 14 log units: the light of the midday sun ($\sim 10^8$ cd/m²) can be as much as ten million times more intense than moonlight ($\sim 10^{-6}$ cd/m²). The process which allows this great extension of retina sensitivity is called *brightness adaptation*.

Four mechanisms underlie the adaptation in such a wide range of luminances:

1. Pupil size
2. Switchover from rods to cones in the passage from scotopic to photopic vision
3. Bleaching/regeneration of the photopigments in the photosensitive cells (rods and cones)
4. Feedback from the horizontal cells to control the responsiveness of the photosensitive cells

These four mechanisms have the combined effect of making the retina more sensitive at low light levels and less sensitive at high light levels, with important consequences for perception.

The pupil size has only a small part in the adaptation process: the luminous flux entering the eye is proportional to the pupil area; therefore, since the pupil diameter ranges from 1 or 2 mm to approximately 8 mm, the luminous flux is modulated by a factor of 16–64. A change of 10 log units of luminance induces the pupil to change in

diameter from approximately 7–8 mm down to approximately 1–2 mm [7]. This range of variation produces a little more than one log unit change in retinal illuminance, so pupillary action alone is not sufficient to completely account for visual adaptation [8].

The neuron net of the retina has a very limited response range: -80 mV to $+50$ mV of graded potential in the non-spiking cells of the retina (rods, cones, and horizontal cells) and 0 to approximately 200 spikes per second for the ganglion cells.

The main parts of the adaptation process are due to the mechanisms (3) and (4) producing changes in retina sensitivity.

The effect of light on photopigments is their bleaching or depletion (photochemical effect) and, after bleaching, there is a regeneration of photopigments (chemical effect). Pigment bleaching makes the receptors less sensitive to light and, at high flux of light, produces a compression in their response. However, pigment bleaching cannot completely account for adaptation because the time-courses of the early phases of dark and light adaptation are too rapid to be explained by pigment bleaching alone [9].

Adaptive processes sited in the neural network of the retina (horizontal cells) have a multiplicative process effectively scaling the input by a constant related to the background luminance. This process acts very rapidly and accounts for changes in sensitivity over the first few seconds of adaptation. A slower acting subtractive process reduces the base level of activity in the system caused by a constant background. This last process accounts for the slow improvement in sensitivity measured over minutes of adaptation [10]. Sensitivity increases in dim light and decreases in bright light inducing a more or less constant range of response of the visual system.

For a given set of visual conditions, the current sensitivity level of the visual system is called the *brightness adaptation level*. These combined phenomena have the effect that the perceived brightness is approximately constant in a wide range of brightness adaptation levels.

The response of photosensitive cells may no longer increase if the light flux is so intense that

the regeneration of the photopigments is not able to counterbalance the bleaching completely. This situation is known as *saturation*.

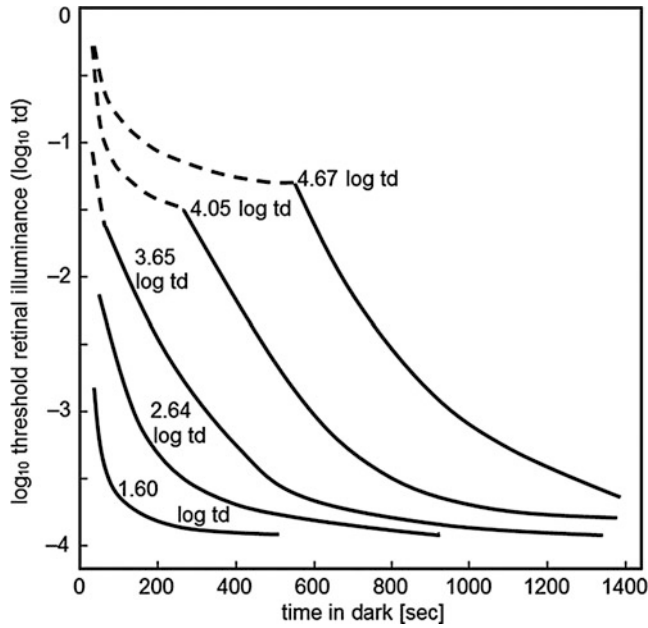
The term brightness adaptation is not defined in the CIE ILV; it is used seldom, and the phenomena here described are generally considered in the light adaptation process.

The Time-course of Adaptation

Dark Adaptation

Visually, dark adaptation is experienced as the temporary blindness that occurs when we go rapidly from photopic to scotopic levels of illumination. The ability to sense small illumination changes develops slowly in the darkness, and the slow time-course of dark adaptation means that HVS is impaired for some minutes when the observer moves quickly from a high level of illumination to a low level. The increment threshold detection of a light during dark adaptation well represents the dark-adaptation process. Test observers are first adapted thoroughly to a uniform background intensity. After the background light is removed, the observer's threshold is periodically measured in darkness. Consider the time-course of dark adaptation given by Haig [11] (Fig. 1).

In this experiment, the observer is first adapted to a high background luminance with the rod system depleted, and then the light is switched off abruptly. In the darkness, the detection threshold is measured continuously over more than 30 min. The detection threshold is the smallest perceptible luminance of a light spot on a black background. Violet light of short wavelength is used. The graph of Fig. 1 shows the detection threshold in the darkness as a function of time. In the first 5 min, after the adapting field is switched off, the threshold drops rapidly, but then it levels off at a relatively high level because the cone system has reached its greatest sensitivity, and the rod system is still not significantly regenerated. After about 7 min, rod system sensitivity comes over that of the cone system and the threshold begins to drop again. A change in slope separates the two curve branches which represent



Adaptation, Fig. 1 Increment threshold detection of a light during dark adaptation as a function of the background adapting illumination. Observers were first adapted thoroughly to a uniform background intensity. Once adapted, the increment threshold is periodically measured in darkness. Five curves show results from five different

widely spaced initial background luminances. Both rod and cone increment thresholds decrease asymptotically over time. The cones (*dashed lines*) adapt to darkness more rapidly and their increment thresholds prevail until rod thresholds (*solid lines*) appear with a slower time descent [11]

the rod and cone systems, respectively. This separation point is known as the Purkyně break (Purkyně shift) (► [Purkyně, Jan Evangelista](#)) [6] and indicates the transition from detection by the cone system to detection by the rod system. Changes in the threshold can be measured out to approximately 35 min, at which point the visual system has reached its absolute levels of sensitivity, and the threshold has dropped nearly 4 log units. The relatively slow time-course of dark adaptation means that vision can be impaired for several minutes when the observer moves quickly from high illumination levels to low ones.

The course of dark adaptation is influenced by the intensity and duration of light preadaptation, and different but analogous curves are measured in correspondence to different preadaptations.

The dark adaptation curves are different if stimuli of different wavelengths are used (Fig. 2). The scotopic (rods) and photopic (cones) spectral sensitivity functions are almost equal over 650 nm (Fig. 3); therefore, the Purkyně

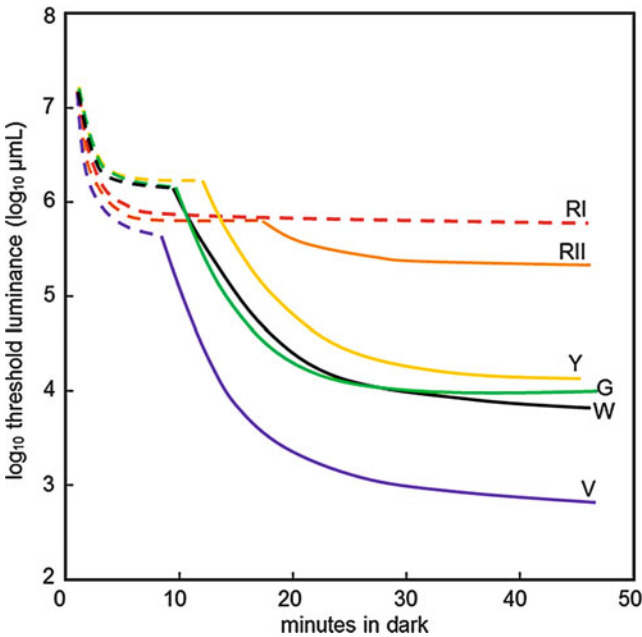
break is not seen. On the other hand, if light of short wavelength is used, the Purkyně break is most prominent because, once the rods have dark adapted, the rods are much more sensitive than the cones to short wavelengths.

Light Adaptation

Visually, light adaptation is experienced as the temporary blindness that occurs when the observer goes rapidly from a scotopic to a photopic level of illumination. The bright light momentarily dazzles the observer, to whom everything appears as a white light because the sensitivity of the receptors is set to dim light. Rods and cones are both stimulated, and large amounts of the photopigment are broken down instantaneously, producing a sensation of glare. During the light adaptation process, the HVS has to adapt quickly to the background illumination to be able to distinguish objects on this background. The sensitivity of the retina decreases dramatically. Retinal neurons undergo rapid adaptation

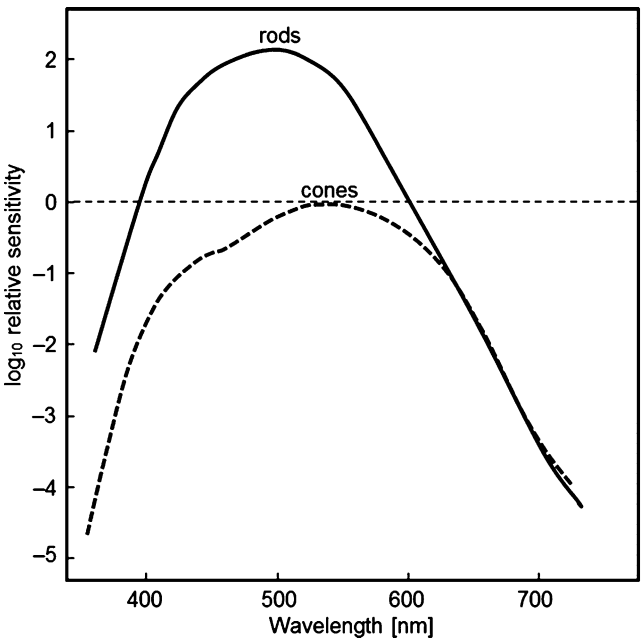
Adaptation,

Fig. 2 Increment threshold detection of a light during dark adaptation as a function of the background adapting illumination using different test stimuli of different wavelengths. Observers were preadapted to 2,000 mL for 5 min. A 3° test stimuli was presented 7° out of fovea toward the nose. The colors were RI (*extreme red*, wavelength of 680 nm), RII (*red*, wavelength of 635 nm), Y (*yellow*, wavelength of 573 nm), G (*green*, wavelength of 520 nm), V (*violet*, wavelength of 485 nm), and W (*white*) (Data from Chapanis [12])



Adaptation,

Fig. 3 Scotopic (rods) and photopic (cones) spectral sensitivity functions from Wald's data [13]



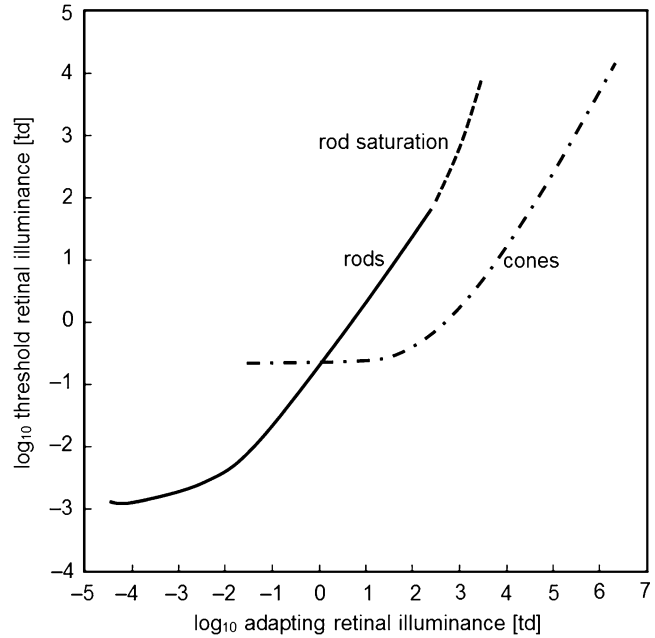
inhibiting rod function and favoring the cone system. Within approximately one minute, the cones are sufficiently excited by the bright light to take over. The process for light adaptation, in which visual accuracy and color vision continue to improve, occurs over 5–10 min. During light

adaptation to photopic vision, rod sensitivity is lost.

The HVS operates over a wide range of luminance levels. The sensitivity defines the state of brightness adaptation. Under low levels, the HVS has a very high sensitivity and can discriminate a

Adaptation,

Fig. 4 Threshold versus retinal illuminance curve, after Hood and Finkelstein [14]. As the background adapting retinal illuminance increases, the detection threshold becomes higher. Light of two different wavelengths is used in this case (580 nm for the test and 500 nm for the background) [13]



luminous spot as small as 100 photons (light quanta) against a black background. In the photopic range, the HVS requires a luminous spot of thousands or millions of photons to be seen against a background of higher illumination. HVS sensitivity is represented by the absolute intensity threshold, which is measured by a psychophysical experiment known as the *threshold versus illuminance* (TVI) experiment. The measured quantity is the retinal illuminance I . This experiment measures the minimum illuminance increment ΔI of a test spot, a sharp-edged circular target in the center of the visual field, required to produce a visual sensation on a uniformly lit background with retinal illuminance I_B , to which the observer is adapted. This can be achieved by placing an observer in front of a background wall of a given luminance (adapting luminance), and, once the adaptation is obtained, by increasing the retinal illuminance of the test spot ΔI from zero until it is just noticeable to the observer. The test phase has to be very quick avoiding any conditioning on the adaptation.

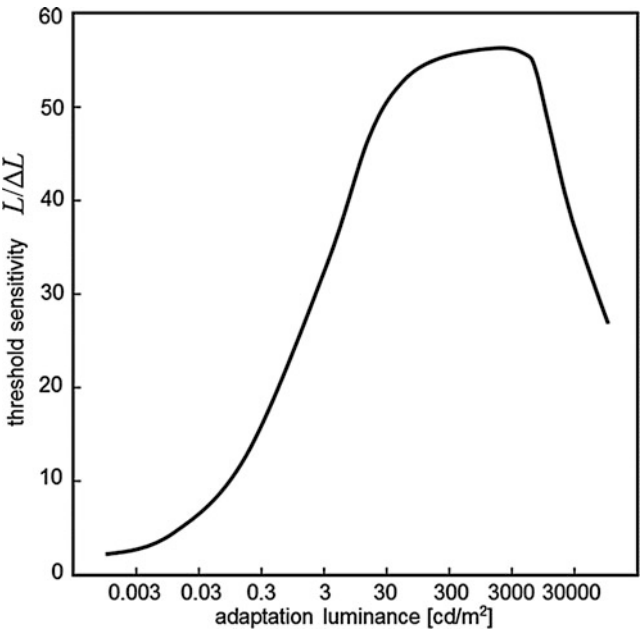
Figure 4 shows the TVI: the threshold of detectable retinal illuminance increment increases as the background adapting retinal illuminance increases (abscissa). The TVI has two branched curves, one

related to rod vision and the other to cone vision. Both curves have an analogous shape. Consider the rod curve, that is, represented by four sections:

1. The TVI below -4 log units is almost constant and the low background luminance does not significantly affect the threshold.
2. The TVI between -4 and -2 log units increases in proportion to the square root of the background retinal illuminance [15].
3. The TVI between -2 and $+2$ log units is in proportion to the background adapting retinal illuminance and the slope $\Delta I/I_B$ is constant. This section of the TVI is known as *Weber's Law*.
4. The TVI over $+2$ log units rises rapidly and the rod system starts to become incapable of detecting any stimulus. This is known as *saturation* and is represented in Fig. 4 with a dotted line [13].

Let us consider the cone curve of a TVI experiment as plotted in Fig. 5, where the plot regards the threshold sensitivity $L_B/\Delta L$ (i.e., the inverse Weber ratio) where L_B is the background luminance. The threshold sensitivity for discriminating small light increments on a background increases as the background luminance increases up to approximately 50 cd/m^2 . Between 50 and

Adaptation, Fig. 5 Plot of contrast sensitivity, i.e., the inverse Weber ratio, as a function of background adapting luminance L_B . The contrast sensitivity increases with the adaptation luminance up to approximately 50 cd/m². The Weber ratio is constant for a higher luminance, from approximately 100 cd/m² up to approximately 10,000 cd/m² [2]



10,000 cd/m², the threshold sensitivity is constant. In this range of luminance Weber’s law $\Delta L/L_B \approx 0.02$ holds true. This result refers to an experiment with a medium test field size (i.e., a size not disturbing the state of adaptation).

These results depend on the test size and retinal eccentricity because the distribution of the rods and cones on the retina are not uniform.

Chromatic Adaptation

Chromatic adaptation refers especially to those transient changes in sensitivity which are ascribable to photopic chromatic stimulation and are reflected in changes in chromatic sensation and perception. Visually, chromatic adaptation is experienced when a sudden illuminant change (e.g., in a room, the passage from daylight to a tungsten light) causes a global change in perceived color which is recognized by the observer as a change in illuminant. In spite of this evident change in the perceptual appearance of the scene caused by significant changes in the wavelength composition of the light reflected from different objects under the new illuminant, the perceived color of the objects remains largely unchanged. This adaptation phenomenon is termed

instantaneous color constancy. In literature, chromatic adaptation and color constancy are often considered as synonymous.

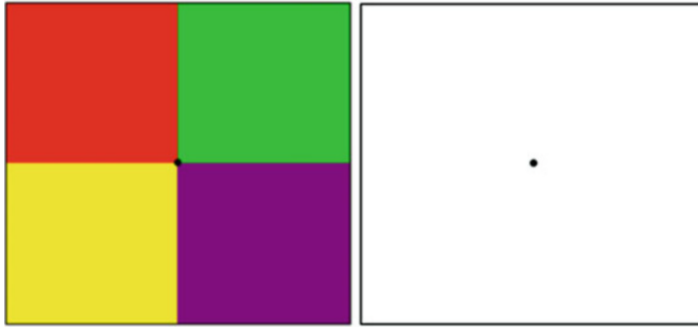
The chromatic, achromatic, and white sensations are a results of chromatic adaptation, and CIE ILV defines:

Adapted White	Color stimulus that an observer who is adapted to the viewing environment would judge to be perfectly <i>achromatic</i> and to have a luminance factor of unity.
---------------	--

NOTE The color stimulus that is considered to be the adapted white may be different at different locations within a scene.

Achromatic Stimulus	Stimulus that, under the prevailing conditions of adaptation, gives rise to an achromatic perceived color.
---------------------	--

NOTE In the colorimetry of object colors, the color stimulus produced by the perfect reflecting or transmitting diffuser is usually considered to be an achromatic stimulus for all illuminants, except for those whose light sources appear to be highly chromatic.



Adaptation, Fig. 6 Illuminate this figure with an intense light (~ 500 lx) and stare at it for approximately 15 s (better with one eye) the *black dot* in the center of the *left* colored image. Then quickly shift your attention onto the *black dot* in the center of the *right square*. The four colored *squares* of the *left* image immediately appear in complementary colors to those of the *left* image. This afterimage is due to a local color adaptation of the retina. Moreover the afterimage has smooth edges, revealing mutual interaction between the contiguous colored patches. To explain the

phenomenon, consider, for example, the region of the retina where the *yellow* field is first imaged. The prolonged exposure to the *yellow* light, that is a mixture of *red* and *green* light, reduces the sensitivities of the L and M cones. When the *yellow* field is replaced by *white* paper, the S cones with not reduced sensitivity generate a blue image. Then a random movement of the attention induces the HVS to recover the sensitivities of the cones and the afterimage fades

Chromatic Stimulus Stimulus that, under the prevailing conditions of adaptation, gives rise to a chromatic perceived color.

NOTE In the colorimetry of object colors, stimuli having values of purity greater than 0 are usually considered to be chromatic stimuli.

Color Adaptation

Visually, color adaptation is experienced when, in chromatic contexts after a prolonged fixation of a scene, a sudden change happens and the view of the new scene appears with overlapped characteristic colored afterimages, influenced by the adapting colors of the preceding scene (Fig. 6).

The term color adaptation is not defined in the CIE ILV; it is used seldom and generally the term *successive contrast* is used.

Cross-References

- [Afterimage](#)
- [Chromatic Contrast Sensitivity](#)

- [CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02](#)
- [CIECAM02](#)
- [Color Contrast](#)
- [Ganzfeld](#)
- [Purkyně, Jan Evangelista](#)

References

1. Standard CIE S 017/E:2011: ILV: International Lighting Vocabulary. Commission Internationale de l'Eclairage. Central Bureau, Vienna (2011)
2. Valberg, A.: Light Vision Color. Wiley, Chichester (2005)
3. Shevell, S.K. (ed.): The Science of Color, 2nd edn. OSA-Elsevier, Oxford, UK (2003)
4. Gegenfurtner, K.R., Lindsay, T.S. (eds.): Color Vision, from Genes to Perception. Cambridge University Press, Cambridge, UK (1999)
5. Dickinson, C., Murray, I., Carden, D. (eds.): John Dalton's Colour Vision Legacy. Taylor & Francis, London (1997)
6. Riggs, L.A.: Chapter 9: Vision. In: Kling, J.W., Riggs, L.A. (eds.) Woodworth and Schlosberg's Experimental Psychology, 3rd edn, pp. 273–314. Holt, Rinehart, and Winston, New York (1971)
7. Pugh, E.N.: Vision: Physics and retinal physiology. In: Atkinson, R.C. (ed.) Steven's Handbook of Experimental Psychology, 2nd edn. Wiley, New York (1988)

8. Spillman, L., Werner, J.S. (eds.): *Visual Perception: The Neurophysiological Foundations*. Academic, San Diego (1990)
9. Crawford, B.H.: Visual adaptation in relation to brief conditioning stimuli. *Proc. R. Soc. Lond. B.* **128**, 283–302 (1947)
10. Adelson, E.H.: Saturation and adaptation in the rod system. *Vision Res.* **22**, 1299–1312 (1982)
11. Haig, C.: The course of rod dark adaptation as influenced by the intensity and duration of pre-adaptation to light. *J. Physiol.* **24**, 735–751 (1941)
12. Bartle, J.N.R.: Chapter 8: Dark and light adaptation. In: Graham, C.H. (ed.) *Vision and Visual Perception*. Wiley, New York (1965)
13. Davson, H.: *Physiology of the Eye*, 5th edn. Macmillan Academic and Professional, London (1990)
14. Hood, D.C., Finkelstein, M.: Sensitivity to light. In: Boff, K.R., Kaufman, L., Thomas, J.P. (eds.) *Handbook of Perception and Human Performance. Sensory process and perception* (5), vol. 1, pp. 1–66. Wiley, New York (1986)
15. Barlow, H.B.: Increment thresholds at low intensities considered as signal noise discriminations. *J. Physiol.* **136**, 469–488 (1957)

Aftereffect

► Afterimage

Afterimage

Simone Gori

Department of Human and Social Sciences,
University of Bergamo, Bergamo, Italy
Developmental Neuropsychology Unit, Scientific
Institute “E. Medea”, Bosisio Parini, Lecco, Italy

Synonyms

Aftereffect; Ghost image; Image burn-in

Definition

An afterimage is an image that continues to be perceived after the physical stimulus that it originated from disappears from the observer’s visual

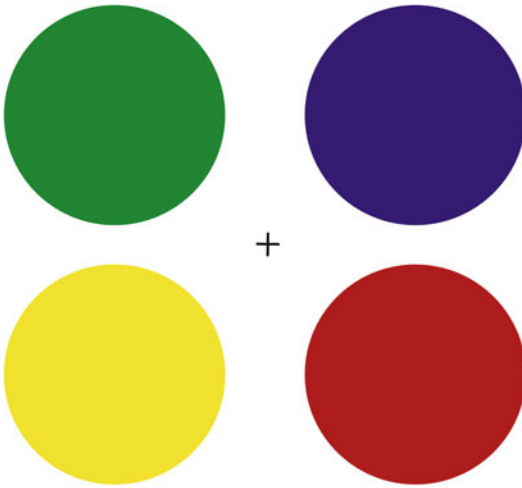
field. The afterimage can be a result of an exposition to a grayscale pattern, a colored pattern, or even to a motion stimulus. There are different kinds of afterimages, for example, the positive one will be the same with the physical stimulus, while the negative one will result in the opposite in terms of luminance, colors, or direction of the physical stimulus that generates it.

Introduction

In everyday experience it is not so uncommon to experience several types of aftereffects. The observation done by Addams [1] while watching the Falls of Foyers, near to the Loch Ness in Scotland, is probably the most famous description of the motion aftereffect outside a scientific laboratory. The motion aftereffect was already observed by Aristotle and Lucretius in the ancient times, but the description done by Addams remains the most cited example: if an observer watches a fall for about 60 s and, after that, she/he moves their gaze to the wood nearby, everything will seem to move upward. Now it is known that motion aftereffect can be so powerful that it can influence the direction of the illusory motion produced by static patterns [2] or even followed after only milliseconds of adaptation [3].

Possible Explanations

However, it is not just adaptation to motion that produces aftereffects. Although an afterimage can be easily experienced after looking at a prolonged grayscale image, the most popular afterimages are probably experienced after looking at colors. Starting at the fixation cross in Fig. 1 for more than 30 s and then moving the gaze to a white surface should produce a color negative afterimage: the disks will appear filled with the opposite colors of the original stimulus on the blank surface. The common opinion in the scientific community is that the afterimages result from neural adaptation, but adaptation is present throughout the visual system, which makes the neural substrate of these phenomena not so easy to isolate. For



Afterimage, Fig. 1 In order to experience a color negative afterimage, the reader should stare at the fixation cross shown for no less than 30 s and after try to move the gaze to a white surface. What should be seen on the blank surface are four circles filled with the opposite colors (the *green* should be now *red*, the *blue* *yellow*, the *yellow* *blue*, and finally the *red* *green*)

long, it was supposed that the afterimages were caused by fatigued cells in the retina responding to light. Then, if an observer will stare at the red color for enough time, the receptors in the retina responding to red will fatigue and will fire less. In color vision, it is generally believed that colors are represented along orthogonal opponent axes: yellow to blue and red to green. Consequently, when the observer will switch over to a white surface, after that the retina receptors for the red are adapted, the visual system will interpret that blank surface as filled with the red complementary color: green. This oversimplified explanation where all the color afterimages are explained at the retina level was successively challenged by several phenomena that required cortical elaboration [4]. In the 1960s it was shown that the rabbit visual cortex presented neural adaptation after prolonged exposition to a stimulus that was comparable to the duration of the aftereffects recorded during the psychophysical experiments [5]. This results, concordantly with other neuropsychological and psychophysical studies [4], moved the possible main physiological substrate of the afterimage in the primary visual cortex (V1). V1 is

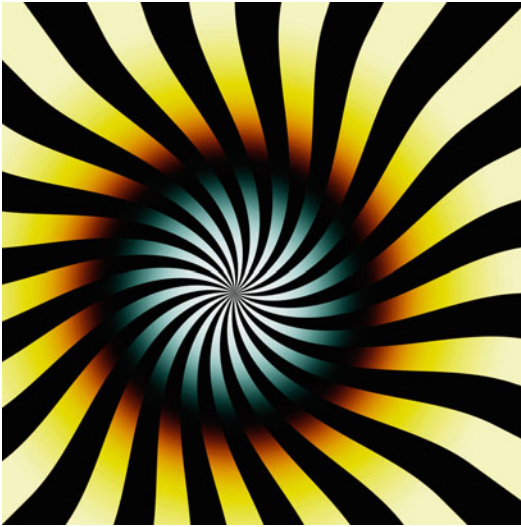
able to also explain the interocular transfer of the aftereffect, which is the possibility to adapt an eye and to observe some effects on the non adapted one.

However, even if low-level cortical processes seemed perfectly appropriated in order to explain the afterimage, the discovery that the top-down processes are able to influence radically both the primary visual cortex and also some subcortical “stations” of the visual system, as the lateral geniculate nucleus (LGN), together with some new types of aftereffects that showed the influence of attention [6, 7], consciousness [8], or an effect that lasts for days, weeks, or even months, seemed to clearly reject an explanation in terms of the short-lived adaptation reported typically in single cells of V1. These effects all suggested that the final percept of the afterimages is probably a complex result of the interaction and the adaptation of several districts throughout the visual system from the retina to the high-level cortical areas. Recently it was shown that the color afterimage signals are generated in the retina, but they are modified by cortical processes [9]. In sum the retinal afterimages face the similar destiny of all retinal signals that in humans are virtually always heavily modulated by the cortex.

An interesting question that can be posed, however, is that whether the afterimages have a function or if they are just a side effect incurred as a minor consequence of having evolved such a complex and superb machinery known as the human visual system. The answer, as it was probably expected, seems to be that the afterimages and the neural adaptation in general are part of complex neural strategies evolved during thousands of years in order to improve our world perception (e.g. [10, 11]).

Breathing Light Illusion

Finally the afterimage (with color or in grayscale, it depends on which version is used) is also at the basis of the family of illusions composed of the breathing light illusion and the dynamic luminance-gradient illusion [12–14]. Approaching these patterns like the one in Fig. 2 by moving



Afterimage, Fig. 2 A complex version of the breathing light illusion. Approaching the patterns by moving one's head toward it makes the spot appear to become larger, more diffuse, and filled with white. On receding from it, the spot's center remains white but the remainder appears smaller, darker, and sharper. In addition, in this specific pattern, an illusory rotation is perceived together with the typical illusory effect of the breathing light illusion. This illusory rotation is the same as experienced in the Rotating-Tilted-Lines Illusion [15–18]

one's head toward it makes the spot appear to become larger, more diffuse, and filled with white. On receding from it, the spot's center remains white but the remainder appears smaller, darker, and sharper. The proposed explanation of the phenomenon is related to the superimposition of the afterimage on the physical stimulus during dynamical viewing [13, 14].

In summary the afterimages are interesting phenomena that helped to understand our brain mechanism, they are able to produce striking illusory effects, and the underlying mechanisms seem to be relevant in producing a more efficient perception of the world.

Cross-References

- [Adaptation](#)
- [Assimilation](#)
- [Simultaneous Color Contrast](#)

References

1. Addams, R.: An account of a peculiar optical phenomenon seen after having looked at a moving body etc. *London Edinburgh Philosoph. Magaz. J. Sci.* 3rd series. **5**, 373–374 (1834)
2. Gori, S., Hamburger, K., Spillmann, L.: Reversal of apparent rotation in the Enigma-figure with and without motion adaptation and the effect of T-junctions. *Vision Res.* **46**, 3267–3273 (2006)
3. Pavan, A., Skujevskis, M.: The role of stationary and dynamic test patterns in rapid forms of motion after-effect. *J. Vis.* **13**, 10 (2013)
4. Thompson, P., Burr, D.: Visual aftereffects. *Curr. Biol.* **19**, R11–R14 (2009)
5. Barlow, H.B., Hill, R.M.: Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature* **200**, 1345–1347 (1963)
6. Reavis, E.A., Kohler, P.J., Caplovitz, G.P., Wheatley, T.P., Tse, P.U.: Effects of attention on visual experience during monocular rivalry. *Vision Res.* **83**, 76–81 (2013)
7. Van Lier, R., Vergeer, M., Anstis, S.: Filling-in after-image colors between the lines. *Curr. Biol.* **19**, R323–R324 (2009)
8. van Boxtel, J.J.A., Tsuchiya, N., Koch, C.: Opposing effects of attention and consciousness on afterimages. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8883–8888 (2010)
9. Zaidi, Q., Ennis, R., Cao, D., Lee, B.: Neural locus of color afterimages. *Curr. Biol.* **22**(3), 220–224 (2012)
10. Greenlee, M.W., Heitger, F.: The functional role of contrast adaptation. *Vision Res.* **28**, 791–797 (1988)
11. Barlow, H.B., Földiák, P.: Adaptation and decorrelation in the cortex. In: Darbin, R., Miall, C., Mitchison, G. (eds.) *The Computing Neuron*, pp. 54–72. Wesley Publishers, Reading (1989)
12. Gori, S., Stubbs, D.A.: A new set of illusions – the dynamic luminance gradient illusion and the breathing light illusion. *Perception* **35**, 1573–1577 (2006)
13. Anstis, S., Gori, S., Wehrhahn, C.: Afterimages and the breathing light illusion. *Perception* **36**, 791–794 (2007)
14. Gori, S., Giora, E., Agostini, T.: Measuring the breathing light illusion by means of induced simultaneous contrast. *Perception* **39**, 5–12 (2010)
15. Gori, S., Hamburger, K.: A new motion illusion: the rotating-tilted-lines illusion. *Perception* **35**, 853–857 (2006)
16. Gori, S., Yazdanbakhsh, A.: The riddle of the rotating tilted lines illusion. *Perception* **37**, 631–635 (2008)
17. Yazdanbakhsh, A., Gori, S.: A new psychophysical estimation of the receptive field size. *Neurosci. Lett.* **438**, 246–251 (2008)
18. Gori, S., Mascheretti, S., Giora, E., Ronconi, L., Ruffino, M., Quadrelli, E., Facoetti, A., Marino, C.: The DCDC2 intron 2 deletion impairs illusory motion perception unveiling the selective role of magnocellular-dorsal stream in reading (dis)ability. *Cerebral Cortex.* **25**(6), 1685–1695 (2015)

Al-Biruni

Eric Kirchner

Kirchner Publications, Leiden, The Netherlands

Abu'l-Raihan al-Biruni was a Persian polymath born in 973 AD. His contributions to mathematics, geography, astronomy, and physics are most well known. During his travels in India, he not only taught Greek science, but in exchange he became acquainted with Indian science. This made him a strong promoter of this science in the Islamic world.

At the court of sultan Mas'ud, al-Biruni wrote an important work on minerals and gemstones, called the *Kitab al-jamahir fi ma'rifat al jawahir* (The book of the multitude of knowledge of precious stones) [1–3]. In this lapidary, al-Biruni gives a detailed description of the colors of many minerals and gemstones. Color is used as a clear way to identify minerals and gemstones. He discusses extensively the slight color differences between minerals originating from different mines or having different degrees of purity, relating them to the effect this may have on their financial value.

The *Kitab al-jamahir* does not contain a separate chapter on color itself or on color ordering. But taken together, the many descriptions of color variations that appear in this book while discussing different types of minerals and gemstones do form a large body of scientific knowledge on color. Two centuries after al-Biruni, another Persian scientist by the name of al-Tusi would indeed collect many of al-Biruni's findings and formulate an impressive color ordering scheme by combining them with other sources [4].

An example of this part of al-Biruni's scientific legacy is found in his description of different types of rubies. According to al-Biruni, the colors of rubies range from sky blue, via lapis lazuli and indigo-blue, to kohl black. Almost exactly the same color series is found in the grand color scheme of al-Tusi, with only the color turquoise having been added. In a similar way, not only many color words in al-Tusi's grand color scheme already occur in al-Biruni's work, but the latter also provided several partial color orderings.

Another interesting aspect of al-Biruni's description of color is that he is one of the first

to verbally describe one of the color dimensions. After having described the color series from sky blue, via lapis lazuli and indigo-blue, to kohl black, al-Biruni characterizes this gradual color change by the word *shab'a* (p. 72 in Ref. [1]) (Vol. 2, p. 42 in Ref. [5]). Originally this word is used for describing the feeling one has after a copious meal, and it is best translated as saturation. In classical Arabic literature and in modern Arabic, the same word became the common word for color saturation (p. 701 in Ref. [6]). Although the color sequence mentioned above shows that the colors in this series do not change only with respect to color saturation in the modern definition of this word, al-Biruni's text is one of the first to relate this term to color [7].

References

1. Krenkov, F.: *Al-jamahir fi ma'rifat al-jawahir*. Deccan, Hyderabad (1936)
2. al-Hadi, Y.: *al-Jamahir fi al-jawahir*. Scientific and Cultural Publications, Tehran (1995)
3. Said, H.M.: *Al-Biruni's Book on Mineralogy – The Book Most Comprehensive in Knowledge on Precious Stones*. Pakistan Hijra Council, Islamabad (1989)
4. Kirchner, E.: Origin and spreading of Tusi's ideas on color ordering. Paper presented at the twelfth congress of the International Colour Association (AIC), Newcastle, 8–12 July 2013. Conference proceedings pp. 1197–1200.
5. Sabra, A.I.: *The Optics of Ibn al-Haytham*. Warburg Institute, London (1989). Two vols
6. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) *Encyclopaedia of Islam*, vol. V, pp. 698–707. Brill, Leiden (1991)
7. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Col. Res. Appl.* 40 (2015) 5–16.

Al-Farisi, Kamal al-Din Hasan ibn Ali ibn Hasan

Eric Kirchner¹ and Seyed Hossein Amirshahi²

¹Kirchner Publications, Leiden, The Netherlands

²AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran

At the end of the thirteenth century, Kamal al-Din Hasan ibn Ali ibn Hasan al-Farisi studied

at the school of Tabriz (Azerbaijan, Iran). His teacher was the famous astronomer Qutb al-Din al-Shirazi, who in turn was a student of Nasir al-Din al-Tusi.

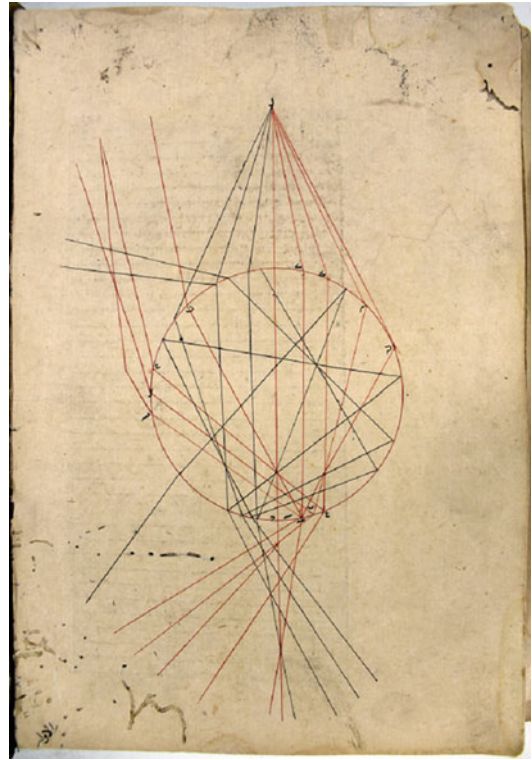
Al-Farisi felt “perplexity” when he found several inconsistencies and errors in the classical works about optics [1]. His teacher then managed to obtain a manuscript copy of the *Kitab al-Manazir* from Ibn al-Haytham, which was brought in “from a very distant land” (probably Egypt). Al-Farisi was greatly impressed by this work and decided to write a detailed commentary on it. He discussed it in great detail and completed it by adding appendices with other optical writings. Al-Farisi also included corrections to these texts, aptly calling the resulting work *Tanqih al-Manazir* (Revision of the [*Kitab*] *al-Manazir*) [2].

In one of al-Farisi’s comments in the *Tanqih*, he described how an object is seen in a certain color under sunlight, but in a different color under moonlight, and yet in another color in the light of fire. From this, al-Farisi concluded that colors are not really present in objects, but depend on illumination [3]. The role of incident light had therefore been changed from being a mere catalyst for color vision as in the ancient Greek theories to being as prominent as in modern color theory [4].

Together with his teacher Qutb al-Din al-Shirazi, al-Farisi tried to find explanations why the famous experiment of Ptolemy, with a spinning top (or mill stone) having sectors painted in different colors, led to new colors [5].

Regarding the colors of the rainbow, Ibn al-Haytham had supported the Aristotelian idea that these were due to a mixture of light and darkness. Al-Farisi rejected this concept. He argued that if it was true, then the colors of the rainbow would be ordered from bright to dark. Also, the secondary rainbow would then have the same color order. Both are not supported by observation [6].

But al-Farisi would become most famous for his experimental study of the formation of rainbow colors. Inspired by Ibn Sina’s work and by Ibn al-Haytham’s *Kitab al-Manazir* and the latter’s treatise on the burning glass sphere, he filled



Al-Farisi, Kamal al-Din Hasan ibn Ali ibn Hasan, Fig. 1 Double refraction and single reflection in a raindrop. Simultaneously with Theodoric of Freiberg, al-Farisi was the first to give the correct explanation of the rainbow in his *Tanqih al-Manazir* (©Library of the Masjid Sepahsalar, Tehran)

a glass sphere with water and considered this as a model for a droplet of rain water in the atmosphere. He then studied the resulting reflection and refraction of light in a dark room (see Fig. 1). This led him to the first correct explanation of the colors of the rainbow, which he described in the *Tanqih*. Interestingly in Germany, at approximately the same time also Theodoric of Freiburg, equally inspired by the *Kitab al-Manazir*, carried out the same experiment and formulated the same conclusions.

It was recently discovered that in his *Tanqih*, al-Farisi included the text of al-Tusi on color ordering, in which five paths were proposed to go from white to black [7]. In this way, this text would become available to many later generations. Al-Farisi also sought for an explanation of

the different orderings of the colors in the primary and secondary rainbow. His dark room experiments let him conclude that the various colors of the rainbow were produced by a superposition of different images as projected after reflections and refraction in the sphere [8]. Thereby, the colors became a function of the positions and luminous intensities of the composing images [9].

The *Tanqih al-Manazir* would become widely spread in the Muslim world, where it was used in academic classrooms, and would be commented upon until the sixteenth century. Therefore, in the Muslim world, this would be the main textbook on optics for more than three centuries. Since it would never be translated into any Western language, its influence on European science would remain marginal at best.

References

1. Kheirandesh, E.: Mathematical sciences through Persian sources: the Puzzles of Tusi's optical works. In: Pourjavady, N., Vesel, Z. (eds.) *Sciences, techniques et instruments dans le monde Iranien (Xe – XIXe siècle)*. Presses universitaires d'Iran and Institut Français de recherches en Iran, Téhéran (2004)
2. Kamal al-Din al-Farisi: *Tanqih al-Manazir*. Deccan, Hyderabad (1928). Two volumes
3. Gätje, H.: Zur Farbenlehre in der muslimischen Philosophie. *Der Islam* **43**, 280–301 (1967)
4. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Col. Res. Appl.* **40**, 5–16 (2015).
5. Vernet, J.: Mathematics, astronomy, optics. In: Schacht, J., Bosworth, C.E. (eds.) *The Legacy of Islam*, p. 483. Clarendon, Oxford (1974)
6. van Campen, M.: *De regenboog bij de Arabieren*. University of Utrecht, Utrecht (1988). MSc Thesis, 95
7. Kirchner, E.: Origin and spreading of Tusi's ideas on color ordering. Paper presented at the twelfth Congress of the International Colour Association (AIC), Newcastle, 8–12 July 2013. Conference proceedings, pp. 1197–1200
8. Rashed, R.: Kamal al Din. In: Gillespie, C.C. (ed.) *Dictionary of Scientific Bibliography*, pp. 212–219. Scribner's sons, New York (1970)
9. Rashed, R.: Geometrical optics. In: Rashed, R. (ed.) *Encyclopedia of the History of Arabic Science*, vol. 2, pp. 643–671. Routledge, London (1996)

Al-Haytham (Alhazen), Abū ‘Alī al-Ḥasan ibn al-Ḥasan ibn

Eric Kirchner

Kirchner Publications, Leiden, The Netherlands

Abu Ali al-Hasan Ibn al-Haytham (Latinized name: Alhazen) made important discoveries in astronomy, mathematics, and optics. His *Kitab al-Manazir* (Book on Optics) made the works of Aristotle, Ptolemy, and Euclid obsolete. He merged their classical theories, combining the mathematical, physical, and physiological aspects into one unified optical theory. The *Kitab al-Manazir* also introduced the distinction between optics and physiology/psychology that is a cornerstone of modern optics and colorimetry.

In his work, Ibn al-Haytham attempted to show how illumination, hue, and saturation combine together into color perception. Although he was not able to establish a complete system of color attributes, his account is largely consistent. “In terms of scope, details and comprehensiveness of treatment, Ibn al-Haytham’s observations on the subject of color perception are unequalled in any single writer before him” (Vol. 2, p. 43 in Ref. [1]).

The *Kitab al-Manazir* would be the dominant text on optics for several centuries to come, both in the East and West. Latin translations appeared from 1200 to 1572 (*Perspectiva*, *De aspectibus*, *Opticae Thesaurus*) and in Italian around 1350 (*Prospettiva*). Basing themselves on what they read in this book, Kamal al-Din al-Farisi and Theodoric of Freiburg around 1300 discovered the correct explanation for the colors of the rainbow. It also made Kepler to formulate in 1604 the theory of the retinal image and Rudolph Snel van Royen to discover a few years later what is now known as Snellius’ law of refraction. In the East, Ibn al-Haytham’s work survived through its thorough treatment and further elaboration by Kamal al-Din al-Farisi (d. 1318).

Ibn al-Haytham was born around 965 AD in Basra (Iraq) where he became a famous mathematician. Caliph al-Hakim of Cairo, impressed by Ibn al-Haytham's claim that he could regulate the flow of the Nile, persuaded him to come to Egypt. However, Ibn al-Haytham soon found out that he was not up to the task. He fell in disgrace and was placed under house arrest for many years. This proved to be beneficent for continuing his studies in mathematics, astronomy, and optics.

According to Aristotle, light was a manifestation of a change of the state of the medium from opaque to transparent. Ptolemy had considered vision to be the result of visual rays, spreading out from the eye that were reflected or refracted by surfaces of objects. Ibn al-Haytham (as well as Ibn Rushd) proposed an alternative view, that light has a much more active role in color vision. It is light which is seen, according to Ibn al-Haytham [2]. Moreover, light is directed toward the eye instead of spreading from it. Ibn al-Haytham was the first to consider light as an entity by itself, traveling from visible objects to the eye, according to mathematical laws originally proposed by Euclid [3]. Regarding the role of the medium for color vision, Ibn al-Haytham adopted the view from al-Kindi (d. 873) that the medium plays a passive role in color vision in the sense that it should not block vision.

In Ibn al-Haytham's theory, color is a distinct property of material bodies. He stated that color and light are distinct but that colors behave exactly as light does in transmission, reflectance, and refraction. Therefore, Ibn al-Haytham's theory of light is at the same time a theory of color [4]. But apart from formulating these theories, Ibn al-Haytham went further and designed many experiments by which he verified the proposal step by step and afterward subjected the results to thorough mathematical analysis [5] (and also Vol. 2, p. xli in Ref. [1]). Nevertheless, Ibn al-Haytham also explored some more philosophical treatments of color and light [5].

Ibn al-Haytham's view of the active role of light in color vision made him investigate many crucial

aspects of color vision. He described how changing the type and intensity of light affects the color appearance of objects and how a stronger light intensity increased perceived color differences [6, 7]. Thus Ibn al-Haytham came close to what has been called a key notion in colorimetry that color is the product of the eye, the light, and the object. In the words of Ibn al-Haytham: "This behavior indicates that the eye observes the colors of colored objects only according to the colors that fall on them" [6, 7].

Ibn al-Haytham was one of the first to use the word saturation as describing an aspect of color. Ibn al-Haytham used this word in his *Kitab al-Manazir*. Generally it refers to the sensation one feels after a copious meal, indeed analogous to the modern English word saturation. In modern Arabic, the same word is still used for color saturation [8]. About this and other words related to color that appear in the *Kitab al-Manazir*, see Vol. 2, p. 43 in Ref. [1].

In other aspects, Ibn al-Haytham followed the classical traditions. He supported the Aristotelian view that the colors of the rainbow are the result of mixing light and darkness and the Ptolemaic interpretation that color mixing in the spinning top is a visual illusion.

References

1. Sabra, A.I.: The Optics of Ibn al-Haytham, 2 vols. Warburg Institute, London (1989)
2. Smith, M.A.: Alhacen's theory of visual perception. *Trans. Am. Philos. Soc.* **91**, 4 (2001)
3. Darrigol, O.: The analogy between light and sound in the history of optics from the ancient Greeks to Isaac Newton. *Centaurus* **52**, 117–155 (2010)
4. Sabra, A.I.: Ibn al-Haytham's revolutionary project in optics: the achievement and the obstacle. In: Hogendijk, J.P., Sabra, A.I. (eds.) *The Enterprise of Science in Islam*, pp. 85–118. MIT Press, Cambridge, MA (2003)
5. el-Bizri, N.: Ibn al Haytham et le problème de la couleur. *Oriens Occidens* **7**, 201–226 (2009)
6. Wiedemann, E.: Zu Ibn al Haithams Optik. *Arch. Gesch. Naturwiss. Techn.* **3**, 1–53 (1910–1912)
7. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Color Res. Appl.* **40** (2015) 5–16.
8. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) *Encyclopedia of Islam*. Brill, Leiden (1991)

Al-Tusi, Nasir al-Din

Eric Kirchner¹ and Seyed Hossein Amirshahi²

¹Kirchner Publications, Leiden, The Netherlands

²AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran



Muhammad ibn Muhammad ibn Hasan al-Tusi, usually known as Nasir al-Din al-Tusi, was born in Tus (Iran) and later worked in Maragha (Azerbaijani) and Baghdad (Iraq). His influence reaches into many fields [1]. His work on reforming Ptolemaic theoretical astronomy would be crucial for later astronomers, including Copernicus. In mathematics he published landmark editions of the works of Euclid and Archimedes and developed trigonometry as a discipline separated from astronomy. He wrote several works on optics [2]. For Shiite theology, al-Tusi wrote an important work on ethics, and he also authored the first systematic treatment of

rationalist theology in 12 imam Shiism, a work still central in Shiite theological education.

Al-Tusi was born in a Twelver Shi'a family. Under patrons at Ismaili courts, he became a famous mathematician. When in 1256 under Mongol rule, Hulagu destroyed the Abbasid Empire, al-Tusi had taken refuge at the last Ismaili stronghold located on the mount of Alamut. After the Mongols destroyed this mountain fortress, Hulagu personally saved al-Tusi's life since the ruler's interest in astrology made him respect al-Tusi's astronomical knowledge. Al-Tusi persuaded Hulagu to support building the first full-scale astronomical observatory in the world at Maragha. With al-Tusi as director, it would collect a mass of observation data during about 50 years, and it would inspire later observatories in Samarkand and India and possibly even Tycho Brahe's observatory in Denmark.

Al-Tusi thoroughly studied the works of Ibn Sina. When after teaching Ibn Sina's color theory one of his students wanted to know more about it, al-Tusi replied in a letter:

Regarding the production of colors from black and white there are numerous paths, from which one gradually walks from white to black. The path through yellow belongs there: First by the mixing of dense and fire, both in small amount, the straw-yellow is produced, then the lemon-yellow, then the saffron-yellow, then the orange-yellow, then the grenade-yellow, then in it the tendency towards black increases, according to the increase in the number of dense particles and the decrease of fire, until it becomes black.

Another path goes through red. First it becomes rosy, then like evening-red, then blood-colored, then purple, then violet, violet-colored. [...]

This all occurs according to the differences of particles in transparency, opacity (density), light and darkness. Now and then one sees a color together with another, and a different color is produced, such as green from yellow and blue, verdigris from green and white. There are infinitely many of such arrangements, and some are often found in small particles of plants and animals. Anyone who observes them is surprised by their number. [3, 4]

Thus, while Ibn Sina had specified three paths from white to black, al-Tusi described five of such paths. They go via yellow, red, green, blue, and gray, as illustrated in the Figure above. Remarkably Forsius in 1611 also described five different

tint/shade scales from white to black. This may all be compared to the German monk Theophilus (ca. 1120), who described how to produce up to 12 grades in a tint/shade scale. The text of al-Tusi shows that he must have considered color space to be two dimensional. Interestingly, at approximately the same time in Latin Europe, Grosseteste had argued on theoretical grounds that color space is three dimensional [5, 6].

Also in his work on minerals and gem stones, the *Tansukhname-yi Ilkhani* (the book on precious stones for the Ilkhan [i.e., for Mongol ruler Hulagu]), al-Tusi describes a color theory [3, 7]. Written primarily by Nishaburi, and largely copied by al-Tusi and Kashani, these texts are the first to describe a limited hue scale. It describes that by mixing blue and yellow in different proportions, colors are produced that change gradually from blue, via green, to yellow. This description represents a great step forward from the Aristotelian point of view that stated that green is one of the colors painters cannot produce. Although several scholars before Nishaburi and al-Tusi had mentioned that green can be produced by mixing blue and yellow, no earlier scholar had described that depending on the mixing ratios, different hues of green are produced [6].

The common opinion among scholars since Aristotle had been that by mixing black and white all colors can be produced. But following Nishaburi, Tusi wrote that “if white color and black color are mixed with each other, an incense-grey color will result.” This had been stated only twice before in history and in much less clear wordings. Clearly this statement, which was a starting point for Newton’s optical work, was not made for the first time by Scaliger in 1557, as is generally thought [8].

References

1. Nasr, S.H.: al-Tusi. In: Gillespie, C.C. (ed.) Dictionary of Scientific Bibliography, vol. 13, pp. 508–514. Scribner’s sons, New York (1970)
2. Kheirandesh, E.: Mathematical sciences through Persian sources: the puzzles of Tusi’s optical works. In: Pourjavady, N., Vesel, Z. (eds.) Sciences, techniques et instruments dans le monde Iranien (Xe – XIXe siècle).

Presses universitaires d’Iran and Institut Français de recherches en Iran, Téhéran (2004)

3. Kirchner, E., Bagheri, M.: Color theory in medieval lapidaries: Nishaburi, Tusi and Kashani. *Centaurus* **55**, 1–19 (2013)
4. Persian edition in Nourani, A.: *Ajvebat al-Masā’el al-Nasīriyah*. Institute for Humanities and Cultural Studies, Tehran (2005). German translation in: Wiedemann, E.: Über die Entstehung der Farben nach Nasir al Din al Tusi. *Jahrbuch der Photographie* **22**, 86–91 (1908). Reprinted in Wiedemann, E.: Aufsätze zur arabischen Wissenschaftsgeschichte, pp. 256–261. Georg Olms, Hildesheim (1970)
5. Smithson, H., Dinkova-Bruun, G., Gasper, G.E.M., Huxtable, M., McLeish, T.C.B., Panti, C.: A three-dimensional color space from the 13th century. *J. Opt. Soc. Am. A* **29**, A346–A352 (2012)
6. Kirchner, E.: Color theory and color order in medieval Islam: a review. *Col. Res.* **40** (2015) 5–16.
7. Razavi, M.: *Tansukhname-ye Ilkhany*. Miras-i Maktub, Tehran (1969)
8. Shapiro, A.E.: Artists’ colors and Newton’s colors. *Isis* **85**, 600–630 (1994)

Anchoring Theory of Lightness

Alan Gilchrist

Department of Psychology, Rutgers University,
Newark, NJ, USA

Synonyms

Achromatic color; Perceived reflectance;
Perceived value

Definition

Lightness refers to the white, black, or gray shade of a surface. The basic problem is that the light reaching the eye from a given surface, as the product of surface reflectance and illumination, does not specify the reflectance of the surface. A black in sunlight can reflect more light than a white in shadow. No computer program exists that can identify the reflectance of an object in a photo or video. Any possible solution must exploit the context surrounding the object. Anchoring theory proposes to solve the problem by grouping

patches of the retinal image into regions of equal illumination, called frameworks, then computing lightness values within each framework using rules of anchoring combined with luminance ratios.

Background

Lightness refers to the perceived shade of white, gray, or black of a surface and is sometimes called perceived reflectance. The reflectance of a surface is the percentage of light it reflects. White surfaces reflect about 90 % of the light that illuminates them while black surfaces reflect only about 3 %. Given the light-sensitive receptor cells in the eye, one might think that the perception of lightness could be easily explained. But the amount of light reflected by a surface, called luminance, depends not only on its reflectance but also on the intensity of light illuminating it. And the intensity of illumination varies from place to place and time to time. Thus, a black surface in sunlight can easily reflect more light than a white in shadow. In fact, any shade of gray can reflect any amount of light.

The ability to perceive the correct shade of gray of a surface despite variations (both spatial and temporal) in illumination is called lightness constancy, or more specifically, illumination-independent lightness constancy. The ability to perceive the correct shade of gray despite variations in the luminance of the background surrounding a target surface is called background-independent constancy. Neither type of constancy is complete. Variations in illumination level produce failures of constancy as described in the traditional literature while variations in background produce failures of constancy that have been called contrast effects.

Summary of the Theory

According to anchoring theory [1], the lightness of a target surface is codetermined by its relationship to the field of illumination in which it is embedded, called the local framework, and its

relationship to the entire visual field, or global framework. The values computed within these local and global frameworks are combined in a weighted average. Within each framework the value is based on the ratio between the luminance of the target surface and the highest luminance in the framework, called the anchor. The weight of a framework depends on its size and complexity. Framework boundaries include corners, occlusion contours, and penumbrae.

Frameworks

Helmholtz [2] suggested that given the luminance of a target surface, the visual system would have to somehow take into account the intensity of light illuminating that surface in order to compute its reflectance. But this idea has never been adequately operationalized. Koffka [3] suggested that fields of illumination can be regarded as frames of reference. The same luminance value that would be computed to be a light gray if it appears within a shadow would be computed to be a dark gray if it appears in sunlight. Anchoring theory borrows Koffka's concept of frames of reference.

Anchoring

In 1948, Wallach [4] proposed the idea that lightness depends simply on the ratio between the luminance of a surface and the luminance of the surrounding region. There is clearly much validity in this proposal. A homogenous disk of constant luminance can be made to appear as any shade of gray from white to black simply by varying the luminance of a homogeneous ring that surrounds it. However, relative luminance per se can only produce relative shades of gray. For example, consider two adjacent surfaces whose luminance values stand in a 5:1 ratio. These could be white and middle gray. But they might also be light gray and dark gray, middle gray and black, or any number of reflectance pairs. To determine the specific shade of a surface requires, in addition to relative luminance, an anchoring rule – that is, a

rule that links some value of relative luminance to some value on the gray scale. Anchoring theory invokes a highest luminance rule according to which the highest luminance within a region of illumination is perceived as white, and this value serves as the standard to which lower luminance values are compared.

Simple and Complex Images

Anchoring theory makes an important distinction between simple images and complex images. A simple image is one in which a single level of illumination fills the entire visual field while a complex image, typical of most images we see, contains multiple levels of illumination. The theory also describes how the rules for simple and complex images are related to each other.

Simple Images

In a simple image, the formula used to predict lightness value of a given target surface is **Lightness** = $T/H * 0.9$, where **T** is the luminance of the target, **H** is the highest luminance in the image, and **0.9** is the reflectance of white.

Geometric relationships, in addition to photometric relationships, also play a role in anchoring. Lightness depends, to a limited extent, on the relative area of surfaces. In general, lightness increases as relative area increases. But according to anchoring theory, this effect of “the larger the lighter” applies only to surfaces darker than the highest luminance and then only when that darker surface fills more than half of the area of the visual field.

Complex Images

While it is not hard to exhaust the rules by which lightness is computed in simple images, the more important goal of a theory of lightness is to explain how lightness is computed in complex images typical of the real world. Anchoring theory makes the claim that complex images are related

to simple images according to two principles. The first is that the same rules of lightness computation found in simple images apply to frameworks embedded in complex images. The second involves a kind of crosstalk between frameworks called codetermination.

Codetermination

The term *codetermination* was coined by Kardos [5], a gestalt psychologist from Budapest. He proposed a friendly amendment to Koffka’s idea of frameworks, observing that the lightness of a target is not computed exclusively within its framework but is influenced by neighboring frameworks as well. Kardos argued that the lightness of a surface is partly determined by its relative luminance in the field of illumination to which it belongs, called the relevant field, and partly by its relationship to a neighboring field of illumination, which he called the foreign field.

Anchoring theory invokes the closely related concepts of local and global frameworks. The global framework is the entire visual field. Unlike relevant and foreign frameworks, which are exclusive, local and global frameworks are hierarchical. The global framework includes the local.

Segmentation

In general, the visual field is segregated into frameworks by identifying illumination boundaries. More specifically, these include cast illumination edges, such as the border of a shadow or spotlight, corners, where the planarity of a surface changes, and occlusion boundaries, where one object partially obscures a more distant object.

Weighting

To implement the concept of codetermination, a provisional lightness value is computed for the target within each framework and then a weighted average of these values is computed. The weight of a framework in this average depends on the size

of the framework and the degree of complexity within it.

The factors of framework size and articulation come from the early work of David Katz [6]. Having invented some of the fundamental methods for studying lightness, Katz conducted experiments testing the factors that lead to strong and weak constancy. He found that the degree of constancy varies in proportion to the size of a framework, which he called field size, and its level of articulation. In his experiments with Burzlaff, level of articulation was operationally defined simply as the number of distinct elements within it.

The concepts of field size and articulation are given a slightly different interpretation in anchoring theory. Rather than factors consistently leading to good constancy, they are treated as factors that determine the relative weight of a framework. This twist resolves an empirical challenge for Katz, namely, that under certain conditions, increasing the articulation level of a framework leads to weaker constancy.

The general formula for a target surface belonging to the global framework and one local framework is

$$\text{Lightness} = W(T/H_l * 0.9) + (1 - W_l)(T/H_g * 0.9)$$

in which **W** is the weight of the local framework, **T** is target luminance, **H_l** is the highest luminance in the local framework, **H_g** is the highest luminance in the global framework, and **0.9** is the reflectance of white.

Staircase Gelb Effect

Anchoring theory was inspired by an empirical finding called the staircase Gelb effect. Gelb had shown that a piece of black paper appears white when suspended in midair and illuminated by a spotlight. In the staircase version, a row of five squares arranged in an ascending series from black to white is placed within the spotlight. This produces a dramatic compression of the perceived range of grays. Although the white square is perceived correctly as white, the black square appears somewhat lighter than middle gray. This

finding seriously undermined several earlier decomposition theories of lightness but suggested an interaction between the bright illumination on the five squares and the dimmer illumination of the room.

According to anchoring theory, each of the squares is assigned its correct value within the local spotlight framework. But relative to the surrounding room, as long as the spotlight is at least 30 times brighter than the illumination level in the room, each square is assigned the same (global) value of white, given that its luminance is as high as, or higher than, a white surface in the room light. The perceived compression of the lightness range is explained by these equal values assigned to the squares relative to the room illumination.

The Scaling Problem

The mapping of luminance values onto lightness values within a given framework involves both anchoring and scaling. If anchoring concerns which value of relative luminance is linked to which value of lightness, scaling concerns which range of luminance values is linked to which range of lightness values. Much as a normal distribution can be characterized by both a measure of central tendency and a measure of dispersion, so a set of luminance values within a framework can be converted to lightness values using an anchoring rule and a scaling metric (although the anchor is located at the top, rather than in the center, of the range).

Wallach's ratio principle assumes a one-to-one scaling metric. That is, if the luminance ratio between two adjacent surfaces is 3:1, then the ratio of their two perceived reflectance values must also be 3:1. This scaling metric is implicit in the formula given above: **Lightness = T/H * 0.9**. It can be considered the default scaling metric and applies to large, well-articulated frameworks typical of everyday scenes.

However, under specific conditions, the range of perceived lightness values can be either expanded or compressed, relative to the range of luminance values. This occurs when the luminance range within a framework is either much

less or much greater than the canonical 30:1 range between white and black (from 90 % to 3 %). Anchoring theory thus includes a scale normalization principle, according to which the range of perceived lightness values tends toward the white–black range. When the luminance range within a framework is less than 30:1 the lightness range is expanded, and when the luminance range exceeds 30:1 the lightness range is compressed. The degree of expansion or compression is proportional to the deviation of the range from 30:1, but the normalization is only partial.

Veridicality Versus Error

In general, theories of lightness have sought to account for lightness constancy, and the degree to which perceived lightness corresponds to physical reflectance. However, an adequate theory of lightness must also explain failures of constancy. The Kardos theory of codetermination was the first lightness theory to confront this problem. The failure of constancy within a relevant field is due to the unwanted influence of the foreign field. Anchoring theory endorses this construction but also broadens it to include a larger class of what can be called lightness errors. This class includes lightness illusions in addition to failures of constancy. Illusions and constancy failures have traditionally been treated separately but have nonveridicality in common.

The most basic and familiar illusion of lightness is called simultaneous contrast. Two identical gray patches are placed, respectively, on adjacent black and white backgrounds. The patch on the black background appears lighter than the patch on the white background. The application of anchoring theory to this illusion is rather simple. The illusion is obviously composed of two frameworks, corresponding to the two backgrounds. These two frameworks taken together can be considered the global framework. For each gray patch, both a local and a global value are computed. In the global framework, both gray patches are computed to be middle gray. But they get different values when computed locally. The patch on the black background is given a value

of white in its local framework, because it is the highest luminance. However, the patch on the white background is given a value close to middle gray (but somewhat darker due to scale normalization). When the local and global values for each patch are combined in a weighted average, the patch on the black background ends up with a slightly higher value. According to anchoring theory, this “contrast effect” is relatively weak because the two local frameworks are weak, due to their lack of articulation.

According to anchoring theory, the simultaneous contrast illusion is not caused by lateral inhibition or any exaggeration of edge differences but rather by the fact that one patch is perceptually grouped with the black background while the other patch is grouped with the white background. Consistent with this grouping approach, several authors have produced reverse contrast illusions. These illusions are variations of simultaneous contrast in which a gray patch on the black background appears darker, not lighter, than an identical gray patch on the white background. This reversed effect is created by making the gray patch a member of a group of patches that are opposite in lightness from the black background. Thus, the gray patch on the black background is part of a matrix of white patches of the same size and shape. The gray patch on the white background is likewise part of a group of black patches. These reverse contrast illusions support the concept of grouping that is fundamental to gestalt theories in general and anchoring theory in particular.

Grouping by Illumination

Although Helmholtz suggested that the visual system must take into account the level of illumination on a given target surface, in fact this is overkill. To compute lightness, the visual system needs to know only which surfaces are under *the same level of illumination*. Thus the segregation of the retinal image into frameworks is equivalent to grouping together those retinal patches that lie under homogeneous illumination. This kind of perceptual grouping must be distinguished from

the more traditional gestalt idea of grouping by which objects are segregated. In fact, grouping by common illumination is orthogonal to the traditional kind, which can be thought of as grouping by common reflectance.

Anchoring in Other Domains

The anchoring problem arises in other perceptual domains including perceived size and perceived motion. Motion is a good example. There is a good deal of evidence that perceived motion depends on relative motion, but to achieve specific values of motion, relative motion must be anchored. Here, the metaphor of an anchor is even more relevant than in lightness. In a matrix of relative motions, which elements will be perceived as stationary? In his brilliant writing on perceived motion, Duncker [7] stressed the importance of frames of reference. And he talked specifically about anchoring rules in a way that appears highly consistent with the anchoring theory of lightness.

Strengths and Weaknesses of the Theory

Anchoring theory has wide applicability within lightness perception. It is consistent with an extensive range of empirical data. In general, it accounts for veridical perception to the degree that perceived lightness is veridical. But unlike most other theories, it also accounts for a wide range of known lightness errors, that is, lightness illusions and failures of constancy. This is important because the pattern of errors shown by humans in lightness perception must be the signature of our visual software.

But anchoring theory is a work in progress. It is not clear whether the current rubric of local and global or the Kardos rubric of relevant and foreign is best. Local/global works effectively for lightness illusions like simultaneous contrast and reverse contrast but creates some difficulties for failures of constancy. Relevant/foreign works better for failures of constancy but creates difficulties for contrast illusions. This and other problems with the theory are discussed in Gilchrist ([1], pp. 354–357).

Cross-References

- ▶ [CIE Whiteness](#)
- ▶ [Color Constancy](#)
- ▶ [Ganzfeld](#)
- ▶ [Helmholtz, Hermann Ludwig von](#)
- ▶ [Katz, David](#)
- ▶ [Perceptual Grouping and Color](#)
- ▶ [Simultaneous Color Contrast](#)

References

1. Gilchrist, A.: *Seeing Black and White*. Oxford University Press, New York (2006)
2. von Helmholtz, H.: *Helmholtz's Treatise on Physiological Optics*. Optical Society of America, New York (1866/1924)
3. Koffka, K.: *Principles of Gestalt Psychology*. Harcourt, Brace, and World, New York (1935)
4. Wallach, H.: Brightness constancy and the nature of achromatic colors. *J. Exp. Psychol.* **38**, 310–324 (1948)
5. Kardos, L.: Ding und Schatten [Object and Shadow]. *Zeitschrift für Psychologie*. **23** (1934), 1–184
6. Katz, D.: *The World of Colour*. Kegan Paul, Trench, Trubner, London (1935)
7. Duncker, K.: Induced motion. In: Ellis, W.D. (ed.) *A Source Book of Gestalt Psychology*, pp. 161–172. Humanities Press, New York (1950)

Ancient Color Categories

David Alan Warburton
Topoi, Berlin, Germany

Synonyms

[Ancient color terminology](#); [Early color lexicons](#);
[Early expression of color in language](#)

Definition

Color terms or partitions of color denotata evidenced in ancient language artifacts.

Words and Hues, Languages and Time: An Overview

The sources for understanding the earliest color terms and categories are from the lands to the east and south of the Mediterranean Sea. Evidence of color categories from *proto-cuneiform*, *Sumerian*, *Egyptian*, and *Akkadian* in Mesopotamia and Egypt (from the end of the fourth millennium BC onwards) is followed (during the second millennium) by *Greek* in the West and *Chinese* in the East. Linguistic terms relating to color are present in all these languages.

What is known about the earliest color categories is derived from artifacts and texts. The use of color goes back at least 100,000 years, but the origins of color vocabulary lie in the period since roughly 8000 BC (=10,000 years ago), and the earliest texts (from ca. 3200 BC) appear millennia later. By comparison with the languages discussed here, virtually all other languages are much younger (e.g., *Hebrew*, *Latin*), or contemporary (e.g., *Eblaitic*, *Hittite*, *Ugaritic*), but linguistically related to the languages discussed here.

Vocabularies in the earliest preserved languages offer representative and definitive evidence concerning the origins of color categorization and its linguistic expression, as well as allowing evaluations of different steps in the process of abstraction and the early linguistic partitioning of perceptual color space. (For linguistic and historical details, see Refs. [1–7].)

Color Terminology

Black and White, Bright and Dark

The earliest color lexicons from languages of the Middle East and the eastern Mediterranean (*proto-cuneiform*, *Sumerian*, *Akkadian*, *Egyptian*, and early *Greek*) have words signifying “dark” and “light” as well as terms denoting something closer to “black” and “white.” Yet like most of the more specific terms for “black” in most ancient languages, even classical *Greek melas* had a semantic range including “black” and “dark” [8] that encompassed some regions described in *English* as “brown.” In general, virtually all of

the linguistic glosses for “black” and “white” had category central exemplars or “foci,” but some also included a broader “light” and “dark.” Despite some overlap, words similar in meaning to *English* “black” and “white” are different from glosses such as “light,” “shining,” “gleaming,” “sparkling,” and “dark” and “gloomy,” respectively, as the latter glosses appear to not have category foci.

Red

In all ancient languages, there is a very clear tendency for a division of the reddish continuum of color experiences into several different hues. While the color typically labeled “red” in modern *English* language usage was among the earliest distinguished in art and written artifacts, the concept of a category of “red” as a distinct linguistic unit was not dominant in the second and third millennia BC. In addition, there is clear evidence, in both earlier *Greek* and earlier *Egyptian*, of a tendency for a “red” and “white” opposition [3, 6, 9]. Moreover, in *Akkadian* and *Greek*, the word for a generic “red” was frequently used as a synonym for “colorful” or “colored” – and there is evidence suggesting this may be the case for *Egyptian* “red” as well.

Green, Green-Blue, and Green-Yellow

The *Egyptian* (*wādj*) and *Akkadian* (*warqu*) color terms that denote what today is referred to as “green” in *English* are derived from the same linguistic root but do not invariably denote the same color, since the Mesopotamian terms *sig* and *warqu* probably included “yellow” as well. The *Akkadian warqu* denoted both “green” and “yellow” appearances and was used to describe both vegetation and gold. The *Akkadian warqu* definitely did not mean “green-blue” or “green-yellow” [2]. Although the *Egyptian wādj* is related to the *Akkadian warqu*, the category centroid, or focus, of the *Egyptian* term was in “green” and neither “green-yellow” nor “green-blue” – and certainly not “yellow” [4]. By comparison, the *Akkadian* color term *ḥašmānum* has been associated with “blue-green” (as well as “light blue”) [10].

There was no generic word for our “green” in *Greek*, although *xlōros* eventually came to mean

something like “green,” but the earliest use of color terms in *Greek* was not specific; green was divided and not dominated by a “green-blue.”

Qīng (“dark,” “green,” or “blue”) has not yet been found in the earliest *Chinese* inscriptions. Since the first millennium BC, *qīng* was used for “dark,” “blue,” and “black”; only slightly later, the word *lǜ*, today’s “green,” also appeared, so that to some extent “green” has since been divided into “light” and “dark” (green) [7].

Blue

Some of the color words preserved in the earliest Semitic languages (e.g., *uqnu*, “lapis lazuli” or “dark blue”) are loanwords for materials from other unknown older languages. Other terms – later shared in different languages – were possibly words (but certainly not the corresponding category “abstractions”) corresponding to “red” and “green” which may have existed in the early Neolithic of the Near East, perhaps 10,000–12,000 years ago, prior to the documentation of language [10].

In *Akkadian* (*uqnu*) and *Egyptian* (*xsbdj*), terms for lapis lazuli designated “dark blue.” In *Greek*, a term (*kyaneos*) for blue appearances is derived from the *Akkadian*. In *Egyptian*, turquoise (*mfkāt*) denoted “light blue.” *Akkadian* used several terms for “light blue” (including *ḥašmanum*, possibly from the *Egyptian* word for amethyst, *ḥsmn*, which was not used as a color word in *Egyptian*). *Chinese* *lán* is a term for “blue” colors but appears quite late (in comparison to, e.g., “red,” “white,” “black,” “yellow”). As a category, the modern *English* term “blue” evolved to ultimately eclipse the distinction (still preserved in *Russian*) between light and dark blue. Through the second millennium BC, color terms are mostly rooted in materials – most of which were later eclipsed with abstract words.

Yellow

Early evidence of terms glossing “yellow” is less common but documented. In *Egyptian*, the word for “gold” (*nb.w*) was occasionally used to represent yellow. The linguistic usage of “gold” (*nb.w*) for “yellow” is not common in *Egyptian*; in *Egyptian* painting, however, the color yellow was

frequently used to depict what was intended to represent gold where required and also the sun on occasion [11]. The usage of gold in *Akkadian* texts with the meaning of “yellow” is rarer than in *Egyptian*, and the form was a simile. In contrast to this earliest material, the later *Greek* words based on *xrusos*, “gold,” were frequently used to designate a color which was most probably “yellow”; significantly, *zanthos* “yellow” is also documented.

Table 1 offers an impression of what can be identified in the way of colors in these earliest languages.

Material Color

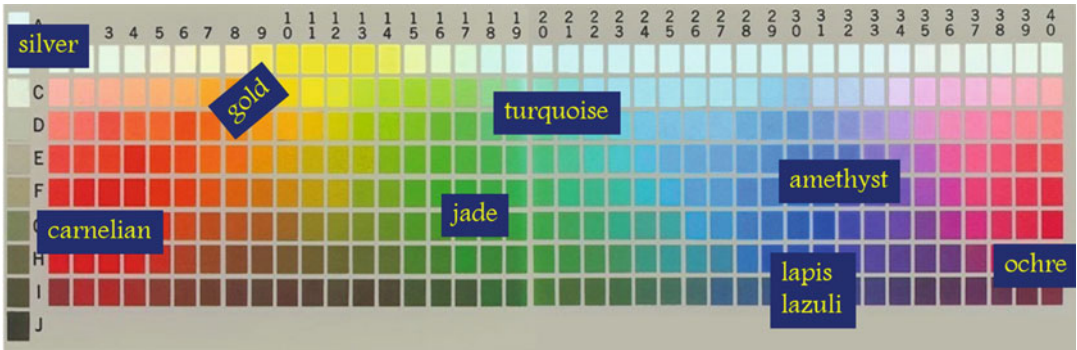
One of the greatest obstacles to understanding the nature of the earliest origins of ancient color terminology (in the Ancient Near East) and the origins of abstraction (in *Greek* and *Chinese*) is appreciating that in the earliest usage the ancients did not classify the world according to modern terminological divisions of the diversity of visible light, but rather that they initially used their perceptions of precious materials to express many of the colors they perceived. Thus, their base was the colors of the materials which they then applied to other domains. This eventually created the basis for abstract color terminology. However, the origins render the discussion complicated. (For references for this section, see [10, 11].)

Figure 1 approximates the relationships between color appearances and color terms.

A good example of color abstraction is “gold.” In several languages, the word for “gold” implies color associations mostly from “yellow” portions of color space but can also imply “orange” (and even “red”) areas of color space. In contemporary terms, the color “gold” is generally considered similar to colors in the yellow range. By comparison, ancient users of color lexicons were unlikely to associate “gold” with “yellow” because ancient languages tended to strictly associate the color term “gold” with the physical materials of gold metals. This suggests variation in the underpinnings of modern and ancient concepts associated with terms denoting golden color appearances.

Ancient Color Categories, Table 1 Summary of identified color terms from the earliest known languages

Language:	Proto-cuneiform	Sumerian	Egyptian	Akkadian	Chinese	Chinese	Greek	Greek
Time period	4th millennium	3rd millennium	3rd–2nd millennia	3rd–2nd millennia	2nd millennium	1st millennium	2nd millennium	1st millennium
Color terms:								
“White”	BAR, ?U ₄	babbar	ḥdj	pešum	bái	bái	re-u-ko	leukos
“Black”	?GI	mi, gíg, ġi ₆	km	šallamu	hēi	hēi	ma-ra	melas
“Bright red”	si/u ₄ , NE ₆	si/u ₄	dšr	sāmu	chi	chì	po-ni-ko-ro	erythros
“Green”	sig ₇	sig ₇	wādj	warqu		qīng		xlōros
“Yellow”	?GI	sig ₁₇	nb.w		huáng	huáng	ka-sa-to	zanthos
“Dark blue”		su ₆ -za-ġin-na	xsbdj	uqniātum			ku-wa-no	kyaneos
“Light green”						lù		prasinos
Other “Blues”			jrtjw, mfkāt	ḥašmānum, pelum, tukiltum		lán		glaukos
Other “Reds”		ša ₄ , su ₉ , sa ₅	rwdj, mss, tjms, mroš	ruššu, ḥuššu		jiàng, hóng	er-ru-to-ro	rhodeis
“White,” “light,” “bright”	(babbar?)	ara, bar, ḥáda, dalla, kára, kug, píriġrín, še-er, tán, zalag	ḥdj, tjḥnt	pešu	míng, qǐ	míng, qǐ		
“Dark”	(gege?)	dara, ge, gíg/gege, kúikku, mí, súš/sú	kk.w	šallamu, tarku				



Ancient Color Categories, Fig. 1 Examples of physical materials related to ancient color terminology displayed in congruent regions of a Munsell Color Chart

Thus, color and material were related – and led to a linguistic partitioning of color in a fashion differing from our own “modern” understanding (as reflected in, e.g., *English*).

The earliest documented use of red (in the form of ochre) dates back 100,000 years – and thus long before the earliest documented written sources. Red and yellow ochre, along with black soot, is

easily recognizable in the Paleolithic cave paintings from the Upper Paleolithic, from ca. 35,000 years ago. Green and blue are strikingly absent until tens of thousands of years later.

Stones with material properties producing green hues obtained prominence beginning approximately 10,000 years ago. Blue lapis lazuli appears gradually in South and Western Asia starting in the sixth millennium; red carnelian and blue-green turquoise are seen slightly later, as is jade in Europe and China. Gold and silver appear since the fifth and fourth millennia (respectively) in Europe and the Near East.

Many kinds of precious materials contributed to the concept of “shining” and “gleaming.” Significantly, the word for “white” or “bright” in *Sumerian* (*babbar*) and in earlier *Egyptian* (*ḥdj*) was the word for silver. In *Greek*, silver played a role in a word for “shining” (*argos*) – with yet another etymology. In his list of Indo-European etymologies, Shields seems to suggest that the abstract terms “bright,” “gleaming,” etc., are the basis for many words that eventually became abstract color words. However, it may be that the evidence suggests the opposite. That is, precious materials (rather than abstract “brightness”) are more likely to be linked to the origins of abstract color words. For example, *Sumerian* *šuba* means “agate” or “precious stone,” but also means “shining.” And *Sumerian* *kug* is the term for the metal silver and “bright” and “white.” Another *Egyptian* word for “gleaming” or “dazzling” can be related to *tjhn.t*, a word later used for faïence, but probably originally referring to a form of naturally generated glass. In the ancient languages, there are many more words for appearance properties of “gleaming,” “shining,” “brilliant,” “bright,” etc., than can be fully inventoried in the present survey. In general, in ancient languages, ideas of shining and color originated from associations with precious materials although many did not lead to color words.

Yet examples of terms for precious materials that did have color meaning were nevertheless abundant: *Egyptian* *ḥdj* means silver – and has been identified as the basic color term for “white” in early *Egyptian*. The *Akkadian* *sāmu* refers to “carnelian” but is what is interpreted as the basic

color term for “red.” *Egyptian/Akkadian* *wādj/warqu* are probably derived (through metathesis) from a widely used (Neolithic?) designation for jade (or a “greenstone”) that later became the basic color term “green” (in *English*; *grün* in *German*, etc.; see below). *Egyptian* *xsbj* and *Akkadian* *uqnu* meant lapis lazuli but were used for “dark blue” – and the latter is related to the *Greek* *kyaneos* for “blue”; *Egyptian* *mškāt* was turquoise but used for “light blue.” *Egyptian* *nb.w* was “gold” but used for “yellow.” *Egyptian* *ḥsmn* meant only amethyst in *Egyptian*, but *Akkadian* *ḥasmānum* designated not only a stone but also a “blue-green” or “light blue” and so on.

Of these materials, several eventually led to abstract color words, but usually only in the languages into which they were imported. This process seems to have begun in the second millennium BC, but only began to have systematic effects from the first millennium BC onwards.

Linguistic Issues in Ancient Color Naming

Black and White, Dark and Light

Significantly, a word for “darkness” shared in *Sumerian* (*kuku*) and *Egyptian* (*kk.w*) means that the word must have been of great antiquity, since it fed into two distinct language families. Yet these glosses for “darkness” were not related to the words used for “black” in either language; the origin of “black” in both languages lay elsewhere. The *Chinese* term for black also meant dark but was otherwise used as an adjectival modifier. The *Chinese* term for white *bái* did not signify “bright” or “clear” (which were *míng* and *qǐ*).

Salience

It is significant that material, not color, properties of several terms are what was apparently salient throughout the history of the languages. In medieval *Coptic*, the early *Egyptian* for “white,” *ḥdj*, is replaced by *ūbaš* (of which the etymological meaning is to “shine”). The salient meaning of *ḥdj* lasting through *Coptic* is not “white,” but the material “silver.” In *Chinese*, the term for “red” is not salient in the sense described by Berlin and

Kay [12], since the common *Chinese* word in the first two millennia of the language is *chì* and not *hóng* (which later replaced *chì*). In Mycenaean *Greek*, the main word for “red” is *po-ni-ko-ro* (later *phonikos*, a loanword referring to the Phoenicians who furnished the red dye) rather than *er-ru-to* (which gave rise to the later basic color term *erythros*).

Etymologies, Materials, and Loanwords

Egyptian and *Akkadian* “green” are most likely the same word and probably at the root of *English* “green” (sharing the radicals *r* and *q/k/g/ġ*). An argument could be made for the diffusion of “red,” where *Akkadian* *ruššu* is probably the same as *Greek* *erythros* and *Italian* *rosso*. Thus, these are ancient concepts that have moved between languages.

As loanwords, precious materials also play an important role. In *English*, lapis lazuli has contributed the words “azure” (derived from *Persian* *lazuward* for lapis lazuli) and “cyan” (originally a loanword imported into *Akkadian* as *uqnu* for lapis lazuli and subsequently to *Greek* *kynaeos*).

Precious Materials, Loanwords, Abstraction, and Grammar

The classical *Greek* term *kyaneos* is derived from *ko-wa-no*, which was probably the Mycenaean *Greek* word for glass paste. The word *ko-wa-no* itself was derived (via *Ugaritic* or *Hittite*) from *uqnu*, the *Akkadian* for lapis lazuli (itself a loanword in *Akkadian*). In the Aegean *ko-wa-no*, the Asian *uqnu* was used to designate the artificial material glass (as opposed to the semiprecious stone lapis lazuli). Thus, this process of linguistic exchange apparently gave birth to the abstract color term *kyaneos* – but only in *Greek* of the first millennium BC. Earlier, its primary role was that of designating a precious material which gave birth to the color term. In *Akkadian*, the nouns lapis lazuli and cornelian appear regularly together in the same texts, along with *huršāsu* and *hurāšānū*, “gold” and “golden.” In Mycenaean *Greek*, this same word appears as *ku-ru-so* “gold,” and later as *Greek* *xrusos*, where it is used as a color word – even though *Greek* had a word for “yellow” which can be traced back to the second millennium.

In *Egyptian*, the first reference to the “sky” as having a color is in *Coptic*, the latest stage of the language, in the first millennium AD. Prior to this, the sky was described as being turquoise or lapis lazuli (rather than having a color itself). Although vegetation was known to be “green,” the word for green is frequently associated with a classifier signifying a stone. Lapis lazuli was treated as a precious stone in these societies, only gold and lapis lazuli had prices higher than silver (which served as money). The statues of the gods in the temples were not made of granite, but rather of gold, silver, lapis lazuli, turquoise, ivory, etc. These materials were the origins of color words.

Although often considered “abstract,” the adjectives for colors in *Sumerian* were never attached to the *nam* abstract determinative. Given the difficulties of understanding adjectives in *Egyptian*, Schenkel used verbs as his criterion, i.e., *ḥdj* “to be white,” *km* “to be black,” *dšr* “to be red,” and *wādġ* “to be green.” It is probably true that these words were verbs, but in some cases, they are also used as adjectives. The noun *xsbdġ* “lapis lazuli” “dark blue” was a metaphor, a simile, and an adjective, but not a verb. The same is true for *mġkāt* “turquoise” used for “light blue.”

In *Akkadian*, color words are all adjectives. However, in the case of *uqnu*, the principal meaning is the noun “lapis lazuli”; an identical adjective means “lapis lazuli color(ed)”; *uqnātu* is an adjectival form with the same meaning. The same is true of *sāmtu* “cornelian” and *sāmtu* “redness” and *sāmu* “red.”

In *Chinese* and *Greek*, color words are largely adjectives. Some of the *Greek* terms can be traced back to Near Eastern materials. Of the languages discussed here, *Chinese* is the only language with a completely abstract color vocabulary, where hues and terms match, more or less. Although silk appears as a component in the writing of some color words in *Chinese*, even as a component, jade played an even more marginal role. Thus *Chinese* color terms never bore primary relations to precious, natural object categories; the evolution of *Chinese* color vocabulary differed fundamentally from that of the West – but the results are similar to those in *Greek*, implying diffusion.

The Mediterranean languages usually had a word for “color” as a phenomenal experience; however, the words were not restricted to a single meaning in terms of hue. In Egyptian, the word *jwn* is probably derived from the designation for a “vein” of ore (meaning the material color which was visible), but it also meant “character” of a person in the figurative sense (what was hidden under the surface). By contrast, although later *Chinese* has such a word, in the earliest *Chinese*, no term for color as such has been discovered; in later *Chinese*, the suffix *-si* is frequently attached to color words.

Theory of Ancient Color Term Emergence

Debate regarding the sequence of emergence in early color terminology is constrained by limited artifact sources and historical evidence of ancient color term sequences. Moreover, color term emergence theories (see also Encyclopedia of Color Science and Technology entry on “► [Berlin and Kay Theory](#)”) that describe color language hierarchies found in contemporary and very recent language data cannot easily be applied to ancient color term emergence as they are based primarily on modern observations of modern phenomena and therefore do not take account of the relevant ancient data discussed here. This makes emergence theories of modern color lexicons problematic as models of both ancient partitions of gradient color continua and ancient color category acquisition sequences.

Warburton [10] argues that the earliest acquisition sequence was not an issue related to individual specific languages, but rather a shared phenomenon based on shared terminology, concepts, and materials. In this scheme, “ochre” (as dark red) will have been the earliest color used (in the Paleolithic, 100,000 years ago), followed by “greenstones” (in the earliest Neolithic, from 10,000 years ago), and “gold,” “silver,” lapis lazuli,” “carnelian,” and “black” (in the following late Neolithic and earliest Bronze Age) and then by turquoise (before 4,500 years ago).

Of extreme importance is that in ancient Egypt, the sun is occasionally called “red” and painted as

such. This does not imply a “red-yellow,” but rather the use of a single term to designate different colors, as with the Mesopotamia “green” used for yellow and green. The reasons behind this do not lie in perception so much as expression (based on gold, bronze, sun, vegetation, etc.).

Loanwords and common etymologies play important roles in emergence of ancient color terms in that, as detailed above, many of the earliest words are related to designations of materials. Terms for “white” and “black” do not stand at the origins, and the red, blue, and green ranges were partitioned before being lumped together. Only later did the further color words begin to crystallize (in the last two millennia). The concept of “blue-green” is not documented in the earliest languages; it appears only recently in *Chinese* and does not universally evoke color significance among informants [7].

Color in Art and Language

Of particular interest for the understanding of the evolution of color terminology is the fact that for *Egyptian*, *Akkadian*, and *Sumerian* (which all date to before the middle of the second millennium BC), it is agreed that the color terminology did not match the colors that were used in the contemporary art. In contrast to this, already for Mycenaean *Greek* and classical *Greek* (thus, from the second half of the second millennium BC onwards), Blakolmer is persuaded that “it is hardly surprising that in the use of color, language and representative art are two sides of the same coin” [6]. Thus, in comparing Bronze Age paintings with Roman-era painting, for Blakolmer, the use of color in painting reflects the development of a vocabulary in the Greek language. By comparison, in *Egyptian* art, there are many colors used in painting that are not accounted for by any kind of vocabulary [11].

Summary

Precious materials played an important role in the development of abstract color terminology,

usually in the form of loanwords during the period well after the earliest recorded languages. Warburton suggests that the materials dominated and that the concept of abstract color terminology never developed in the ancient Near East; yet the flow of loanwords meant that abstract terminology may have already been embryonic in the Aegean by ca. 1400 BC [11].

There is a clear difference between the descriptive uses of color in literature and economic texts. Although the sources for various languages are not equally balanced in terms of genre, it is significant that there is little evidence of color from the earliest administrative texts in proto-cuneiform whereas color terminology appears in the administrative texts written later in all other languages (where it also appears in other genres of text).

The gradual emergence of color as a means of categorization was thus a historical development that contributed to the emergence of abstraction in relatively recent historical times. In the art of the Paleolithic and Neolithic societies, few colors appear, whereas there is an abundance of color in the Near Eastern Bronze Age – well beyond what appears in evidence from language artifacts. By comparison, in the Greek world, the use of color in art and language marched hand in hand. It is clear that the Greeks acquired color concepts and terminology from the Near East through the diffusion of materials, ideas, and terminology. These in turn influenced the color lexicons of more recent languages.

3. Schenkel, W.: Die Farben in ägyptischer Kunst und Sprache. *Z. Ägyptische Sprache* **88**, 131–147 (1963)
4. Schenkel, W.: Color terms in ancient Egyptian and Coptic. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modelling*, pp. 211–228. Benjamins, Amsterdam (2007)
5. Bulakh, M.: Basic color terms from proto-semitic to old ethiopic. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color: Interdisciplinary Multilevel Modelling*, pp. 247–261. Benjamins, Amsterdam (2007)
6. Blakolmer, F.: Zum Charakter der frühägäischen Farben: Linear B und Homer. In: Blakolmer, F. (ed.) *Österreichische Forschungen zur Ägäischen Bronzezeit 1998: Akten der Tagung am Institut für Klassische Archäologie der Universität Wien 2.–3. Mai 1998*, pp. 225–239. Wiener Forschungen zur Archäologie, Vienna (2000)
7. Xu, W.: *A Study of Chinese Colour Terminology*. Lincom, Munich (2007)
8. Liddel, H.G., Scott, R., Jones, H.S.: *A Greek-English Lexicon*. Oxford University Press, Oxford (1958)
9. Baines, J.: Color terminology and color classification: ancient Egyptian color terminology and polychromy. *Am. Anthropol.* **87**, 240–262 (1985)
10. Warburton, D.A.: The theoretical implications of ancient Egyptian colour vocabulary for anthropological and cognitive theory. *Lingua Aegyptia* **16**, 213–259 (2008)
11. Warburton, D.A.: Colors in bronze age Egyptian art and language. In: Brinkmann, V., Primavesi, O., Hollein, M. (eds.) *Circumlitio: The Polychromy of Antique and Mediaeval Sculpture*, pp. 170–187. Liebighaus, Frankfurt (2010)
12. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)

Cross-References

- [Berlin and Kay Theory](#)
- [World Color Survey](#)

References

1. Englund, R.K., Nissen, H.J.: *Die Lexikalischen Listen der Archaischen Texte aus Uruk*. Gebr Mann, Berlin (1993)
2. Landsberger, B.: Über Farben im Sumerisch-Akkadischen. *J. Cuneif. Stud.* **21**, 139–173 (1967)

Ancient Color Terminology

- [Ancient Color Categories](#)

Apparent Magnitude – Visual Magnitude

- [Apparent Magnitude, Astronomy](#)

Apparent Magnitude, Astronomy

Ken Tapping

D.R.A.O., National Research Council Canada,
Penticton, BC, Canada

Synonyms

Apparent magnitude – visual magnitude;
Opacity – extinction

Definitions

- Elevation: angle between the geometrical horizon (as indicated by a spirit level) and the direction of the star or other object in question.
- Extinction: The degree to which starlight is weakened by the medium through which the light travels from the star to the observer. It is similar in meaning to “opacity” but is usually expressed in magnitudes. Zero means no extinction; otherwise the extinction indicates how many magnitudes fainter the object looks due to opacity of the interstellar medium and the atmosphere.
- Opacity: The degree to which light (or other optical radiation) is blocked when passing through the material in question. It varies between zero for something that is completely transparent and infinity for something completely opaque.
- Orthochromatic: This term refers to a type of photographic emulsion that was widely used in the earlier days of photography. It was very sensitive to blue and green light but insensitive to red light.
- Photometry: The measurement of quantities referred to light (visible optical radiation) as evaluated according to a given spectral luminous efficiency function (of the human eye). In astronomy it applies to the measurement of the brightness of astronomical objects.
- Photosphere: The layer in a star’s atmosphere from which almost all of the heat and light are radiated. Because it lies at a fairly abrupt

change in optical thickness, it looks almost like a solid surface and is often referred to colloquially as the “star’s surface.”

- Transmissivity: The fraction of the input energy passing through an attenuating device or layer. It varies from unity, for a completely transparent medium, to zero, for something completely opaque.
- Zenith angle: Angle between the local zenith and the direction in question.

Background

At the most fundamental level, the apparent magnitude of an object is a measure of how bright it looks. This is a function of the properties of the light source, its distance, the nature of the medium through which the light travels to the observer, and inevitably the properties of the observer or device used to make the measurement. When minimizing the subjective effects and trying to make the measurement more descriptive of the source and less related to the local circumstances of the observer, the parameter in question inevitably becomes less “apparent” but more useful. The history of “apparent magnitude” reflects such an evolution.

At a qualitative level, at a particular time and location, under the atmospheric conditions existing at that time, a specific observer can assess that some objects in the sky look brighter than others and possibly estimate by how much. This information may be useful to other observers present at the same location at the same time. However, recording such information for later consideration is difficult, and passing it in any usable form to another observer at another location for comparison of observations is even more so.

The development of the concept of “apparent magnitude” marks the beginning of an evolution toward producing quantities that are more intrinsic to the object being observed and can be combined with other observations. In the process, the term “apparent” becomes more a designator of a particular type of data rather than something literally true.

The flux from a distant star arriving at the top of the Earth's atmosphere is given by

$$F_{00}(\lambda) = \alpha(\lambda) \frac{W^*(\lambda)}{4\pi R^2} \quad (1)$$

where W^* is the energy output of the star, R its distance from us, and α is the transmissivity of the interstellar medium along the path from the star to Earth. Note the wavelength dependence of these parameters. In this discussion the “00” suffix identifies a quantity measured at the top of the Earth's atmosphere.

Because the same value of F_{00} would be incidentally equal upon all points at the top of the Earth's atmosphere, it would be common to all terrestrial observers. Therefore, in these discussions, F_{00} can be considered as the starting point.

In the development of the apparent magnitude concept, a number of issues arise. Firstly, since this is a comparative system, a defined calibration datum is required. In addition, starlight is not monochromatic; the atmosphere's transmissivity (or opacity) varies with wavelength and so does the sensitivity of the detectors, sensors, or measurement devices used.

Quantifying Apparent Brightness

The concept of assigning “magnitudes” to stars was first proposed by the ancient Greeks, and has been attributed to Hipparchos [1], who was a Greek astronomer, geographer, and mathematician who lived about 190–120 BC. Accommodating the natural nonlinear response of our eyes to brightness, he assigned a magnitude scale where the brightest stars were categorized as being of the First Magnitude ($m = 1$), and the faintest stars visible to the unaided eye were defined as being of the Sixth Magnitude ($m = 6$), with each intervening magnitude corresponding to half the brightness of the next brighter class and twice the brightness of the next fainter. In this system, if the energy fluxes reaching from two stars of magnitude m_1 and m_2 are respectively F_1 and F_2 , the relationship between these quantities is

$$\frac{F_1}{F_2} = 2^{-(m_1 - m_2)} \quad (2)$$

This classification scheme for the apparent magnitude of stars was used for some time. Ptolemy used it in *The Almagest*, [2], his compendium of mathematical and astronomical information. Since the magnitudes as defined are exponents, taking logarithms of Eq. 2 puts it into a more convenient form

$$\begin{aligned} \log_{10} \left(\frac{F_1}{F_2} \right) &= -\log_{10}(2)(m_1 - m_2) \\ &= -(0.30102996 \dots)(m_1 - m_2) \end{aligned} \quad (3)$$

so

$$(3.321928 \dots) \log_{10} \left(\frac{F_1}{F_2} \right) = -(m_1 - m_2) \quad (4)$$

The logarithm of the ratio of brightnesses is simply related to the magnitude difference. However having irrational numbers as the scaling factors is inconvenient, and in 1856 Norman Robert Pogson [3] redefined the magnitude scale in a more useful form. In Hipparchos' system, a first magnitude star is defined as being 64 times brighter than one of the sixth magnitude. Pogson changed the definition of the magnitude scale, defining a sixth magnitude star to be 100 times fainter than a first magnitude star. So the ratio of brightness between two adjacent magnitudes became

$$P = 100^{\frac{1}{5-1}} = 2.5118864315 \dots \quad (5)$$

This number has become known as *Pogson's Ratio* (P). Despite the apparent messiness of this number, it has an advantage in calculations. P is almost always used in the forms $\log_{10}(P)$ and $1/\log_{10}(P)$, which equal exactly 0.4 and 2.5, respectively. Thus, Eqs. 2 and 3 become

$$\frac{F_1}{F_2} = P^{-(m_1 - m_2)} \quad (6)$$

$$\begin{aligned}\therefore \log_{10}\left(\frac{F_1}{F_2}\right) &= -\log_{10}(P)(m_1 - m_2) \\ &= -0.4(m_1 - m_2)\end{aligned}\quad (7)$$

so

$$(m_1 - m_2) = -2.5\log_{10}\left(\frac{F_1}{F_2}\right)\quad (8)$$

Since the magnitude system describes relative brightnesses, it needs to be based upon a calibration reference: a star to which a magnitude value has been assigned. Pogson first chose Polaris (also known as the Pole Star, or North Star) as a calibrator, to which he assigned a magnitude of 2. However, when this star proved to be variable in brightness, he selected Vega, which he defined as having an apparent magnitude of 0. In practice, Vega is a better choice because it is available to observers over much of the world, and since it varies in elevation with time at any given site, other stars at similar elevations at that time can easily be set up as secondary or transfer calibration standards.

The nature of the magnitude scale makes it able to accommodate stars brighter than magnitude 0, which gives them negative magnitudes. For example, Sirius, the brightest star in the constellation of Canis Major, and the brightest star in our sky other than the Sun, has a magnitude of -1.5 and the Sun a magnitude of about -26 . Fainter stars, with magnitudes greater than 6 can also be accommodated. Binoculars or telescopes will reach objects far fainter than the eye can discern, which therefore have magnitudes greater than 6.

Wavelength Issues

Stars are not sources of monochromatic light; they emit a broad spectrum extending across the nominal 400–800 nm band of visible wavelengths and beyond in both the longer and shorter wavelength directions. The apparent color of a star depends upon the wavelength at which the peak brightness occurs, which in turn depends upon its photospheric temperature. The consequent star colors

range from red to blue (and into the infrared for very cool stars).

The interstellar and atmospheric transmissivities are wavelength dependent, and in general so is the sensitivity of devices typically used to measure light intensity. This variability is particularly marked in the case of the human eye, which may differ between individuals and also with the same individuals, depending upon health, fatigue, etc. A classic example of the need to take this wavelength dependence into account is evident in old photographs of the constellation of Orion, made using orthochromatic photographic emulsions, which have extremely low sensitivity to the red end of the visual spectrum. Betelgeux (α Orionis), the brightest star in the constellation (a red giant), is not visible, whereas Rigel, a blue-white star, is very conspicuous. A more extreme, although non-astronomical example is in a photograph (again using an orthochromatic emulsion) of the cab of an early steam locomotive, looking into the open firebox, which shows no fire at all, although the boiler pressure gauge shows a vigorous fire must be present. To maintain the utility of the apparent magnitude observations, it is necessary to deal with wavelength dependencies, but this needs to be done in a manner that is both standardized and does not overly complicate the measurement process.

Stellar Spectral Types

The concept of “stellar spectral type” is, to a large extent, an indicator of a star’s photospheric temperature, i.e., its color. This way to describe stars is therefore intricately mixed with the magnitude/color issue, so a short description is included here. Stars come in a wide range of colors, brightnesses, and masses and with different spectral line signatures, so it was deemed useful to develop some sort of classification framework toward providing a coherent overall picture of star types and stellar evolution. The first attempts were made at a time stellar astrophysics was not a highly developed subject. One early system to classify stars used alphabetic labels. However, as knowledge improved, the trend was to classify on the basis

of photospheric temperature and spectral line signatures. Some classes were found to be redundant, and the sequence of remaining classes had to be rearranged, so they were no longer in alphabetical order. The new system is in order of decreasing temperature and became O, B, A, F, G, K, M, L, T, Y. This arbitrary sequence of letters is remembered using the mnemonic “Oh Be A Fine Girl and Kiss Me. Like This? Yes!” This sequence is

further subdivided using numerical interpolation (e.g., G2, A5, M0, etc.). The classes and their corresponding photospheric temperatures and colors are summarized in Table 1. In the case of the Y0 class at 600 K, these cool stars are usually described as “brown,” although the color would be more accurately described as an extremely dull red or perhaps not visible at all at optical wavelengths since its spectral peak lies in the infrared region.

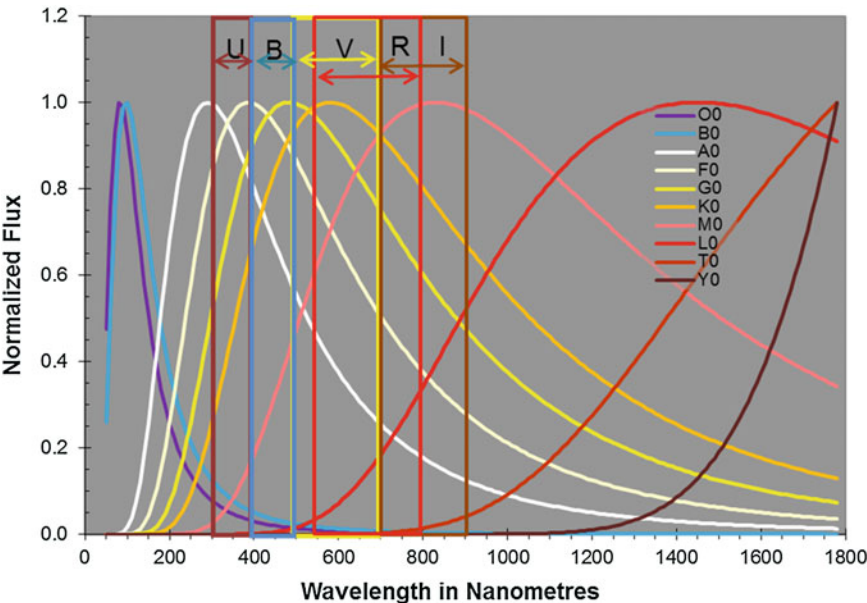
Apparent Magnitude, Astronomy, Table 1 Photospheric temperatures/colors and spectral types

Photospheric temperature (K)	Spectral type	Color
35,000	O0	Blue
30,000	B0	Blue white
10,000	A0	White
7,500	F0	Pale yellow
6,000	G0	Yellow
5,000	K0	Orange
3,500	M0	Red
2,000	L0	Dull red
1,200	T0	Very dull red
600	Y0	“Brown”

The UBV(RI) Magnitude System

Stellar spectra are generally (almost) blackbody spectra corresponding to the stars’ photospheric temperatures, with superimposed absorption lines. Magnitude determination applies to the thermal emission only. Figure 1 shows blackbody spectra for the photospheric temperatures and spectral types listed in Table 1.

The required amplitude and wavelength-related information is included in a star’s spectrum. Stellar spectra are the primary data source for most stellar astrophysics. However, each



Apparent Magnitude, Astronomy, Fig. 1 Blackbody spectra corresponding to the photospheric temperatures of stars of spectral types O0, B0, A0, F0, G0, K0, M0, L0, T0,

and Y0. The nominal bandwidths of the U, B, V, R, and I filters are shown. The spectra are all normalized

Apparent Magnitude, Astronomy, Table 2 Nominal filter bandwidths for stellar magnitude measurements

Band designation	Band name	Low end (nm)	High end (nm)	Bandwidth (nm)
U	Ultraviolet	320	400	80
B	Blue	400	500	100
V	Visible	500	700	200
R	Red	550	800	250
I	Infrared	700	900	200

spectrum contains a lot of information, and when studies involving large numbers of stars, or where measurements that are relatively simple and reproducible are needed, an alternative system would be useful.

To meet these needs, Johnson and Morgan [4] developed the three-filter, UBV system, which is still widely used. This method uses magnitude determinations made using three standard filters, designated U (ultraviolet), B (blue), and V (visible) respectively. To better describe cool, red stars, two additional bands have been added: R (red) and I (infrared). The objective is to describe the thermal spectrum using a small list of standardized parameters. The nominal wavelength ranges (bandwidths) are listed in Table 2.

The shapes of the filter passbands for devices currently in use are far from rectangular, but there has been an effort to at least standardize the filters to improve comparability of data between sources. The I and R bands are not so cleanly delineated. The I band overlaps the V band, and the R band overlaps the I band. Some institutions have added additional filter bands to suit their particular research needs.

Table 1 also shows the wavelength ranges passed by the U, V, B, R, and I filters. These rectangular passbands are not representative of the passbands of existing filters. The bandwidths are also shown using arrows, so the overlapping bands are more clearly discernible.

For most purposes, the U, B, and V magnitudes are sufficient, and it is possible to describe the photospheric thermal spectrum of most stars using V and the magnitude differences U–B and B–V. Since the magnitudes are logarithmic quantities, the differences actually describe brightness ratios, i.e.,

$$(U - B) = -2.5 \log_{10} \left(\frac{P_U}{P_B} \frac{\Delta\lambda_U}{\Delta\lambda_B} \right) \quad (9)$$

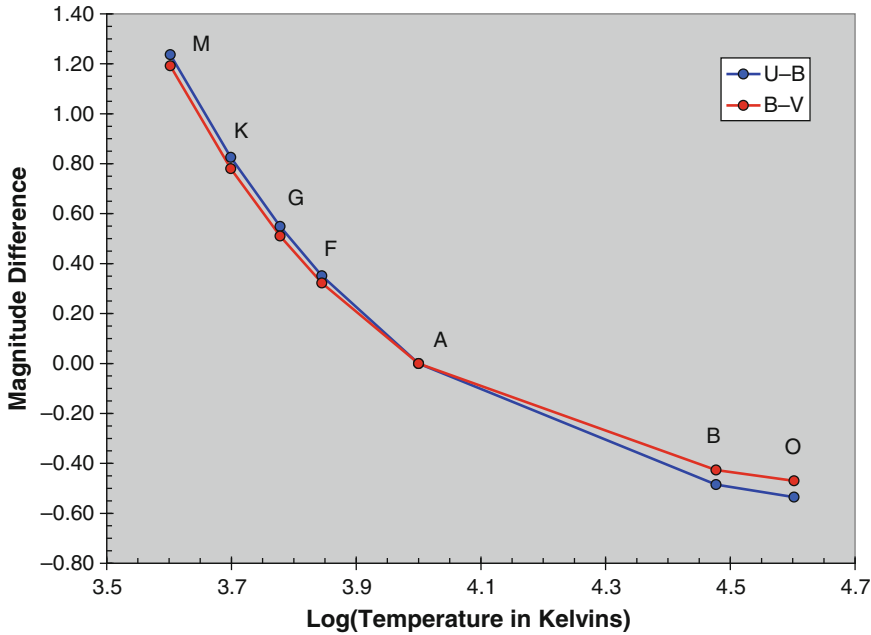
and

$$(B - V) = -2.5 \log_{10} \left(\frac{P_B}{P_V} \frac{\Delta\lambda_B}{\Delta\lambda_V} \right) \quad (10)$$

where P is the total power passing through the filter. Conventionally, the values are offset so that a star of spectral type A0 is characterized by U–B and B–V = 0. If the filter passbands are assumed to be rectangular, the spectra in Fig. 1 produce the U–B and B–V values shown in Fig. 2. It can be seen that these values provide a good estimator for the star's photospheric temperature. The letters above the data points on the graph are the spectral types of stars with those particular photospheric temperatures. Since stellar spectral types are temperature dependent, the UVB photometric system provides an estimator for spectral type.

However, in reality, filter passbands are not rectangular, and stellar flux densities may vary dramatically with wavelength, so simply averaging over assumed rectangular passbands will not produce useful results. Unfortunately, making filters with rectangular passbands is very difficult, and the more effort dedicated to achieving this also makes the filters harder to reproduce, which is a critical issue where standardization between instruments and observatories is important.

For this reason, a set of standardized filters are used. These are generally those used by Johnson and Morgan [4]. They do not have rectangular passbands but have the great advantage of being



Apparent Magnitude, Astronomy, Fig. 2 U-B and B-V plotted against the logarithm of the photospheric temperature for the blackbody spectra in Fig. 1. The

spectral types of the stars modeled are shown. In each case the types are M0, K0, etc.

fairly easy to reproduce. The wide adoption of these standard filters for stellar photometry makes it possible to combine data from different sources.

magnitude”: a magnitude measurement made using a V filter.

Absorption and Scattering

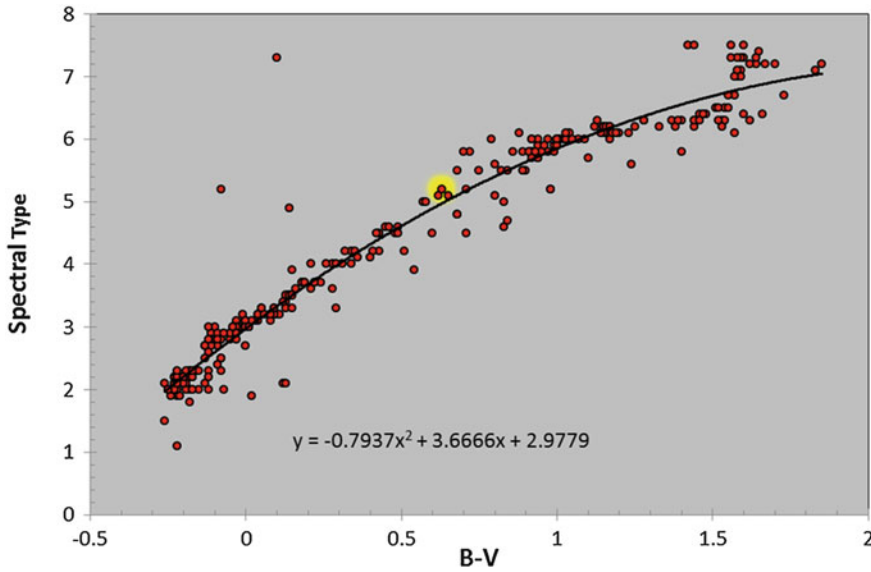
B-V and Stellar Spectral Types

Figure 3 shows a plot of spectral type against B-V for the stars listed in the bright star section of the Observer’s Handbook of the Royal Astronomical Society of Canada (RASC) [5]. For convenience in plotting, the spectral types O0, B0, A0, F0, G0, K0, and M0 are represented by the numbers, 1, 2, 3, 4, 5, 6, and 7 respectively. The subdivisions within each spectral class are coded as a decimal; for example, stars of spectral classes O5, A0, F6, G2, K5, and M3 would respectively be represented as 1.2, 3.0, 4.6, 5.2, 6.5, and 7.3. The point surrounded by the yellow halo is the Sun. The empirical equation can be used to estimate the spectral type from B-V.

Today, the term “apparent magnitude” is generally taken to be equivalent to “visual

To reach terrestrial observers, the starlight has to travel through the interstellar medium and then through the Earth’s atmosphere. Both of these media contain dust, molecules, and atoms that absorb or scatter some of the starlight, with the extent of these processes usually being wavelength dependent. The result is generally a weakening and reddening of the light. There are three main processes:

- Scattering at wavelengths coinciding with molecular or atomic resonances, where the directional light drives electrons to higher energy states and then is reradiated isotropically.
- Rayleigh scattering, which is scattering by particles smaller than the wavelength of the light (mainly atoms and molecules). This is



Apparent Magnitude, Astronomy, Fig. 3 Numerical representation of spectral types plotted against B–V values for the stars in the RASC list of brightest stars

strongest at shorter wavelengths and is what causes the sky to look blue and sunsets to be rich in reds and golds.

- Scattering off larger particles. This shows much less of a variation with wavelength. In the atmosphere the main scattering agents are aerosols, composed of small droplets of water or pollutants. This mode of scattering is less likely in interstellar space but may occur where there are clouds of larger particles. Fog and mist are good examples.

If an intrinsically stellar measurement is needed, then both the interstellar and atmospheric effects on the starlight have to be determined and removed. Although even the densest interstellar clouds are more rarefied than the best achievable laboratory vacuum, the very large distances the light has to traverse means it has to pass through an enormous amount of material. However, in this article the emphasis is on making measurements that are comparable between observers and observing sites, so the measurements need only be referred to the top of the Earth's atmosphere, so interstellar effects are common to all observers and not further discussed here.

The attenuation imposed by the atmosphere is by processes similar to those attenuating and scattering the light in the interstellar medium. However, an important difference is that in the case of the interstellar medium, the light path from star to the Earth is common to all observers and changes very slowly with time, whereas in the case of the atmosphere, the rotation of the Earth changes the light path through the atmosphere quickly enough to offer a means to estimate the stellar magnitude that would be observed in the absence of atmospheric attenuation. It is assumed that the transmissivity of the atmosphere is related to the total mass of atmosphere that the light passes through. If we look in the direction of the zenith (zenith angle = 0), we are looking through a certain total mass of air:

$$M_0 = \int_0^\infty \rho(h)dh \quad (11)$$

where $\rho(h)$ is the atmospheric density as a function of height. The “0” suffix indicates a quantity measured when looking at the zenith. The *transmissivity* of the atmosphere (the fraction of light arriving at the top of the atmosphere that

reaches the ground) from an object at the zenith is defined as

$$F_0(\lambda) = F_{00}(\lambda)\exp(-\eta(\lambda)M_0) \quad (12)$$

where η is an absorption parameter and M_0 the mass of air through which the light is passing. The flux density from the star is a function of wavelength and so is the atmospheric attenuation.

Assuming the magnitude measurements are restricted to zenith angles smaller than about 65° (elevation angles larger than 25°), the atmosphere can be treated as a stratified slab. In this case, the air mass through which the light passes is a simple function of zenith angle:

$$M_z = M_0 \sec(z) \quad (13)$$

It is unlikely that there will be any need to separate M_0 and γ , so they can be combined into a single parameter $\beta(\lambda)$, and so for a zenith angle z

$$F_z(\lambda) = F_{00}(\lambda)\exp(-\beta(\lambda) \sec(z)) \quad (14)$$

We have two ways to define the effect of the atmosphere on the light. The *transmissivity* (α) is the fraction of the incident light reaching the ground:

$$\alpha = \frac{F_z(\lambda)}{F_{00}(\lambda)} \quad (15)$$

An alternative is define an *opacity* (ξ), where

$$\xi(\lambda) = \beta(\lambda) \sec(z) \quad (16)$$

So

$$\begin{aligned} \alpha &= \exp(-\xi) \\ \text{or} \\ \xi &= -\log(\alpha) \end{aligned} \quad (17)$$

As the atmosphere ranges from completely transparent to completely opaque the transmissivity (α) ranges from one to zero and the opacity (ξ) goes from zero to infinity.

The slab atmosphere model offers a means to estimate the magnitude of the star in the absence

of the atmosphere, using a method attributed to Pierre Bouguer and discussed by Dufay [6]. Using Eqs. 7 and 12,

$$m_z(\lambda) - m_{00}(\lambda) = -2.5 \log_{10} \left(\frac{F_{00} \exp(-\beta(\lambda) \sec(z))}{F_{00}} \right) \quad (18)$$

So

$$m_z(\lambda) = m_{00}(\lambda) + 2.5 \log_{10}(e) \beta(\lambda) \sec(z) \quad (19)$$

where m_z is the measured stellar magnitude at a zenith angle z and m_{00} is the magnitude of the star before it is attenuated by the atmosphere. Equation 16 shows a simple, linear relationship between m_z and $\sec(z)$. Therefore, if a series of magnitude measurements for a star are made over a range of zenith angles, a plot of m_z against $\sec(z)$ will have a slope $2.5 \log_{10}(e) \beta(\lambda)$ and intercept $m_{00}(\lambda)$, the magnitude of the star as it would be measured at the top of the Earth's atmosphere. This process is repeated for the U, B, and V filter bands, to obtain the U, B, and V magnitudes corrected for atmospheric absorption.

Summary

“Apparent magnitude” is literally a measure of how bright something looks. As such it includes the effects of observing conditions, locale, and the nature of the observer. In addition, stars come in a range of colors, depending on their temperatures. To make this measure useful to all terrestrial observers, it has to be less apparent and more objective. This is achieved by making stellar magnitude measurements using three filters, each covering a standard wavelength range, designated U, B, and V. The latter, the visual magnitude, covering the bandwidth 500–700 nm, replaces apparent magnitude.

References

1. Taub, L.C.: Hipparchus. In: John Lankford, J. (ed.). A History of Astronomy: An Encyclopedia. New York, Garland (1997)

2. Pederson, O.: *A Survey of the Almagest*. Springer. ISBN: 978-0-387-84825-9 (2011)
3. Pogson, N.R.: Magnitudes of 36 of the minor planets for the first day of each month of the year 1857. *MNRS* **17**, 12 (1856)
4. Johnson, H.L., Morgan, W.W.: Fundamental stellar photometry for standards of spectral type on the revised system of the Yerkes spectral atlas. *Astrophys. J.* **117**, 313 (1953)
5. Garrison, R.F., Karmo, T.: The brightest stars. In: Chapman, D. (ed.) *Observer's Handbook of the Royal Astronomical Society of Canada*, RASC Publications, Toronto, pp. 273–283 (2013)
6. Dufay, J.: *Introduction to Astrophysics*. The Stars Dover Press, Chapters 1, 2 and 3, Dover Press, New York, (1964)

Appearance

José Luis Caivano¹ and Paul Green-Armytage²

¹Secretaria de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

²School of Design and Art, Curtin University, Perth, Australia

Synonyms

[Color appearance](#); [Perceptual appearance](#); [Total appearance](#); [Visual appearance](#)

Definition

Visual appearance refers to how things look to the visual system. Objects have physical properties, and in this regard, optical properties are usually most relevant together with size, texture, shape, etc. But what finally matters for vision is the appearance that objects have for an observer (human observers will be mainly referred to in this article; however, this is not restrictive to human beings since there are many other living species that have similar visual processing systems). The American Society for Testing and Materials (ASTM) as well as the International Commission on Illumination (CIE) define appearance as the aspect of visual experience by which

objects are recognized [1, 2]. The ASTM specifies, furthermore, the definition of appearance from a psychophysical point of view: “perception in which the spectral and geometric aspects of a visual stimulus are integrated with its illuminating and viewing environment.” The physical properties have, of course, influence on the way things look, but also the neurophysiology and psychology of vision play a very important role in determining how objects and the world appear to humans. Thus it is important to understand visual appearance not merely in regard to physical aspects but as a sensation produced by the visual system when processing stimuli channeled by means of light – visible radiation. Visual appearance, understood as a visual sensation, i.e., what humans see, has certain attributes, which are also visual sensations. These include color, visual texture, gloss, and transparency. In the ASTM definition the “spectral aspects” refer, of course, to color, and the remaining aspects (including the perceived shape, size, etc. of objects) are usually considered as “geometric aspects.” From these visual sensations observers get information about the environment and build their knowledge of the external world, which allows them to interact with the environment and other individuals.

Overview

Historical Review of Studies on Visual Appearance

Under the heading of visual appearance a number of aspects are included such as color, texture, gloss, transparency, and also the shape and size of perceived objects or stimuli. The psychologist David Katz was probably the first to understand the various kinds of phenomena that always accompany the perception of color [3]. He describes different modes of color appearance: surface color, film color, volume color, specular colors, transparent colors, perception of luster, etc. By the middle of the twentieth century, Arthur Pope realized that, in order to define a color with accuracy, i.e., giving account for different modes in which a color may appear, more than the three usual variables (hue, lightness, saturation) are needed [4, p. 28].

In 1953, the Committee on Colorimetry of the Optical Society of America published a book in which visual perception is classified into eleven “attributes of modes of appearance” [5]:

1. Brightness (or Lightness)
2. Hue
3. Saturation
4. Size
5. Shape
6. Location
7. Flicker
8. Sparkle
9. Transparency
10. Glossiness
11. Luster

Lightness, hue, and saturation (attributes 1, 2, 3) are the usual variables by which color sensations are described. Size, shape, and location (attributes 4, 5, 6) are spatial categories of visual perception, which can be included in the category of spatial form. Flicker and sparkle (attributes 7, 8) involve temporal variations in the perception of light. Finally, transparency, glossiness, and luster (attributes 9, 10, 11) refer to the perception of different spatial distributions of light, aspects that will be included in the category of “cesia” to be described in the next section.

Richard Hunter proposed a classification of the geometric attributes of visual appearance [6]. He defined six different types of gloss and developed instruments for the measurement of some of these phenomena: goniophotometers, diffuse-reflection and specular-reflection meters, glossmeters, diffuse and specular-transmission meters, etc. Before Hunter, August H. Pfund had already described an instrument for measuring gloss [7].

The American Society for Testing and Materials has standardized the measurement of various physical aspects related to visual appearance [8–11] and published a glossary of standard terminology [1].

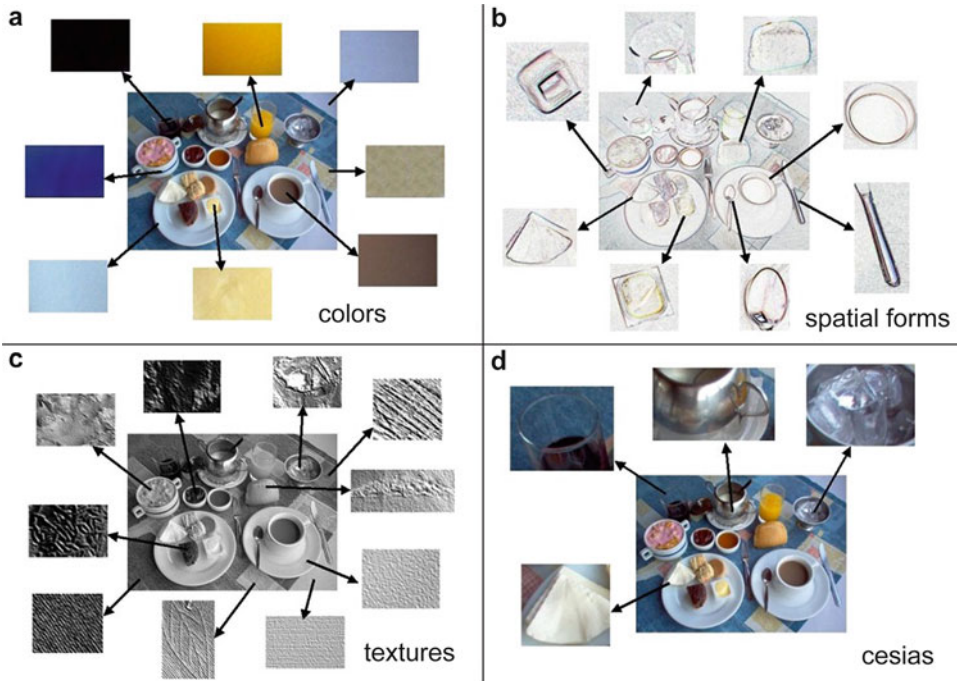
Robert Hunt has developed a hypothesis for the measurement of color appearance under different conditions of illumination and observation, taking color as the central point of his study [12]. Sven

Hesselgren observed that visual sensations, such as luster, reflection, and gloss, are not perceived as belonging to the color of an object but as something separate from color [13]. Ralph Evans, like Pope, also realized that the three variables normally used to define color are not enough to characterize color under different modes of appearance [14]. Evans concluded that it would be necessary to define at least five variables. In these conceptions, the phenomena of appearance have always been considered as aspects that accompany color sensations.

Other researchers have studied, either partially or globally, the aspects of visual appearance related to the different modes of distribution of light in space. Fabio Metelli tackled the problem of transparency from a phenomenological point of view, establishing a clear difference between physical and perceived transparency and developing a model that predicts the conditions under which the perception of achromatic transparency is possible [15]. Osvaldo Da Pos, following Metelli, enlarged the model to study the perception of chromatic transparency [16].

John Hutchings studied the phenomenon of translucency and its importance for the visual qualities of food. In articles published in 1993 and 1995 he integrated all the aspects in a model of total appearance [17, 18]. However, this model is mainly oriented to the analysis of food [19] and for this reason includes also nonvisual aspects of appearance such as smell, taste (flavor), and also texture as felt in the mouth, in addition to visual texture. Paul Green-Armytage proposed a three-dimensional model to organize what he called “qualities of surfaces” [20, 21]. Subsequently he introduced the term “tincture,” borrowed from heraldry, to encompass color, texture, and cesia [22]. Robert Sève tackled the specific problem of gloss [23]. Michael Brill used the model of Metelli to formulate a series of rules for the perception of chromatic translucency [24].

In 2006, a technical committee of the CIE (Commission Internationale de l’Eclairage, International Commission on Illumination) published a report to set a framework for the measurement of visual appearance [2].



Appearance, Fig. 1 The same scene with different visual features selected and categorized: (a) Different *colors*. (b) Different *spatial forms* or shapes that appear defined by their borders. (c) If attention is paid to certain regularities or repetition of elements on the objects, the notion of

texture can be applied. (d) Other visual characteristics such as the transparency of the glass, the white matte opacity of the paper towel, the metallic luster of the sugar pot, the glossy and semitransparent aspect of the ice cubes, and other similar ones have been given the name *cesia*

A Classification of Visual Appearance into Four Categories

It is possible to analyze a given scene and try to understand how most humans separate and categorize (by means of concepts and words) the main visual aspects involved. If the features shown in Fig. 1a are selected, people would agree on saying that they are different colors. If other features are selected, such as those shown in Fig. 1b, people will call them shapes (spatial forms). If the features selected in Fig. 1c are shown, people would call them textures. Finally, for the features selected in Fig. 1d the name *cesia* has been proposed.

Color can be defined as the aspect of visual appearance that results from the perception of different spectral compositions of light; it is what people are referring to when they say that some object or light stimulus is red, green, yellow, blue, etc.

Spatial forms, or shapes, are built in consciousness from the perception of borders and their integration, so that they enclose or delimitate a portion of space. Spatial forms can be one-dimensional (a straight line, a curved line, an irregular line, etc., which separate one portion of space from another), two-dimensional (a triangle, square, circle, irregular surfaces, etc.), or three-dimensional (a cube, a prism, a pyramid, a tetrahedron, a sphere, irregular volumes, etc.). They can have different shapes (as usually defined by the previous names), but they can also differ in size (length, surface, volume) and proportion (defined by the relationships among the dimensions of the shape: length, width, height).

Visual textures arise from the perception of sets of small elements that appear repeated in a surface or volume. The category of texture applies when these elements have no individual significance

and are interpreted as making a whole, building a pattern that can be more or less regular or irregular. Visual textures can be distinguished and described by different variables or dimensions. These variables include density (the elements of the texture can be sparse, dense, or intermediate), predominant directionality (the elements can vary in direction or have no predominant directionality), and, of course, the shape and size of the relatively small elements whose repetition composes the texture. Visually perceived textures may be two-dimensional (without relief) or three-dimensional (with relief), the latter being also perceived by touch.

Beyond color, spatial forms, and textures, there are some other visual appearances that depend on the way in which differences in the spatial distribution of light are perceived. It is these spatial distributions of light that cause things to be seen as transparent or opaque, glossy or matte, or with intermediate appearances between these. Since there has not been a single global name or category to refer to these aspects, the neologism *cesia* has been coined for them.

Cesia

“Cesia” is a term invented around 1980 by César Jannello, to designate the perceptions resulting from the different spatial distributions of light. This concept was developed mainly by Caivano [25–28]. Cesia encompasses all the visual sensations from transparency to translucency, matte opacity to mirror-like appearance (passing through glossy appearance), and lightness to darkness. Cesiums are related to color in the sense that both are different aspects of the perception of light, and color deals also with lightness and darkness. But there are aspects that cannot be described by the three classical color dimensions of hue, lightness, and saturation. In this sense, cesia constitutes a category that can be added to the categories of color, visual texture, and spatial form in order to extend the possibilities of describing and classifying visual appearance. Three dimensions or variables can define cesia with respect to the perception of how light interacts with objects: permeability (from opaque to transparent), diffusivity (from regular to diffuse),

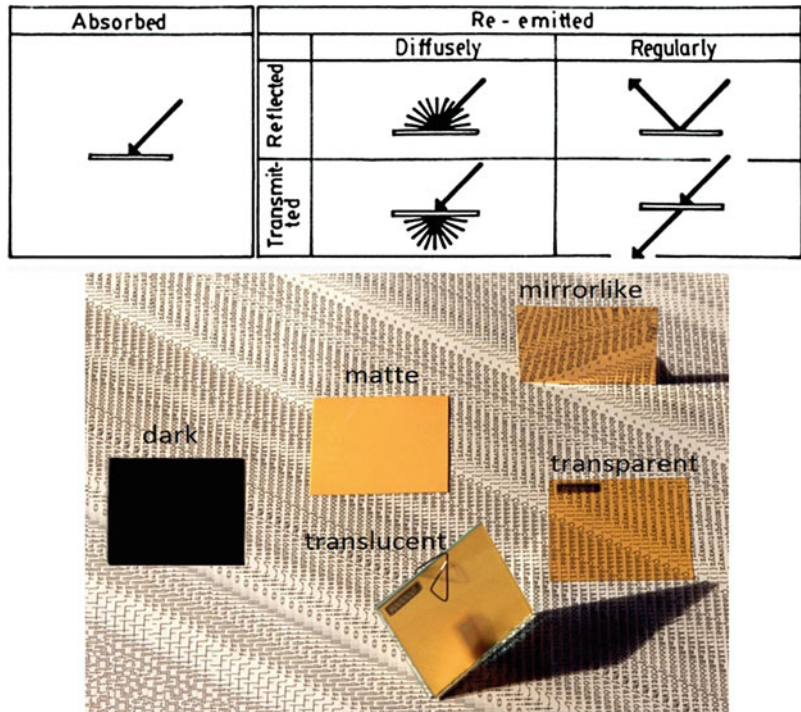
and darkness (from light to dark). Light can be reflected or transmitted by objects, either diffusely or regularly, and it is the perception of these phenomena that is covered by the term “cesia.” For instance, a surface that reflects light in a diffuse way will normally be perceived to be matte. If a surface reflects light with a certain specular component it will be perceived as glossy or mirror like. If it transmits light diffusely it will be seen as translucent, and if it transmits light in a regular way it will be seen as transparent. Finally a surface that absorbs most of the light will appear dark or black (Fig. 2).

A three-dimensional model has been proposed to organize the three variables of cesia (permeability, diffusivity, and darkness). The model is a kind of inverted pyramid, with four vertices on the top for the sensations of matte appearance, mirror-like appearance, translucency, and transparency. The connections among these form scales with a variation of permeability (or the opposite, opacity) in one direction and scales with a variation of diffusivity (or the opposite, regularity) in a perpendicular direction. From any of these points, scales going down in lightness (or increasing darkness) connect to the point that is the place for black (Fig. 3).

In both color and cesia the relationship between stimulus and sensation is not fixed but depends on three main factors – illumination, object, and observer –, and is affected by other factors such as visual context, adaptation, contrast, etc.

Color and cesia are closely connected because of their relationship with light; both are different aspects of the perception of light that contribute to the visual appearance of objects. The dimension of lightness/darkness is shared by color and cesia, being the link that connects both phenomena. Color and cesia interact thereby expanding the countless number of different visual appearances that humans are able to perceive. When Ludwig Wittgenstein, in his *Remarks on Color* [29], is concerned with the different types of white, yellow and golden, gray and silver, “black” mirrors, etc. and when he says, “Opaqueness is not a property of the white color. Any more than

Appearance,
Fig. 2 Cesia: processes followed by light falling on objects, and the way these stimuli are usually perceived



transparency is a property of the green,” he is dealing, in reality, with this aspect that can be called cesia.

Models to Organize Different Aspects of Visual Appearance

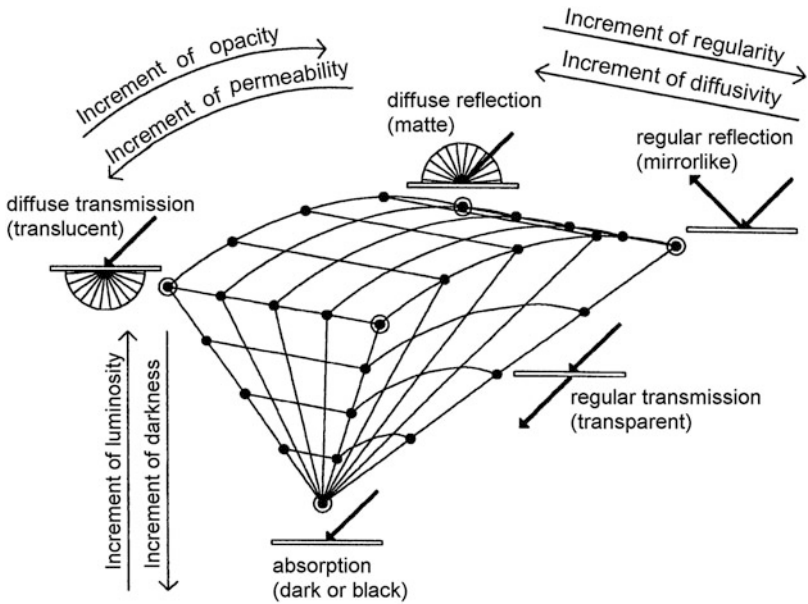
Roberto Daniel Lozano, one of the members of the CIE technical committee 1-65, proposed a new approach to appearance characterization, with a model that classifies visual appearance according to three main categories (color, cesia, and spatiality), which in turn includes subcategories and shows the interrelations and intermediate aspects among them [30].

More recently, Paul Green-Armytage proposed a model to link different modes and different aspects of appearance [31]. The model shows relationships between the surface, volume, and illuminant modes of appearance as well as transparency, translucency, opacity, texture, gloss, metallic luster, and luminosity. Caivano identifies the five vertices of his inverted pyramid model of cesia (Fig. 3) as “primary sensations” [26]. These are opaque white, translucent white, transparent

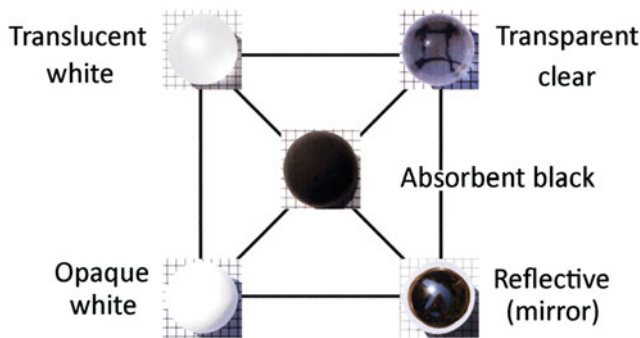
clear, mirror-like, and black. Caivano’s three-dimensional inverted pyramid can be presented in two dimensions (Fig. 4).

This two-dimensional representation of Caivano’s model is at the center of Green-Armytage’s model. Green-Armytage makes a distinction between gloss and metallic luster (mirror-like), noting that scales can be constructed from a matte surface to surfaces that are increasingly glossy in one direction and from a matte surface to surfaces with increasing metallic luster in another direction. He also notes that scales can be constructed with increasing texture from three different starting points: smooth matte, smooth gloss, and smooth metallic luster. Matte, glossy, textured, and metallic aspects of appearance are all characteristic of the surface mode of appearance. The volume mode can vary between clear (transparent) and cloudy (translucent). In the illuminant mode light sources can appear in varying degrees of brightness with a notional maximum that can be described as glare. To Caivano’s five primary sensations Green-Armytage adds three more: gloss white, textured white, and white

Appearance, Fig. 3 The three-dimensional model of cesias



Appearance, Fig. 4 The classification of visual appearance proposed by Lozano (reproduced with permission for R. D. Lozano)



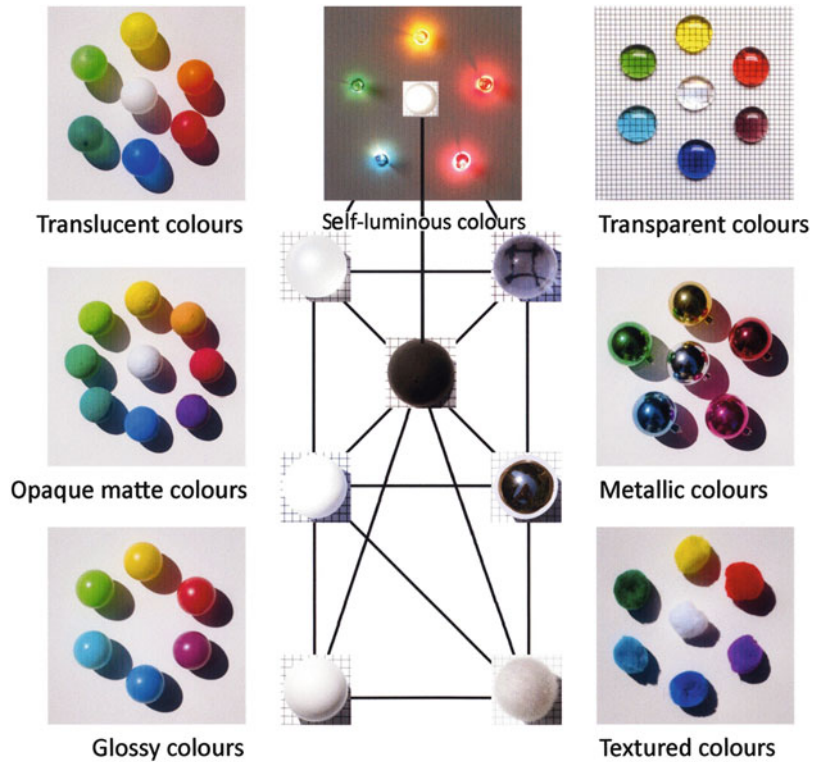
light glare. Each of the seven other primary sensations can be connected with scales to black, and these scales can, in turn, serve as the achromatic axes of order systems for colors that are opaque, translucent, transparent, and metallic (mirror-like), as well as glossy, textured, and self-luminous. Green-Armytage’s model, with color circles to correspond with the different primary sensations, is shown in Fig. 5.

Increasing apparent intensities of illumination can be represented on lines which converge at white light glare. The coherence of a light source, which affects appearances, can also be represented

on the diagram. “Hard” light, as from direct sun-light and under which objects cast clearly defined shadows, is represented on the line connecting transparent clear to white light glare, and ‘soft’ light, as from a heavily overcast sky and under which no shadows are cast, is represented on the line connecting translucent white to white light glare.

The various models described here can go some way to increasing awareness of the different aspects of appearance, and how they can be related, but no model can encompass the full experience of seeing the world in space and time.

Appearance, Fig. 5 Two-dimensional representation of Caivano's inverted pyramid model of cesia



Cross-References

- [Achromatic Color](#)
- [Anchoring Theory of Lightness](#)
- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [Diffuse Lighting](#)
- [Gloss Meter](#)
- [Glossmeter](#)
- [Goniophotometer](#)
- [Opacity – Extinction](#)
- [Texture Measurement, Modeling, and Computer Graphics](#)
- [Tinture](#)
- [Transparency](#)
- [Unique Hues](#)

References

1. ASTM, American Society for Testing and Materials: Standard Terminology of Appearance. Standard E 284-01. ASTM, West Conshohocken (2001)
2. CIE, Commission Internationale de l'Eclairage, Technical Committee 1-65: CIE Technical Report: A Framework for the Measurement of Visual Appearance. CIE Central Bureau, Vienna (2006)
3. Katz, D.: Der Aufbau Der Farbwelt, 2nd ed. of Die Erscheinungsweisen Der Farben Und Ihre Beeinflussung Durch Die Individuelle Erfahrung. Johann Ambrosius Barth, Leipzig (1911/1930). English translation by MacLeod, R.B., Fox, C.W.: The World of Colour. Keagan Paul, Trench, Trubner, London (1935)
4. Pope, A.: The Language of Drawing and Painting. Harvard University Press, Cambridge, MA (1949)
5. Optical Society of America, Committee on Colorimetry: The Science of Color. Crowell, New York (1953). chapter 2
6. Hunter, R.S.: The Measurement of Appearance. Wiley, New York (1975)
7. Pfund, A.H.: The measurement of gloss. J. Opt. Soc. Am. **20**, 23–26 (1930)
8. ASTM, American Society for Testing Materials: Standard Recommended Practice for Selection of Geometric Conditions for Measurement of Reflectance and Transmittance. Standard E 179-73 (1974)
9. ASTM, American Society for Testing and Materials: Standard Guide for Selection of Geometric Conditions for Measurement of Reflection and Transmission Properties of Materials. Standard E 179 (1990)

10. ASTM, American Society for Testing and Materials: Standard Test Method for Visual Evaluation of Gloss Differences Between Surfaces of Similar Appearance. Standard D 4449 (1990)
11. ASTM, American Society for Testing and Materials: ASTM Standards on Color and Appearance Measurement, 5th edn. ASTM, Philadelphia (1996)
12. Hunt, R.W.G.: Measurement of color appearance. *J. Opt. Soc. Am.* **55**(11), 1540–1551 (1965)
13. Hesselgren, S.: *The Language of Architecture*. Studentlitteratur, Lund (1967)
14. Evans, R.M.: *The Perception of Color*. Wiley, Nueva York (1974)
15. Metelli, F.: The perception of transparency. *Sci. Am.* **230**(4), 90–98 (1974)
16. Da Pos, O.: *Trasparenze*. Icone Editrice, Padova (1990)
17. Hutchings, J.B.: The concept and philosophy of total appearance. In: Nemcsics, A., Schanda, J. (eds.) *Colour 93, Proceedings of the 7th Congress of the International Colour Association*, vol. C, pp. 55–59. Hungarian National Colour Committee, Budapest (1993)
18. Hutchings, J.B.: The continuity of colour, design, art, and science (I and II). *Color. Res. Appl.* **20**(5), 296–312 (1995)
19. Hutchings, J.B.: *Food Color and Appearance*, 2nd edn. Aspen Publishers, Gaithersburg (1999)
20. Green-Armytage, P.: Colour and other aspects of appearance. *Spect. Newsl. Colour Soc, Aust.* **6**(3), 1–11 (1992)
21. Green-Armytage, P.: Beyond colour. In: Nemcsics, A., Schanda, J. (eds.) *Colour 93, Proceedings of the 7th Congress of the International Colour Association*, vol. A, pp. 155–162. Hungarian National Colour Committee, Budapest (1993)
22. Green-Armytage, P.: Tincture – a new/old word for the appearance of things. *J. Sch. Des.* **2**, 16–23 (1993)
23. Sève, R.: Problems connected with the concept of gloss. *Color. Res. Appl.* **18**(4), 241–252 (1993)
24. Brill, M.H.: The perception of a colored translucent sheet on a background. *Color. Res. Appl.* **19**(1), 34–36 (1994)
25. Caivano, J.: Cesia: a system of visual signs complementing color. *Color. Res. Appl.* **16**, 258–268 (1991)
26. Caivano, J.: Appearance (cesia): construction of scales by means of spinning disks. *Color. Res. Appl.* **19**, 351–362 (1994)
27. Caivano, J.: Cesia: its relation to color in terms of the trichromatic theory. *Die. Farbe.* **42**(1/3), 51–63 (1996)
28. Caivano, J.: An atlas of cesia with physical samples. In: *AIC Color 97, Proceedings of the 8th Congress*, vol. I, pp. 499–502. Color Science Association of Japan, Kyoto (1997)
29. Wittgenstein, L.: *Bemerkungen über die Farben (1950–1951)*. English translation by McAlister, L., Schättle, M.: *Remarks on Colour*. University of California Press, Berkeley (1977)
30. Lozano, R.D.: A new approach to appearance characterization. *Color. Res. Appl.* **31**, 164–167 (2006)
31. Green-Armytage, P.: A model to link different modes and different aspects of appearance. In: *AIC 2011 Interaction of colour and light, Proceedings of the Midterm Meeting of the International Color Association*, pp. 165–168. Pro/colore, Zurich (2011)

Arc Lamps

► Carbon Arc Lamp

Architectural Lighting

Peter Raynham¹ and Kevin Mansfield²

¹University College of London, London, UK

²UCL Institute for Environmental Design and Engineering, London, UK

Synonyms

[Lighting design](#)

Definition

Architectural lighting design is the design and implementation of daylight and electric lighting provision in urban and architectural environments. This process requires conceptual thinking, design development, and illuminating engineering to produce lit environments for people's work and leisure.

Lighting Quality

The built environment in which people live and work is composed of a variety of building types. In a shopping mall retail units are expected to be colorful and attractive. In an art gallery it is necessary to be able to appreciate the color and detail in a painting, and when walking in the city people expect to remain safe and secure. All these

impressions can be dramatically affected by the lighting. Boyce has suggested that *bad* lighting “does not allow you to see what you need to see, quickly and easily and/or causes you discomfort” [1]. If the lighting designer overcomes these problems then they can produce at best *indifferent* lighting. Only when the configuration of the windows and surfaces or the characteristics of the lighting installation “lift the spirits” can the environment be described as *good* lighting.

The task of the lighting designer or a lighting engineer is to determine the character of a lit space whether it be a street, a townscape, or a room and then to arrange the sources of light, be they daylight or electric light, to achieve this desired character.

Objectives of Design

The purpose of lighting is to provide for the needs of people who are going to use the built environment that is being created, and any person designing the lighting will wish to create a quality lit environment. However this raises the question about what constitutes good-quality lighting. To answer that question the needs of the people who are going to use the building that is to be lit require consideration. In general terms the users of a building are going to want lighting so that they can see things easily, they will require the lighting to set the correct ambience, and if they are going to remain in the building for long periods they may well benefit from the biological impact that the lighting may have on their endocrine systems. These topics are covered in more detail in the following sections.

Light for Seeing

Before considering what sort of lighting is required it is first essential to consider what the occupants of a building need to see and what the consequences are of them not being able to see. Consider two cases: a surgeon undertaking heart surgery and an office junior doing a filing task. Both of them need to be able to see small objects;

in the case of the surgeon this may be small blood vessels whereas for the office junior he has to read the small text on the documents he is working with. However, the office junior is normally dealing with high-contrast printed text, while the surgeon has the problem of distinguishing objects that are very similar in color and reflectance. Moreover, if the surgeon makes a mistake it could kill a patient, but an error in filing can be easily rectified. Thus it is normal to have more than 10 times as much light available in an operating theatre compared to an office.

Selecting an appropriate level of lighting for a given visual task is bound to be complex as a number of factors need to be considered. These include the complexity of the visual task, the visual acuity of the person performing the task, the length of time for which the task is performed, and the consequences of any mistakes. These factors need to be set against the cost of providing the lighting and the other negative consequences of providing lighting such as energy use and associated greenhouse gas emissions.

To help lighting designers in the selection of appropriate lighting conditions there are a number of Codes and Standards available which provide recommendations for the lighting conditions considered necessary for a whole range of different tasks. These documents have evolved over a number of years and in general provide a consensus view of the lighting needs in a typical installation. However, the lighting requirements for every installation are potentially different, and so the guidance provided by Codes and Standards should usually be regarded as a starting point from which the design may be developed; however, there are occasionally situations where the national regulations insist that particular illuminance values are used in given circumstances. Good examples of such codes are the *IES Handbook* [2], the *SLL Code for Lighting* [3], and *ISO 8995-1:2002* [4].

As well as considering the required illuminance necessary for a particular task it is also important to consider other factors such as freedom from glare and the necessary color properties of the light sources used. Most Codes and Standards also give advice on these issues.

Setting the Ambience

Every lighting installation sends a message to the users of the building; it is the art of good lighting design to ensure that this message complements the message sent by the architecture and that the overall message is appropriate to the use of the building. Consider the lighting in a supermarket and compare it with the lighting in an upmarket boutique. In many ways the lighting has a similar job to do: it must show off the goods in the store so that people choose to buy them. However, the style of the lighting is radically different as supermarket-style lighting in the boutique would send entirely the wrong message to potential customers.

There is no simple way of providing lighting that perfectly provides the correct message for a given environment. There have been many studies into how to characterize the lit environment of places that are considered to be well lit, and in general the results have been very limited in validity to the environment in which they were assessed. The section below on design considerations gives some basic advice, but it is always going to involve a series of complex subjective judgments to get the lighting right, and the reliance on simple codes will not suffice.

Biological Impact

The nonvisual or biological effects of light have their biggest impact on the regulation of the human sleep/wake cycle. Exposure to high levels of light may be used to shift the time of peoples' body clocks and may also be used to boost the alertness of people. These biological effects happen naturally to people in an outdoor environment; however, in many interiors there is not sufficient light to provide the necessary stimulus to the human endocrine system.

The most logical way to resolve the problem is to provide interiors with good levels of daylight. However, where that is not possible, high levels of electric light with increased amounts of blue light content may be used.

A Framework for Design

The objective for anyone trying to design lighting is to provide good-quality lighting. The construction of a new building or the refurbishment of an existing one comprises a series of construction stages from design through construction to the final use of the building or area. Lighting should be designed by someone who has the ability to make the necessary aesthetic judgments about how the lighting will impact on the space being lit while at the same time engineering the lighting equipment into a functioning system. Therefore the lighting designer or lighting engineer can make a particular contribution or intervention within the *design* phase. Here a lighting design concept is proposed where outline proposals for the arrangement of structural and building services is made. The design is then developed in detail to decide on the relationship between the electric lighting systems and the daylighting, and finally a technical specification is drawn up to coordinate the various components and architectural elements.

At each of these stages the designer will be considering a variety of factors. Veitch has suggested that designers should consider the well-being of the individual user, the integration with the architecture, and the economics of the lighting installation [5]. Loe has proposed a useful variation in a framework for design [6] which comprises the following elements (Fig. 1).

Lighting for Visual Function

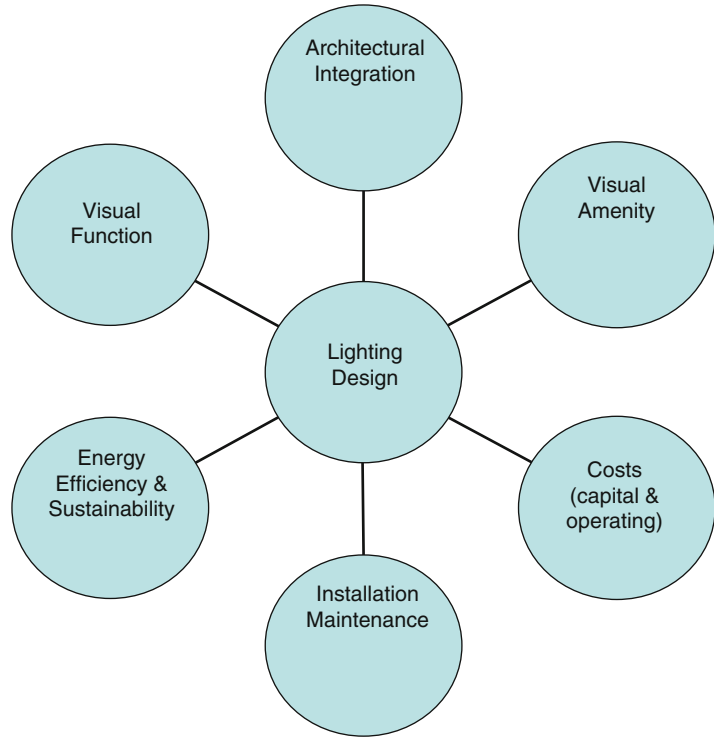
The lighting should enable users to undertake visual tasks without visual discomfort. Recommendations are made concerning the level of light on the task (illuminance) and the uniformity of light on the task. It is also important to prevent unwanted brightness in the visual field (the control of glare).

Lighting for Visual Amenity

People prefer an architectural room or urban space to contain a degree of *visual lightness* (a sense of lightness and airiness) and *visual interest* (a variation in light and shade). The color appearance of the space should be appropriate for the

Architectural Lighting,

Fig. 1 A framework for lighting design (after Loe). The relative weight given to each element of the framework depends upon the requirements of the lighting design



application, e.g., warm and inviting or cool and professional. The objects and people in the space need to be lit pleasingly.

Energy Efficiency and Sustainability in Lighting

The lighting installation must contribute to sustainability by maximizing the use of daylighting and ensuring that energy-efficient light sources and light fittings are specified. Reducing energy consumption is a matter of reducing the power load of the lighting components and the hours of use. This can be achieved by the effective use of lighting control systems (switching, dimming, and linking to daylight levels).

Lighting Installation Maintenance

The lighting designer or engineer must specify a lighting installation that can be easily maintained by the client or user. It should be easy to clean and relamp the fittings to ensure that the installation continues to deliver the appropriate lighting. Poor maintenance of an installation will

lead to some of the lighting points failing, and this sends the message that no one cares about the space.

Lighting Costs

The cost of the installation needs to be controlled, and a distinction is made between *capital* costs (for the original lighting equipment) and *running* costs (the cost of operating the lighting system as designed).

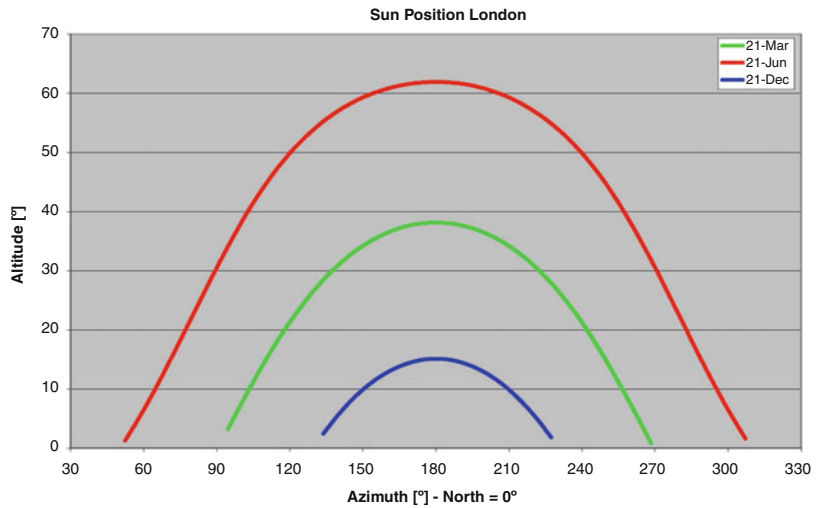
Architectural Integration

An understanding of the intent of the architect and the character of the space in terms of dimensions, finishes, textures, and materials is essential. Only then can electric lighting systems be properly integrated into the fabric of the building to produce the appropriate lighting effect.

These factors are described in greater detail in the *SLL Lighting Handbook* [7]. It should be noted that there are occasionally legal requirements for lighting in certain areas.

Architectural Lighting,

Fig. 2 A plot of the sun's position in the sky for London (51°N Latitude) denoted by its azimuth angle in plan and altitude above the horizon



A

Daylighting

It is useful to think of the elements of the city and the town and the constituent buildings as being under the vast dome of the sky. In many climates there will be days in which this dome is covered by a thick layer of cloud: the fully overcast sky, which might be termed the *diffuse* component. As the conditions change, the clouds begin to roll away and the sun comes out. This source has a very different character: a sharp beam and the ability to create shadows: this may be called the *direct* component. These two components of daylighting are varying continually throughout the day and throughout the year.

The lighting designer works at different scales of analysis. For example, the designer might seek to optimize the spacing between apartment blocks to ensure that certain apartments are not restricted in the amount of light that they receive. The designer might also seek to ensure that courtyards or plazas receive enough sunlight or are shaded in certain climates. Stereographic sun paths are available to show the passage of the sun in the sky (Fig. 2).

The lighting designer can also work in conjunction with the architect to ensure that the individual rooms and spaces within a building receive sufficient daylighting. The traditional way to pierce the solid wall of a building is by a vertical window. The clerestory window and the rooflight provide ways of allowing daylight to enter the

building to provide a different character to the space. This is because they are facing different parts of the sky. The atrium is a useful device, historically and in contemporary buildings, to let natural light deep into the building. It is also possible to “borrow” light from one space to another using transparent partitions.

Contemporary buildings have a framed steel or concrete structure in which the structure defines the architecture. The introduction of daylight into the building thus becomes a result of the structural decisions that have been made. This is the province of facade engineering whereby the transparent partitions that clothe the structure are then subject to varying architectural treatment in terms of the choice and specification of the glazing (solar control or electrochromic materials) and whether shading is to be added (louvers and *brise-soleil*). It is at this point that daylight availability is estimated by calculation of the daylight factor or methods of calculation of cumulative daylight exposure.

When designing the daylight in a building it is also important to consider the quality of the view out that the windows may give.

Electric Lighting

There are a plethora of new technologies for lighting, e.g., light-emitting diodes (LEDs), and a

variety of lamp technologies including high-pressure discharge lamps, fluorescent lamps, and incandescent lamps. The designer has the ability to control both the intensity and the color of the light that is emitted from these sources.

Consider the public square in the town or city. An appropriate choice of lamp on an amenity lighting column can attractively reveal the features on the faces of fellow street users to ensure safety and security, and the appropriate color choice can dynamically reveal the city.

The architectural façade can take a variety of forms: neoclassical, turn of the century, international modern, or contemporary. It is the task of the lighting designer to reveal the characteristics of the architecture without distortion, to complement and enhance the architectural detail. It is important to ensure also that each architectural element is read as a whole: from the base to the roof. Again the judicious choice of color and intensity of the light source in conjunction with an appropriate lighting fitting can ensure that the architectural intention is maintained.

Moving indoors the choice of light source remains critical in terms of its color, its life, and its efficiency. Such sources are housed within light fittings (*luminaires*) – used individually or more typically in arrays to provide layers of lighting.

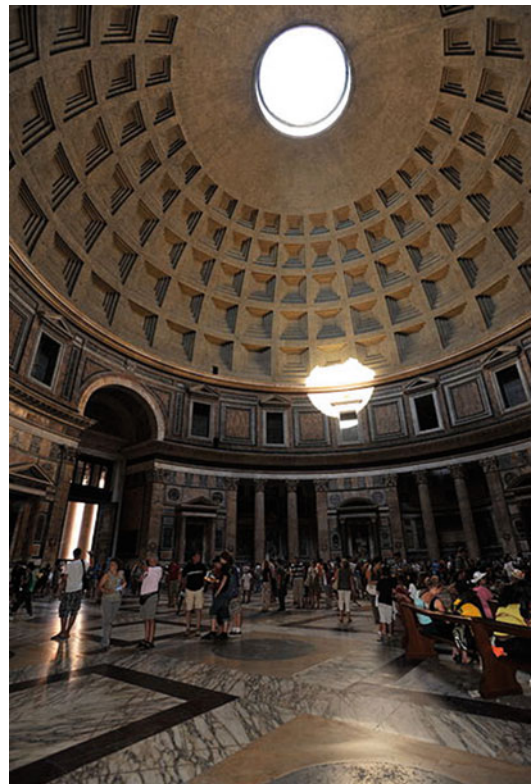
An individual luminaire can be used to accent individual features, and arrays of luminaires (linear and modular in form) can be arranged within the various architectural spaces to provide lighting for tasks and to reveal the architecture of the space with its characteristic finishes and materials. Care is taken by the designer to ensure that the user of the space is not subject to unwanted glare from the fittings and that the overall brightness of a sequence of spaces from one to the next is kept within reasonable bounds (*adaptation*).

The design of an individual luminaire is a complex process in itself. It must be designed to house a particular light source which might be spherical, linear, or elliptical. The flux from the lamp must be directed into the required direction by the use of a reflector, and then the beam from the lamp must be further refined by optical elements such as diffusers, lenses, or louvers. The fitting must be electrically safe, must not overheat,

and be mechanically strong: these are common requirements even though the physical form of the luminaire might be quite different, *cf.* a downlighter in a conference center to a street-lighting bollard. The designer must also consider the appearance of the fitting when it is lit and when it is unlit and ensure that the fittings integrate well with the architectural character of the space.

Endpiece

Architectural lighting has great power. Think of the dramatic daylighting of the Pantheon in Rome (Fig. 3) or the harmonious integration of the lighting of the city of Lyon at night (Fig. 4) or the dynamic lighting of a modern luminous media façade (Fig. 5). But with that power comes



Architectural Lighting, Fig. 3 The Pantheon in Rome
http://commons.wikimedia.org/wiki/File:Pantheon_Rome.jpg (CC BY 2.0)

Architectural Lighting,

Fig. 4 The city of Lyon at night (Fabian Fischer (CC BY 2.0))



A

Architectural Lighting,

Fig. 5 A luminous media façade <https://www.flickr.com/photos/ell-r-brown/6258523229> (CC BY 2.0)



responsibility that can be accepted by the designer in designing lighting that promotes human comfort and satisfaction.

Cross-References

- [Light-Emitting Diode, LED](#)
- [Lighting Controls](#)

References

1. Boyce, P.R.: Lighting quality: the unanswered questions. Paper presented at the first CIE symposium on lighting quality, Ontario, 9–10 May 1998
2. The Lighting Handbook: Reference and Application, Illuminating Engineering Society of North America, New York (2011)
3. SLL Code for Lighting, Society of Light and Lighting, London (2012)

4. ISO 8995-1:2002 (CIE S008/E:2001) Lighting of Work Places Part 1: Indoor
5. Veitch, J.A.: Commentary: on unanswered questions. Paper presented at the first CIE symposium on lighting quality, Ontario, 9–10 May 1998
6. Loe, D.L.: Lighting quality: an exploration. Paper presented at the first CIE symposium on lighting quality, Ontario, 9–10 May 1998
7. SLL Lighting Handbook, Society of Light and Lighting, London (2009)

Art and Fashion Color Design

Larissa Noury

Colour-Space-Culture, Paris, France

Synonyms

Color harmonies for fashion design; Modern art and fashion design; Painted dress; Tactile painting and haute couture

Definition

Art and fashion color design concerns the harmonies of color associations of textile and other materials used for a couturier's aesthetic project. It is an illustration of collaboration of artists and couturiers on the harmonies of color associations in order to create a personal style. They were more closely tied at the turn of the twentieth century than they are today. Artists did not see the difference between creating an original work of art, such as a painting, and designing a textile pattern that would be reproduced many times over. Each was a valid creative act in their eyes. There are a lot of vivid illustrations of the centuries-long love affair between fashion and art of color. Couturiers are past masters at capturing the contemporary *zeitgeist* in their designs, while artists have frequently used clothing as a way to give all-round expression to their aesthetic ideas.

The Role of Art and Fashion

Beginning from the nineteenth century, major changes emerged both in the role of fashion and

in the place of art in society. Growing affluence and new social structures gradually turned art, color, and fashion into ways of expressing personal taste and identity. This section presents a historic panorama of the color in the fashion design during the twentieth century, highlighting certain symbolic movements such as the Art Nouveau, the Russian avant-garde, the modernism, pop art, or the kinetic art. It lists the color's harmonies for modern fashion design and describes the tools, ranges, ► [palettes](#), and techniques which allow personalizing a dress with fantasy and subtlety.

Art has often been a major source of inspiration for dress designers of the twentieth century. Some creations by Liberty, Sonia Delaunay, Liza Schiaparelli, Coco Chanel, Yves Saint Laurent, Givenchy, and Jean-Paul Gauthier can be mentioned. In their practice, the art of color is conveyed to fashion. The designers paraphrase the masters of painting. This connection between art and fashion is particularly obvious starting in the late nineteenth century.

Fashion and Identity of the Color Palette

Belle Epoque

The artistic principles of the Art Nouveau combine the personal style of expression where the sinuosity of the shape harmonized well with the mosaic of contrasted colors. The synthesis of practical considerations and of aesthetic natural forms and lines is characterized by color palettes of a great richness. The modern time influenced a lot the development of the fashion design and produced profusion of the ► [polychromy](#) in it. The large selection of paintings by Giovanni Boldini, Auguste Toulmouche, James Tissot, Jean Beraud, and Alfred Stevens – visual artists who reflected the best on the fashion style of the Belle Epoque – is an example of what could be called fashion inspiration in fine arts. Artworks were created in realistic manner and have not even a touch of stylization typical of modernist art. Figures depicted in these paintings are precise and realistic. However, it would be fair to say that style itself is the subject matter instead of a

particular model. Fashion is the main theme and inspiration for these paintings.

Beautiful symbolic women's dresses by famous Austrian artist Gustav Klimt followed, in which he introduced his frescoes and paintings of portrait. Klimt himself drew the blue caftan, a long tunic of oriental style, and many other dresses for Flöge Emilie, his companion for many years, who owned a fashion house in Vienna.

After a visit of a workshop in Vienna, the avant-garde fashion designer Paul Poiret brought the idea of mixing art and fashion to Paris. He opened a Martine school in 1911, a place that was also attractive for artists. He then employed Parisian artists such as Lepape, Ibibe, and Erte on fashion illustrations. He employed the artist Raoul Duffy to design fabric prints and to invent tissues. He went to art galleries and showed his artistic sensibilities by preferring Impressionist paintings at a time when they were new and unappreciated by the public. Poiret became very interested in modern art and said, "I have always liked painters. It seems to me that we are in the same trade and that they are my colleagues". The couturier considered himself as an artist first.

Russian Constructivism and Suprematism

The artists of Russian avant-garde in 1915–1935 initiated probably the most intensive and creative art and architectural movement of the twentieth century and became a significant source of any art movement since that. The constructivists defined the chromatic surfaces as fundamental colored elements where the straight lines, the rectangular forms, and the principal colors (yellow, red, blue, black, and white) are used to make a unified composition. The color interdependence simultaneously with the community of ideas formed by the uniformity of geometric contours and the absence of force of gravity, the balance of the unit, and the dynamism of the parts, all form a kind of representation of an ideal cosmos.

Varvara Stepanova, Alexandre Rodchenko, Liubov Popova, and others show their creativity thorough re-energizing new forms and meaning in art and dress design. The key fragments of Russian revolutionary creativity still glow like

radium, living on futuristic art and design into the imaginations of some most influential couturiers of the twentieth and twenty-first centuries. Christian Dior *haute couture* in 2002 was inspired by the color ► [palette](#) of Malevich painting.

Simultaneous Contrast and Sonia Delaunay

In the 1920s, abstract painting inspired a variety of fabric designs, mainly based on the simultaneous contrast effect, by the successful designer and artist Sonia Delaunay [1]. Married to Robert Delaunay and friend of artists like Mondrian, Arp, Vantongerloo, and Kandinsky, she was a member of the contemporary artistic avant-garde in her own right. It was her own abstract paintings that she translated into rhythmic designs composed of squares, lines, circles, diagonals, and color planes. In all, Delaunay created over 2,000 of these fabric designs, around 200 of them produced especially for the fashion house Metz & Co in Amsterdam.

Surrealism and Elsa Schiaparelli

Fashion designer Elsa Schiaparelli, Coco Chanel's main rival in the 1920s and 1930s, produced clothing and hats heavily influenced by Surrealism [2]. Her sweaters incorporating knitted ties or sailor collars were a sensation, and she worked in close cooperation with artists like Salvador Dalí and Jean Cocteau. An example of her work with Dalí is her famous lobster dress.

Coco Chanel and Japanese art

The designer's passionate interests inspired Coco Chanel's fashions [3]. Her apartment and her clothing followed her favorite color ► [palette](#), shades of beige, black, and white. Elements from her art collection and theatrical interests likewise provided themes for her collections. The ornament of the dress, in both pattern and color palette, resembles the Asian lacquered screens which the designer loved and collected. In convergence with the Art Deco line, the modernist impulse was married with pure form and Japanese potential.

Neoplasticism and the Bauhaus

Piet Mondrian changed the face of modern art. His influence extends to painting, sculpture, graphic design, and fashion. In search of plastic harmony,

he introduced a universal language of shapes and ► [primary colors](#) that goes beyond the painting; Mondrian was the central and most famous figure of the De Stijl movement. This style was baptized as neoplasticism and intended to achieve real objectivity by releasing the work of art from its dependence on the momentary individual perception and temperament of the artist. Yves Saint Laurent has created his famous dress with Mondrian's color composition. The color in fashion design was developed by the school of Bauhaus and was also very enriching and interesting [4]. The relationship between form and color within the framework of visual perception was defined starting from the color theories of Kandinsky, Klee, Itten, and Albers in this school.

Op Art, Minimalism, and Pop-Art in Fashion

Since the Second World War, there has also been frequent interaction between art and fashion. In the 1950s, Karel Appel produced signed fabric designs, in the 1960s the Bijenkorf department store sold dresses inspired by op art, and since then the emergence of minimalist art has given rise to a widespread taste for sober, often asymmetrical designs. For instance, Jean-Charles de Castelbajac was inspired by Andy Warhol and his pop art *Campbell's Soup* painting. Victor Vasarely, Yvaral, and other representatives of optical art, as well as Friedensreich Hundertwasser's art, continue to influence modern fashion design and modern life [5].

In 2007, Christian Dior designed a unique piece, hand painted and enhanced with spectacular embroidery, Manteau Suzurka-San. Dior *haute couture* was inspired by *The Great Wave of Kanagawa*, emblematic work of the Japanese artist Hokusai. Focusing on the relationship between art and creations of the house Dior, it is possible to say that the original works have been in one way or another influenced by different artists. Certain highlighted artworks influenced not only the collections *haute couture* and *pret-à-porter* but also the unique world of fine jewelry, perfumes, and accessories. Because they reflect light and velocities, colors and shapes of the paintings of Sonia Delaunay emphasized the dancing model dressed in her parade by Ungaro Fall-Winter 2003–2004. Yves Saint Laurent revolutionized fashion and

gave woman the freedom of movement that has inspired artists, poets, and painters. “The profession needs an artist to exist,” he said. He loved painting and painters such as Matisse, Mondrian, Braque, Picasso, and Van Gogh. Saint Laurent began in 1988 with Georges Braque, whose famous birds seem to fly stuck to the bride's dress. Then, he designed a jacket inspired by *Iris* of Van Gogh. It took 800 h to sew the whole Van Gogh. Flakes, tubes, seed beads, and ribbons were all embroidered by hand to make the effects and lighting of the canvas. Yves Saint Laurent influenced in some way other couturiers. For instance, stylist Erin Fee has been inspired by flowered textile design and some kind of “camouflage” for his fashion collection. Jean-Charles de Castelbajac, on the other hand, brings a joyful color palette from BD art graphics. Nowadays, fashion designs have increasingly been regarded as autonomous works of art. Some creations by designers like Viktor & Rolf now go so far in that direction as to be scarcely wearable at all. Viktor & Rolf spread their cut-up couture in the magazine *Dazed and Confused*.

Tactile Color and Haute Couture

Nowadays, the collaboration between artists and fashion designers continues. The personal artist's inspiration comes from tactile painting. Jean Marie Pujol, couturier who worked with Dior and Yves Saint Laurent, designed several dresses to be painted by Larissa Noury as a means to perpetuate the art-and-fashion marriage [6]. With this personal style, a series of hand-painted dresses were created and presented during the exhibition at Faubourg's gallery in Paris, amid boutiques of Christian Dior, Jean Paul Gauthier, Chanel, and other creators of fashion.

Cross-References

- [Color Contrast](#)
- [Color Harmony](#)
- [Palette](#)
- [Polychromy](#)
- [Primary Colors](#)

References

1. de Leeuw-de Monti, M., Timmer, P.: *Colour Moves: Art and Fashion by Sonia Delaunay*. Thames & Hudson, London (2011)
2. Martin, R.: *Fashion & Surrealism*, (reprint edn.). Rizzoli International Publications, New York (1996)
3. Degunst, S.: *Coco Chanel: Citations*. Éditions du Huitième Jour, Paris (2008)
4. Fiedler, J., Feierabend, P.: *Bauhaus*. Editions Konemann Verlag, Cologne (2000)
5. Popper, F.: *From Technological to Virtual Art*. Leonardo Books. MIT Press, Cambridge (2007)
6. Pujol, J.-M.: *Haute Couture*. Éditions du Palais, Paris (2011)

Artistic Rendering

► Non-Photorealistic Rendering

Assimilation

Simone Gori
Department of Human and Social Sciences,
University of Bergamo, Bergamo, Italy
Developmental Neuropsychology Unit, Scientific
Institute “E. Medea”, Bosisio Parini, Lecco, Italy

Synonyms

Bezold–Brücke effect; von Bezold spreading effect

Definition

Color assimilation is somewhat considered the opposite of the color contrast: the color is perceived in the direction of the hue of the surrounding color, whereas in color contrast the perceived color tends to be the complementary color. Assimilation can be experienced also in grayscale stimuli and it is called lightness assimilation or brightness assimilation. In this case a gray is perceived lighter if it is close to a light object and

darker if it is close to a darker object which is, again, the opposite of the simultaneous contrast.

Introduction

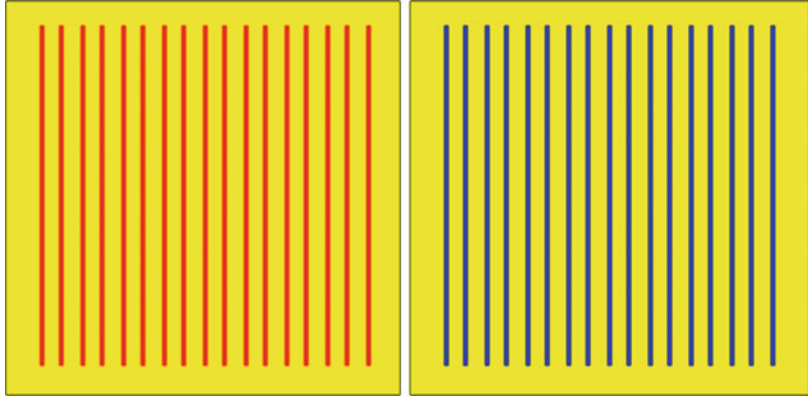
According to a naive physicalism, the physical stimulus should present a direct relationship with its mental correlate. Visual illusions (e.g. [1–3]) indicate, however, that physical manipulations of a stimulus do not directly determine the perceptual experience [4]. The existence of visual illusions has been already reported by ancient philosophers and they were commonly considered counterintuitive singularities [5], which demonstrate the active work of the sensorial system involved in the stimulus processing [6, 7]. It was proved that also non-human animals can experience visual illusions [8, 9]. An important family of this illusion is the brightness illusion and this also shows a similar counterpart in colored stimuli. It is often assumed that perceiving a surface as a source of light depends just on its physical radiant emission. However, the Persian natural philosopher Ibn Al-Haytham (circa 965–1040 AD), known as Alhazen, stressed the subjective nature of color sensation and argued that color appearance was partly due to a mental process in his description of the simultaneous contrast [10].

Simultaneous contrast can be described as follows: a gray target surrounded by a bright inducer that appears darker than its physical value. If the same gray target will be surrounded by a dark inducer, the results will be the opposite: the target will be perceived lighter than its physical value. The simultaneous contrast can be enhanced by blurring the boundaries of the inducer [11–14]. The simultaneous contrast is observed also in color stimuli where the color of the target tends to the hue of the opposite color of the inducer.

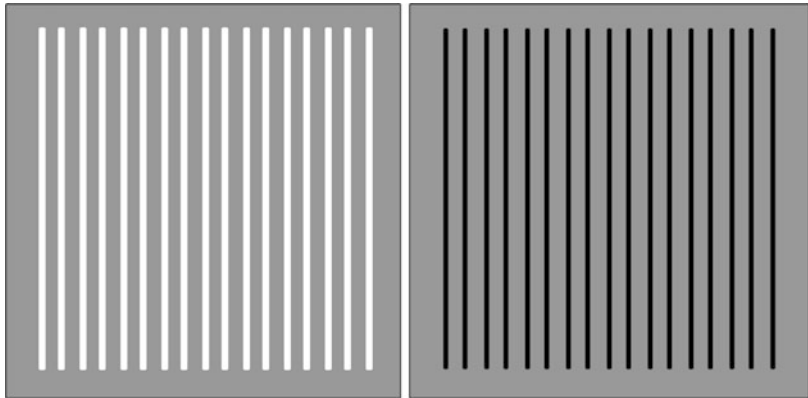
Assimilation Description

The brightness assimilation can be considered the opposite of the simultaneous contrast because under specific conditions a gray target will appear lighter when bordered by a brighter inducer and

Assimilation, Fig. 1 The color assimilation display: the yellow background on the left is perceived reddish while the yellow background on the right is perceived bluish. The two yellows are physically the exact the same hue



Assimilation, Fig. 2 The brightness assimilation display: the gray background on the left is perceived lighter while the gray background on the right is commonly perceived darker. The two gray values present the exact same luminance



vice versa. The color assimilation is a similar phenomenon in which a colored target tends to be perceived as similar in hue to the inducer color. Figure 1 shows a typical color assimilation display: the yellow background on the left is perceived reddish, while the yellow background on the right is perceived bluish; however, the two yellows are exactly the same hue. Figure 2 shows brightness assimilation: the gray background on the left is perceived lighter, while the gray background on the right is commonly perceived darker; however, the two gray values have the exact same luminance. von Bezold [15] probably described first this effect in 1874. Based on the seminal observations by von Bezold, it seems that reducing the inducers' size and increasing their number (which implies increasing their spatial frequency and density together) will lead to a shift from simultaneous contrast to assimilation. Consequently, it seems clear that the physical variables responsible for this shift are the spatial

frequency and the density of the inducers. It is known, for example, that those two variables also affected the size perception [16]. Recently, it was demonstrated that even the perception of beauty is influenced by the spatial frequency [17]. However, the questions that remain open are: at which level of the brain processing are those variables interpreted in this specific way? And how does the visual system transform those variables in the final percept?

Possible Explanations for Assimilation

One approach to explain the assimilation is that bottom-up, peripheral mechanisms are sufficient to produce this percept. Assimilation could be indeed produced by the fact that the retinal input can be imagined as a blurry image; consequently, if the visual system will consider the retinal input of a display like the one in Figs. 1 or 2, the final

percept will be in the direction of assimilation. However, considering how important is the cortical process in the final percept, it is hard to believe that at least in typically developed individuals the cortex will simply “accept” the retinal signal without any post processing in order to provide the final percept. Other proposals that consider the assimilation exclusively a bottom-up process are based on local averaging of luminance within large neurons’ receptive fields. The receptive fields are small in the fovea and larger in the periphery and they also increase their size going up in the cortical hierarchy [18]. Specifically, two possible mechanisms have been proposed: a neuronal spatial integration [19] or neural weighted averages [20], suggesting that the primary anatomical site for assimilation could be spatially close but still outside V1. Both mechanisms may result in assimilation when the physical stimulus is a similar pattern to the one showed in Figs. 1 and 2. However, some predictions based on this explanation fell short in front of the lab tests. For example, DeValois and DeValois [21] suggested that stronger assimilation should be found for color in comparison with grayscale stimuli, because of the lack of lateral inhibition in chromatic receptive fields. However, the results by de Weert and Spillmann [22] showed that it was not the case.

Another approach that is not necessarily opposite to the bottom-up interpretation suggests that assimilation is primarily generated by more central mechanisms of visual processing, such as figure–ground segmentation [23, 24] and observer expertise [25]. It is also important to note that assimilation received interest based on the White illusion [26]. It has been suggested that assimilation depends on the existence of T-junctions that produce a perception of figure–ground segregation [27, 28]. T-junctions seem, for example, to affect also illusory motion [29]. However, this explanation was not supported by the lab test, which demonstrated that assimilation effects can also be seen in versions of White’s display where T-junctions have been completely removed [30]. More recently, Soranzo et al. [31] supported the central mechanism explanation for assimilation by testing that in stroboscopic conditions. In

2010 Rude [32] proposed an intriguing computational neural model that includes the effect of the top-down attentional control in explaining the assimilation effect.

In summary the assimilation is an interesting phenomenon that is still searching for a convincing explanation that can keep all the experimental results under the same theoretical umbrella. However, several important steps were done in order to explain how the brain interprets these patterns.

Cross-References

- [Afterimage](#)
- [Al-Haytham \(Alhazen\), Abū ‘Alī al-Ḥasan ibn al-Ḥasan ibn](#)
- [Anchoring Theory of Lightness](#)
- [Color Contrast](#)
- [Complementary Colors](#)
- [Simultaneous Color Contrast](#)

References

1. Gori, S., Giora, E., Stubbs, D.A.: Perceptual compromise between apparent and veridical motion indices: the Unchained-Dots illusion. *Perception* **39**, 863–866 (2010)
2. Gori, S., Giora, E., Yazdanbakhsh, A., Mingolla, E.: A new motion illusion based on competition between two kinds of motion processing units: The Accordion-Grating. *Neural Netw.* **24**, 1082–1092 (2011)
3. Kitaoka, A., Ashida, J.: Phenomenal characteristics of the peripheral drift illusion. *Vision*. **15**, 261–262 (2003)
4. Gori, S., Giora, E., Pedersini, R.: Perceptual compromise between apparent and veridical motion indices: the Unchained-Dots illusion. *Acta Psychol. (Amst)* **129**, 399–409 (2008)
5. Gregory, R.L.: Knowledge in perception and illusion. *Phil. Trans. R. Soc. London* **352**, 1121–1127 (1997)
6. Eagleman, D.M.: Visual illusions and neurobiology. *Nat. Rev. Neurosci.* **2**, 920–926 (2001)
7. Spillmann, L.: Phenomenology and neurophysiological correlations: two approaches to perception research. *Vision Res.* **49**, 1507–1521 (2009)
8. Gori, S., Agrillo, C., Dadda, M., Bisazza, A.: Do Fish Perceive Illusory Motion? *Scientific Reports*. **4**, 6443 (2014) doi:10.1038/srep06443.
9. Agrillo, C., Gori, S., Beran, M.J.: Do rhesus monkeys (*Macaca mulatta*) perceive illusory motion? *Animal Cognition*. **18**(4), 895–910 (2015)

10. Kingdom, F.A.A.: Simultaneous contrast: the legacies of Hering and Helmholtz. *Perception* **26**, 673–677 (1997)
11. Agostini, T., Galmonte, A.A.: New effect of luminance gradient on achromatic simultaneous contrast. *Psychon. Bull. Rev.* **9**, 264–269 (2002)
12. Gori, S., Stubbs, D.A.: A new set of illusions – the dynamic luminance gradient illusion and the breathing light illusion. *Perception* **35**, 1573–1577 (2006)
13. Anstis, S., Gori, S., Wehrhahn, C.: Afterimages and the breathing light illusion. *Perception* **36**, 791–794 (2007)
14. Gori, S., Giora, E., Agostini, T.: Measuring the breathing light illusion by means of induced simultaneous contrast. *Perception* **39**, 5–12 (2010)
15. von Bezold, W.: *Die farbelehre im Hinblick auf Kunst und Kuntsgewerbe*. Brunswick; Westermann. (1874)
16. Giora, E., Gori, S.: The perceptual expansion of a filled area depends on textural characteristics. *Vision Res.* **50**, 2466–2475 (2010)
17. Vannucci, M., Gori, S., Kojima, H.: The spatial frequencies influence the aesthetic judgment of buildings transculturally. *Cogn Neurosci.* **5** (3–4), 143–149 (2015) doi:10.1080/17588928.2014.976188.
18. Yazdanbakhsh, A., Gori, S.: A new psychophysical estimation of the receptive field size. *Neurosci. Lett.* **438**, 246–251 (2008)
19. Helson, H.: *Adaptation-level Theory*. Harper & Row, New York (1964)
20. Reid, R.C., Shapley, R.: Brightness induction by local contrast and the spatial dependence of assimilation. *Vision Res.* **28**, 115–132 (1988)
21. DeValois, R.L., DeValois, K.K.: *Spatial Vision*. Oxford University Press, Oxford (1988)
22. de Weert, C.M.M., Spillmann, L.: Assimilation: asymmetry between brightness and darkness? *Vision Res.* **50**, 2466–2475 (1995)
23. Musatti, C.: Forma e assimilazione. *Archivio Italiano di Psicologia.* **9**, 213–269 (1931)
24. de Weert, C.C.M., van Kruysbergen, N.: Assimilation: central and peripheral effects. *Perception* **26**, 1217–1224 (1997)
25. Kanizsa, G.: *Organization in Vision: Essays on Gestalt perception*. Praeger, New York (1979)
26. White, M.: A new effect of pattern on perceived lightness. *Perception* **8**, 413–416 (1979)
27. Anderson, B.L.: A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. *Perception* **26**, 419–453 (1997)
28. Todorović, D.: Lightness and junctions. *Perception* **26**, 379–394 (1997)
29. Gori, S., Hamburger, K., Spillmann, L.: Reversal of apparent rotation in the Enigma-figure with and without motion adaptation and the effect of T-junctions. *Vision Res.* **46**, 3267–3273 (2006)
30. Yazdanbakhsh, A., Arabzadeh, E., Babadi, B., Fazl, A.: Munker–White-like illusions without T-junctions. *Perception* **31**, 711–715 (2002)
31. Soranzo, A., Galmonte, A., Agostini, T.: von Bezold assimilation effect reverses in stereoscopic conditions. *Perception* **39**, 592–605 (2010)
32. Rudd, M.E.: How attention and contrast gain control interact to regulate lightness contrast and assimilation: a computational neural model. *J. Vis.* **10**, 1–37 (2010)

Aura

► Scintillating Scotoma

Automotive Lighting

Stephan Volker

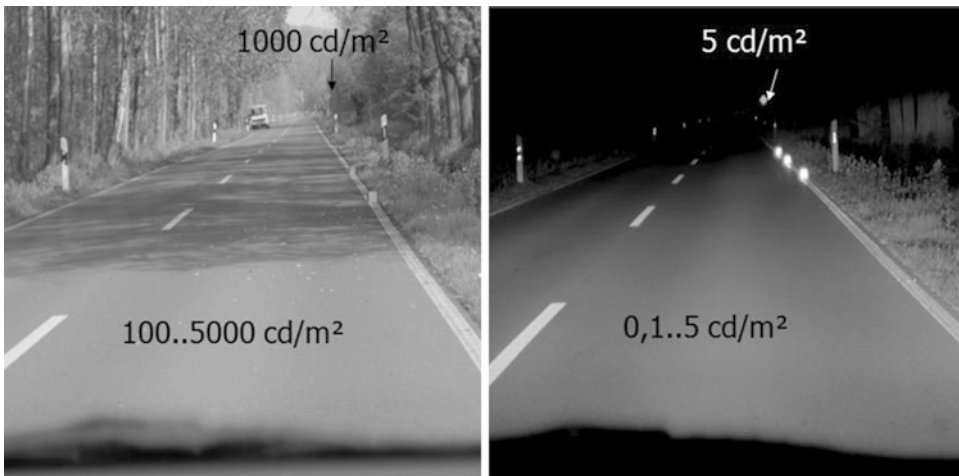
University of Technology Berlin, Berlin,
Germany

Synonyms

[Car lighting](#); [Headlamps](#); [Perception by night](#)

Definition

Light is needed for road safety at night, supporting the driver's orientation and early recognition of obstacles. This light is produced by the headlamp. However, the same light useful for the driver will also glare oncoming traffic. Thus, the light of headlamps increases and reduces road safety at the same time. The main task of headlamp designers is to solve the optimization problem between a good view for the driver and as little glare as possible for oncoming traffic. But not only the own headlights are necessary for a good orientation of the driver by night. Signal lamps of other cars give information about the change of direction, reducing of speed, drive back, and the contour of cars. So automotive lighting deals with the construction of headlamps and signal lamps based on perception conditions by night.



Automotive Lighting, Fig. 1 Compare the luminances of night and day of the same street [1]

Overview

The following article will give a short overview about automotive lighting. It is focused on car lighting and does not lit the interaction between car and street lighting. It starts with the perception conditions by night. The knowledge of this field helps to understand the special construction of current and future headlamps.

Perception at Night Time

Figure 1 shows the same street during the day and at nighttime. The gray values represent luminances. The day-to-night luminance ratio is 1:1,000. Additionally, the quantity of information is much higher during the day than it is at night. The daylight viewing distance of 200 m and more opposes to no more than 65 m in darkness – very often even lower. However, the driving speed is nearly the same. The density of traffic is lower in the middle of the night – still in winter between 17 and 20 h, traffic density is nearly the same as in summer. It is clear that perception conditions at nighttime are completely different from daylight conditions.

What does that mean for the perception? The visual power functions that followed give an answer. Figure 2 left shows the contrast sensitivity

and Fig. 2 right the visual acuity as a function of adaptation luminance.

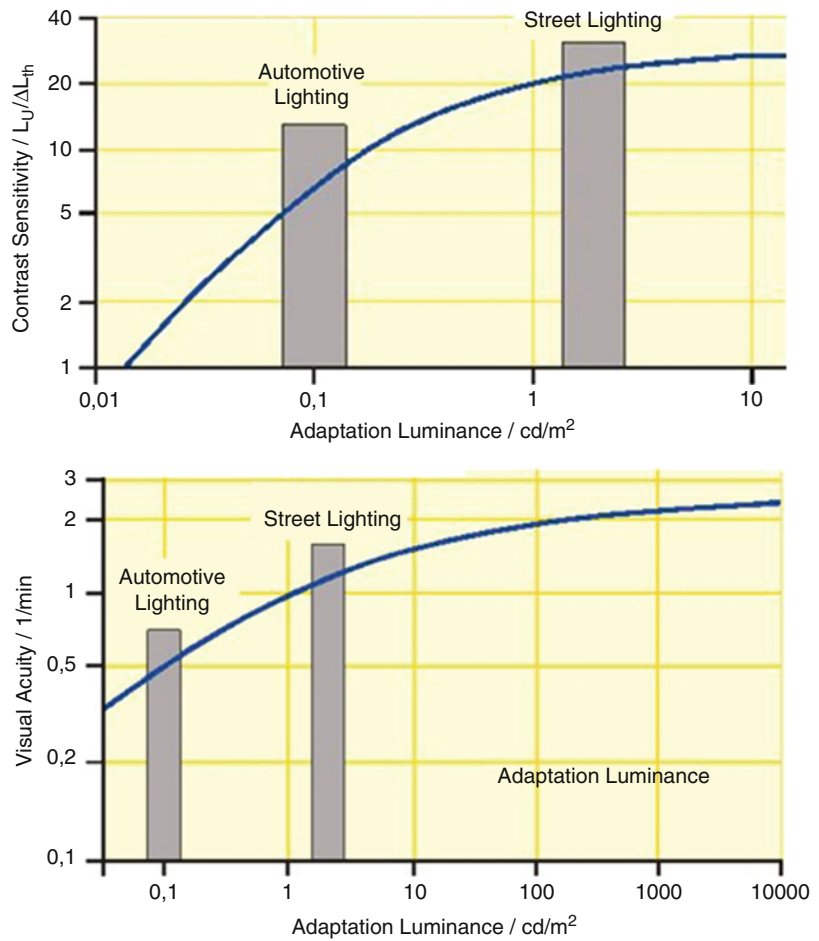
Because the luminance of the road is not constant, the right adaptation level has to be defined. Figure 3 gives an impression in which area the driver is looking for objects. According to this data, the accumulation point of view is in 30–60 m ahead of the car where luminance levels are measured to 0.01–0.1 cd/m^2 depending on the road surface and the luminance intensity of the headlamp. Consequently, the contrast sensitivity and the visual acuity (Fig. 2) are much lower compared to daytime or road lighting running conditions.

Another very important point is glare for oncoming traffic caused by car headlamps. Figure 4 shows increasing of the vertical illuminance of an oncoming car. Figure 5 presents the threshold contrast depending on the vertical illuminance at the eye. The threshold contrast is doubled if the legally required illuminance value of 0.5 lx is reached. Because light above the horizontal line is needed for recognition of traffic signs, bridges, tunnels, and so on, a typical optimization task has to be solved.

Construction of Headlamps

The aim of a good headlamp is to bring as much light as possible to the road surface beyond a

Automotive Lighting,
Fig. 2 Contrast sensitivity
 (up) and visual acuity
 (down) as function of
 adaptation luminance
 (~road luminance)

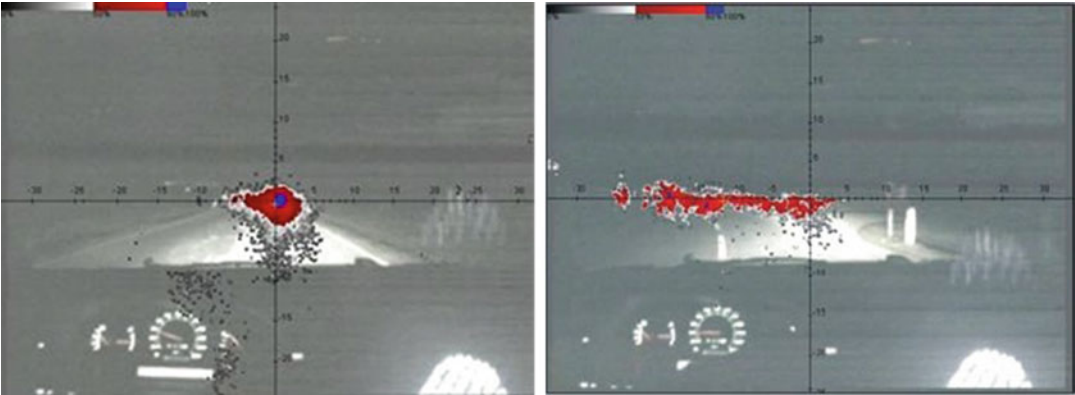


distance of 30 m. The headlamp should not glare oncoming traffic additionally. So a limitation of light above the horizontal line is necessary. That leads to the typical light distribution of a low beam headlamp shown in Fig. 6 (up). The illuminance distribution is demonstrated in false color on a measuring screen in 25 m distance. Because of the dynamic swiveling of a car and production tolerances in the headlamp, the cutoff line has to be tilted down by a minimum of 1° . Only if there is no oncoming traffic can the driver use high beam headlamp; see Fig. 6 (down). High beam is used in Europe approximately between 5 % and 15 % at night. This very rough estimate shows the limitation of good driving conditions.

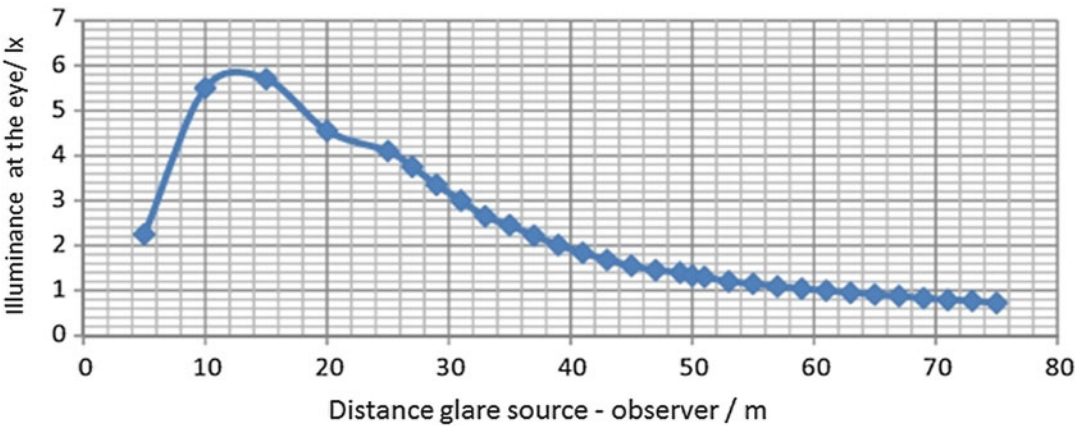
What is needed to get such a light distribution?
 The headlamp has to produce:

1. A hot spot with more than 10 lx in 25 m
2. More than 1 lx in an area between $\pm 30^\circ$ (left/right) and -10° to 0° (up/down)
3. A cutoff line with a gradient of 20:1 on the vertical lines by $\pm 5^\circ$
4. Less stray light above the cut-off-line (< 0.4 lx for halogen lamps and < 0.5 lx for discharge lamps)

Light sources with high luminances are needed to create a bright hot spot. So halogen lamps, gas discharge lamps, or LED is used. For

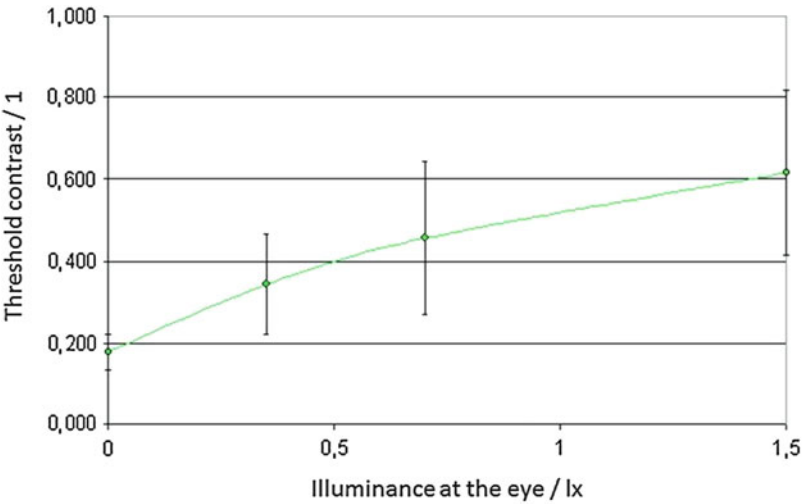


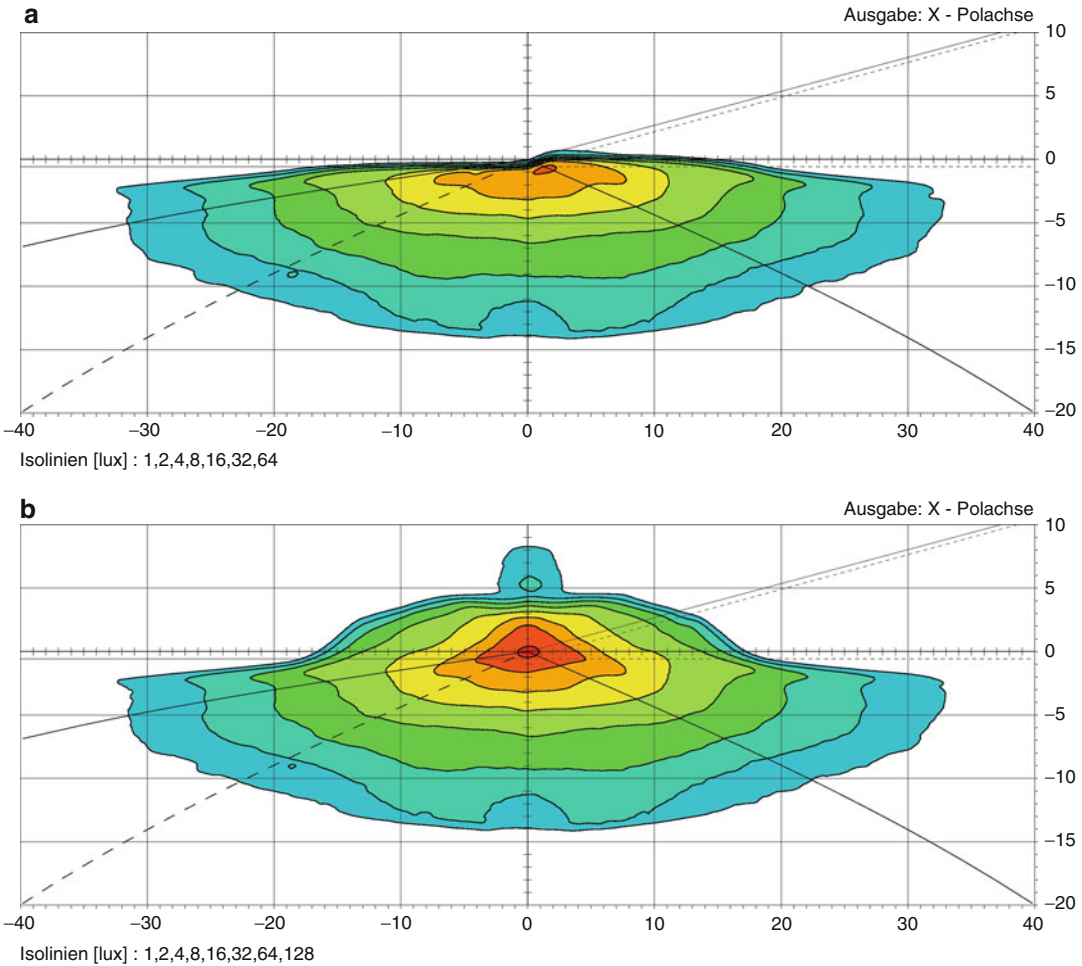
Automotive Lighting, Fig. 3 Accumulation point of view on a straight road (*left*) and a bending road (*right*) [2]



Automotive Lighting, Fig. 4 Vertical illuminance at the eye produced on oncoming car is situated in different distances [3]

Automotive Lighting, Fig. 5 Threshold contrast as function of different illuminance level on the eye [4]





Automotive Lighting, Fig. 6 Illuminance distribution of a low beam (*up*) and a high beam (*down*) on a 10-m-wall measured in 25 m presented in false color [Hella]

the shape of the cutoff line, the light source should have an aspect ratio of 5:1. Figure 7 shows the luminance pictures of a halogen lamp and a LED module fulfilling this requirement.

Generally, in addition to the lamp, a reflector and/or a lens is necessary to distribute the light in the desired direction. The following standard reflectors are used:

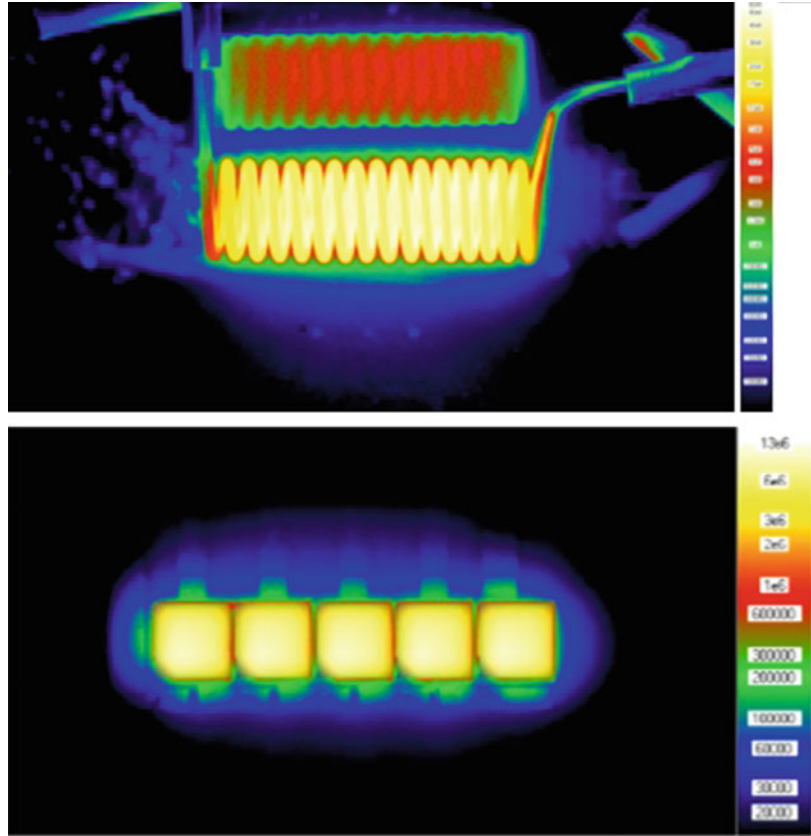
1. Parabolic (lenses in cover glass) – Fig. 8a
2. Freeform reflector (transparency of cover glass) – Fig. 8b
3. Ellipsoid (projection system with projection lens and shutter) – Fig. 8c.

Figure 8c presents the simulation of the light distribution in a projection system (down) and demonstrates the beam of a real system (up). The comparison of a reflection and a projection system is illustrated in Table 1.

Every headlamp has to fulfill the demands and regulations of ECE (for Europe, Asia, Australia, South America) or SAE (for North America). These standards describe the photometric and geometric requirements of headlamps and are

Automotive Lighting,

Fig. 7 Luminance picture of a halogen bulb (*up*) and an LED module (*down*)



adapted to the state of the technological development.

Construction of Signal Lights

A headlamp supports the driver's vision and provides information about oncoming traffic and obstacles. Further lighting equipment on a car is necessary to signalize position, speed, and change of direction. All lighting equipment not supporting vision but to be seen is called signal lights or signal lamps. Stop lamps, rear lamps, day time running lamps, marking lamps, and position lamps belong to this group. Figure 9 shows some lamps.

What are the requirements for signal lamps? The luminance of the light source should not be as high as for a headlamp. The light distribution has to be very wide. Figure 10 shows the relative luminous intensity.

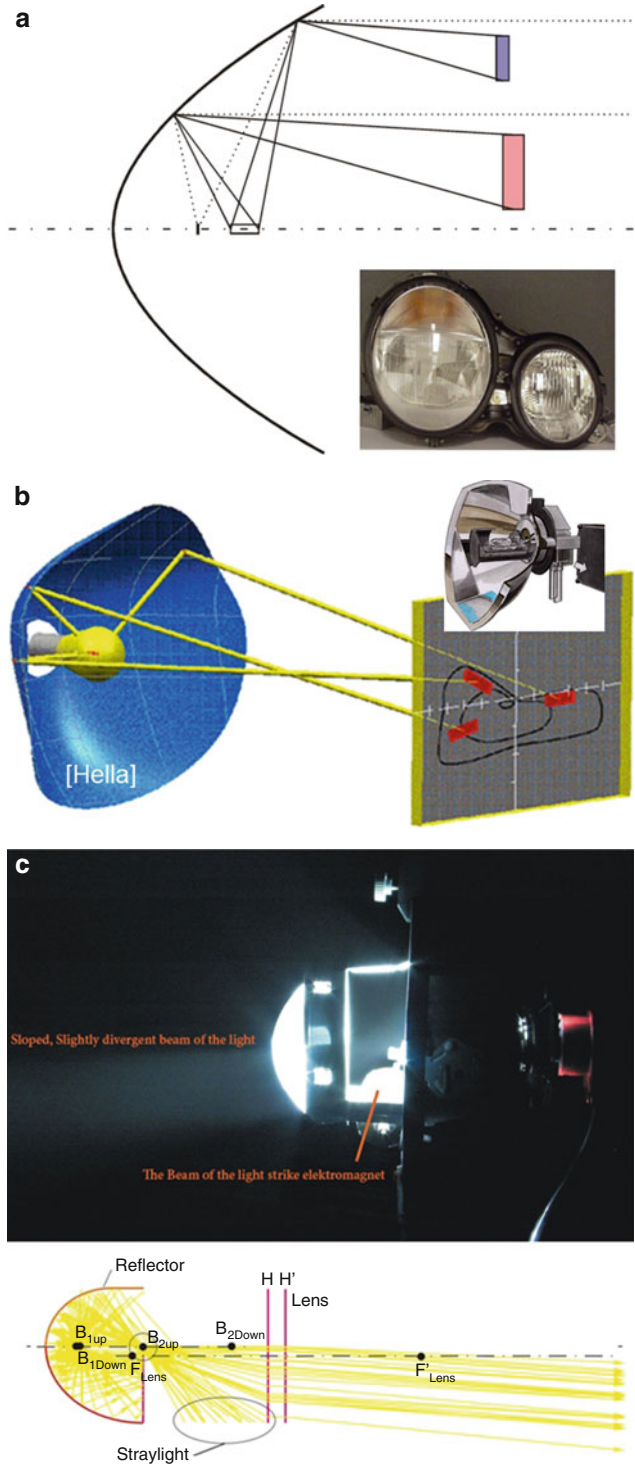
The usage of white light for signal functions is prohibited. Signals have to be well defined. Different colors, different luminance, and blinking are used to indicate different information (Fig. 11). For example, a higher luminance is used for a stop lamp than for a rear lamp. Blinking yellow light (for ECE) or red light (for SAE) shows that the driver will change his/her direction.

Vision for Automotive Lighting

Headlamps will change significantly in the next decade. The headlamps presented above cannot completely solve the task to produce enough light for the driver with no glare for oncoming traffic. Considering the maximum viewing distance of 65 m, the maximum speed should not be more than 65 km/h without high beam. At a higher speed, these headlamps do not provide enough light on the left side to see objects in

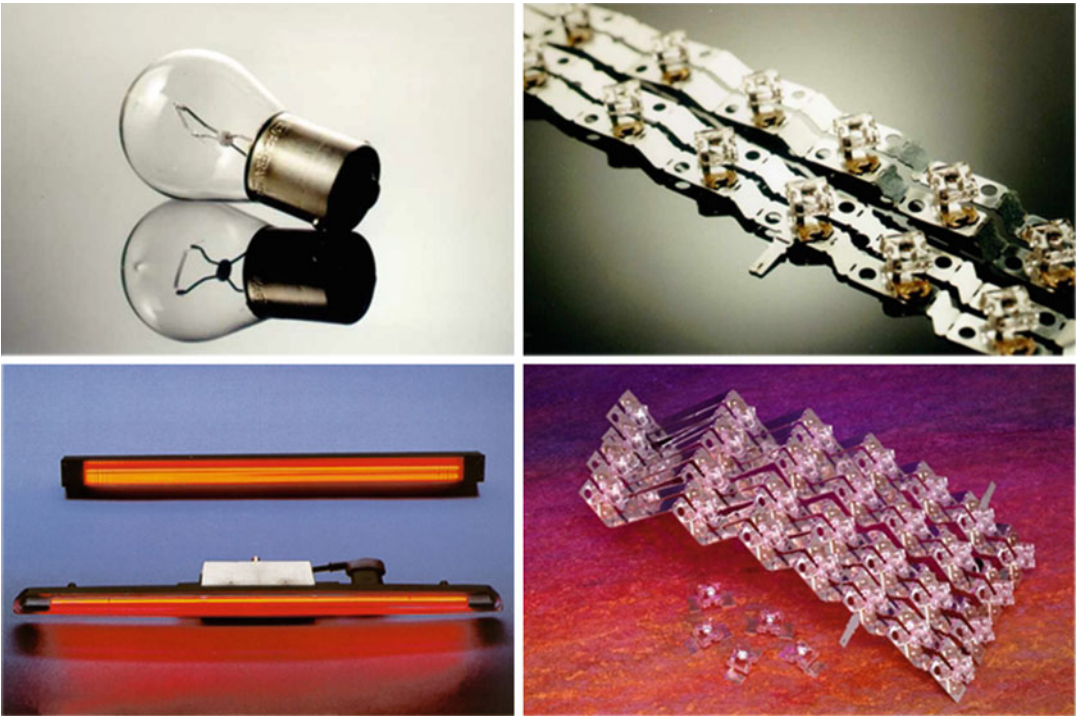
Automotive Lighting,

Fig. 8 Different reflector systems (a): parabolic [Hella]; (b): freeform [Hella]; (c): ellipsoid [5]



Automotive Lighting, Table 1 Comparison of reflection and projection system

	Reflection system	Projection system
Cover glass	With lenses/transparent	Transparent
Luminous surface	Large	Small
Optical system	Large reflector + cover glass with lenses	Small reflector + shutter + projection lens + transparent cover glass
Efficacy	High	Low
Reflector type	Parabolic/freeform	Ellipsoid/freeform
Depth	Small	Large
Height/width	Large	Small



Automotive Lighting, Fig. 9 Different signal lamps (incandescent lamps, LED, Neon tubes) [Hella]

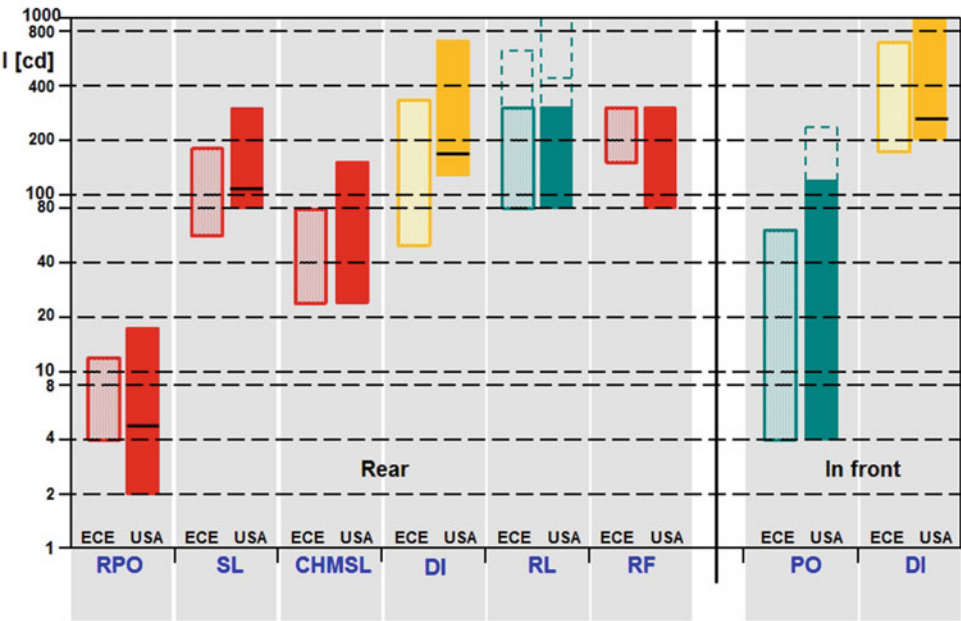
time to react. Development of headlamps should look into providing high beam all the time without causing glare for the oncoming driver. Some prototypes of such headlamps have been available lately.

A glare-free headlamp needs a sensor system and actuating element. The aim of such a

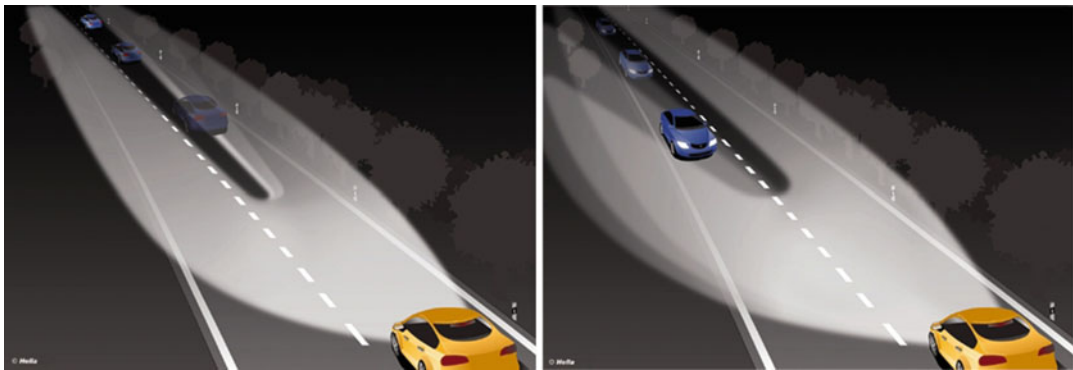
system is to detect oncoming traffic and to blank out the light in the direction of the oncoming cars immediately (Fig. 12). Radar, lidar, infrared scanners, or cameras are used as sensors. A LED array (Fig. 13) or a dynamic AFS system (actor) is responsible for the adaptive light control. This way, it is possible to bring enough light

Automotive Lighting, Fig. 10 Relative luminous intensity as a function of angle (maximum luminous intensity = 100)

	30°	20°	10°	5°	0°	5°	10°	20°	30°
15°									
10°				20		20			
5°		10	20		70		20	10	
0°			35	90	100	90	35		
5°		10	20		70		20	10	
10°				20		20			
15°									



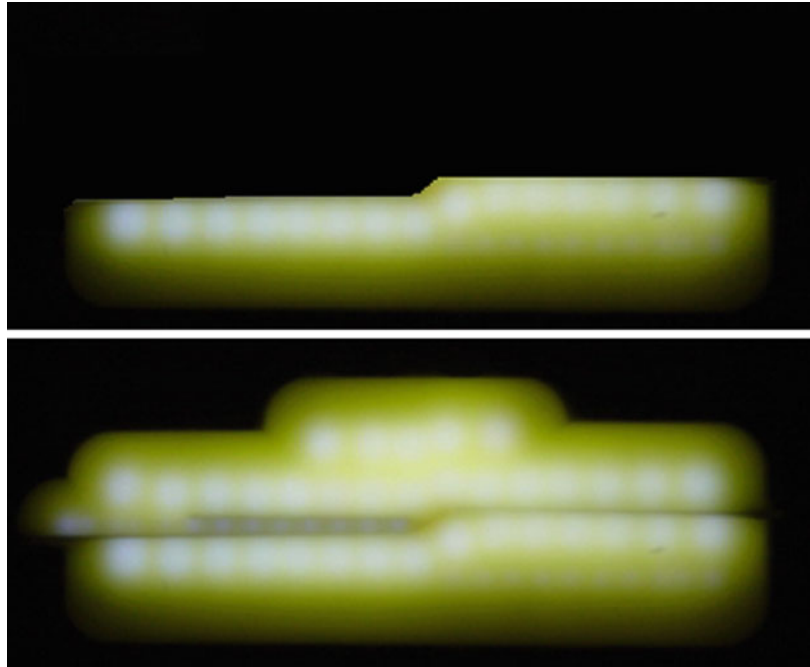
Automotive Lighting, Fig. 11 Luminous intensity and light color of different signal function [Hella]



Automotive Lighting, Fig. 12 The function of a glare-free headlamp [Hella]

Automotive Lighting,

Fig. 13 LED array – *up*
low beam; *down* high
beam [5]



A

right and left of an oncoming car without glaring the driver.

With this “glare-free” technology in a large number of cars, severe accidents with pedestrians and bicycles at night could be reduced dramatically. To achieve this, further technological improvements are necessary as well as cost reduction and the employment of these systems in a large number of cars. “Intelligent light for safety!” should be a political, economic, and human objective for automotive lighting in the future.

Cross-References

- [Headlamps](#)
- [Road Lighting](#)

References

1. Völker, S.: Hell- und Kontrastempfindung – ein Beitrag zur Entwicklung von Zielfunktionen für die Auslegung von Kraftfahrzeug-Scheinwerfern. Habilitationsschrift Universität Paderborn, 2006; 231 S
2. Stahl, F.: Kongruenz des Blickverlaufs bei virtuellen und realen Autofahrten -Validierung eines Nachtfahrssimulators; Diplomarbeit L-LAB/TU Ilmenau 2004
3. Michenfelder, S.: Ermittlung des Einflusses einer ambienten Innenraumbelichtung auf das Kontrastsehen des Fahrzeugführers. Diplomarbeit, LTIK am KIT Karlsruhe, 2010
4. Freyer, M.: Einfluss der Blendbeleuchtungsstärke und der Leuchtdichte auf die Blendung und den Schwellenkontrast bei homogenen und inhomogenem Umfeld. Diplomarbeit, L-LAB 2004
5. Strauß, S.: Theoretische und experimentelle Untersuchungen zum Einsatz gepulster Halbleiterlichtquellen in der Kraftfahrzeugbeleuchtung. Paderborn, Universität, Dissertation, 2007.

Autosomal Dominant Inherited Color Vision Deficiency

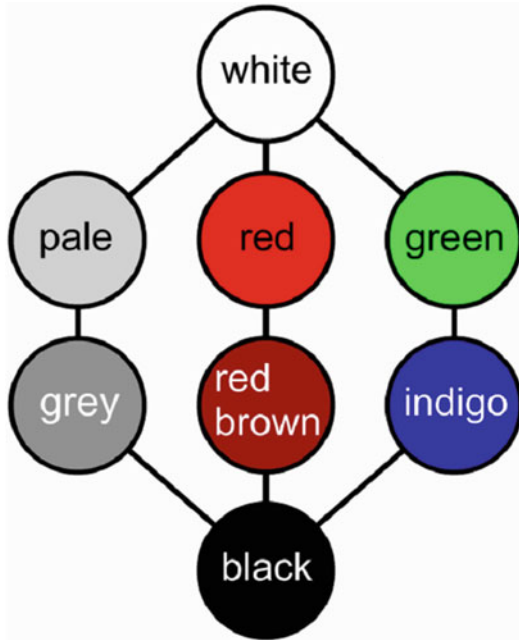
- [Tritanopia](#)

Avicenna, Abū 'Alī al-Ḥusayn ibn 'Abd Allāh ibn Sīnā

Eric Kirchner¹ and Seyed Hossein Amirshahi²

¹Kirchner Publications, Leiden, The Netherlands

²AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran



Iran's Abu 'Alī al-Ḥusayn ibn 'Abd Allāh ibn Sīnā (Latinized name: Avicenna) was a Persian polymath and probably the most influential natural philosopher in all of Islamic history [1]. He was born around 980 AD near Bukhara (present-day Uzbekistan) and died in Hamadan (Iran). Already at the age of 16, famous physicians worked under his direction, and at the age of 18, Ibn Sina mastered the contemporary knowledge of the various sciences [2]. His massive *Qanun fi-'l-tibb* (Canon of Medicine) would be the main medical text in the East and West until the seventeenth century.

The *Kitab al-Shifa* (The cure [of ignorance]) is an equally immense four-part encyclopedia on mathematics, physics, and metaphysics [3, 4]. For several centuries this would be the main text on Aristotelian natural philosophy. Here, Ibn

Sina displayed his Aristotelian sympathies mainly by refuting alternative theories. For theories on light and color, this meant that Ibn Sina rejected the so-called extramission theories advocated by Ptolemy and Euclid. Ibn Sina gave good arguments why it is absurd to assume that vision occurs by visual rays emerging from the eye. Instead, Ibn Sina explained vision in terms of forms transmitted from the visible object to the eye (i.e., by intromission). In this respect, he followed Aristotle just like Ibn Rushd [5]. It would be Ibn al-Haytham, a contemporary of Ibn Sina, who would formulate a successful fusion of the optical theories of Ptolemy, Euclid, Galen, and Aristotle. It would be an intromission theory.

The *Kitab al-Shifa* contains a full chapter on color [6, 7]. In chapter 4 section 3, Ibn Sina was the first to break with the Aristotelian notion that all colors can be ordered along a one-dimensional line, writing:

... the fact that white gradually passes to black by three paths. The first is via pale [...], at first it progresses to pale, from there to grey, and continuing in this manner until black is obtained [...]. There is also another path proceeding [from whiteness] toward red, and from there to red brown, thereafter to black. The third path is the one going to green, from there to indigo, thereafter to blackness. [4, 7]

Thus, Ibn Sina introduced what in modern terms would be called a two-dimensional color order (cf. *the illustration above*) [8]. It would become known in Christian Europe as well, for example, through its inclusion in the *Speculum maius* from Vincent de Beauvais (1244) [7]. Ibn Sina's texts on color were also widely discussed in the Islamic world. When more than two centuries after Ibn Sina's death a student of the astronomer al-Tusi formulated questions about Ibn Sina's color ordering, al-Tusi further elaborated on this theory [9].

In his commentary, Ibn Sina also criticized the Aristotelian ideas on color mixing. For example, he disagreed with the Aristotelian tradition that claimed that green is composed of red and purple. According to Ibn Sina, a composition of red and purple does not produce green because from mixing red with purple, a color is produced that

is brighter than purple but more purple than bright red. Instead, according to Ibn Sina green is formed by a mixture of yellow, black, and indigo blue [10]. Further, Ibn Sina made a clear distinction between the brightness of a light source (*lux*, in the Latin translation of Ibn Sina’s work) and the “splendor” shining from an object (*lumen*) [11].

Finally, Ibn Sina was highly critical about what was known about the cause of the rainbow. He wrote that he was “not satisfied with what our friends the Peripatetics [i.e., Aristotle and his followers] have to say about the rainbow.” He continued by reporting some of his observations that cannot be explained by that theory, only to conclude that he himself also had nothing to add, except by a suggestion that the cause of these colors might be inside the eye of the observer [5].

References

1. Lindberg, D.C.: Theories of Vision from al-Kindi to Kepler. University of Chicago Press, Chicago (1976)
2. Anawati, G.C.: Ibn Sina. In: Gillespie, C.C. (ed.) Dictionary of Scientific Bibliography, pp. 494–498. Scribner’s sons, New York (1970)
3. Ibn Sina: Kitab al-Shifa. In: Muntasir, A., Zayid, S., Ismail, A., Madkur, I. (eds.) at-Tabi’iyyat 5; Al-ma’adin wa-l-athar al-‘ulwiyya. Cairo (1964). Also Qasim, M., Madkur, I. (eds.) at-Tabi’iyyat 4; Fi l-af’al wa-l-infi’alat. Cairo (1969). Latin translation: van Riet, S. (ed.) Avicenna Latinus: Liber de Anima I-II-III. Édition critique de la traduction latine médiévale. Brill, Louvain-Leiden (1972)
4. Bakos, J.: Psychologie d’Ibn Sina (Avicenne) d’après son oeuvre ash-Shifa. Académie Tchèque des Sciences, Prague (1956). Vol. 1, 105–113; Vol. 2, 75–81
5. Sabra, A.I.: Optics, Islamic. In: Strayer, J.R. (ed.) Dictionary of the Middle Ages, pp. 240–247. Scribner’s sons, New York (1982–1989)
6. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) Encyclopedia of Islam. Brill, Leiden (1991)
7. Kuehni, R.G., Schwarz, A.: Color Ordered – A Survey of Color Order Systems from Antiquity to the Present. Oxford University Press, Oxford (2008)
8. Kirchner, E.: Color theory and color order in medieval Islam: a review. Col. Res. 40 (2015) 5–16.
9. Kirchner, E., Bagheri, M.: Color theory in medieval lapidaries: Nishaburi, Tusi and Kashani. Centaurus 55, 1–19 (2013)
10. van Campen, M.: De regenboog bij de Arabieren. University of Utrecht, Utrecht (1988). MSc Thesis, 95
11. Lindberg, D.C.: Optics, Western European. In: Strayer, J.R. (ed.) Dictionary of the Middle Ages, vol. 9, pp. 247–253. Scribner’s sons, New York (1989)

B

Balance

► [Color Harmony](#)

Bartleson, C. James

Robert W. G. Hunt
Department of Colour Science, University of
Leeds, Leeds, UK



C. James Bartleson was an American color scientist who made very important contributions to colorimetry and visual science.

Bartleson graduated as an Associate of Photographic Science at the School of Photography at Rochester Institute of Technology (RIT), Rochester NY, USA, in 1951. He received a Ph.D. at the City University, London, England, in 1977.

After high school, he enlisted in the marines and served as a frogman using photography. After his discharge in 1948, he spent a year doing freelance aerial photography and then enrolled in the School of Photography at RIT to obtain his degree of Associate of Photographic Science in 1951. There followed a 3-year period working on the new Ansco Plenachrome system at the Pavelle Color Corporation in New York City.

In 1952, he moved to Rochester, NY, to start a career with the Eastman Kodak Company. His first assignment was in the Color Control Department under Ralph Evans but his flair for research was noticed and in 1957 he moved to the Physics Division where he started his research on color and tone reproduction which led to the publication of a series of very significant papers, one of the most outstanding, in 1967, coauthored with Ed Breneman, being on the effect of light and dark surrounds on apparent contrast. His papers attracted various awards, and he was one of the first recipients of the C.E.K. Mees Award. He was

also author, coauthor, or editor of several books, including the five volume series on “Optical Radiation Measurements.” He was also much sought after as a lecturer, and his award of the British Color Group’s Newton Medal was accompanied by a masterly presentation.

In 1967, his outstanding reputation led him to be chosen to establish a research facility for the MacBeth Company, later the MacBeth Color and Photometry Division of the Kollmorgen Corporation, in Newburgh, NY.

When his time at Kollmorgen came to an end in 1974, the following year, at the then age of 45, he enrolled in the School of Ophthalmic Optics and Visual Science at The City University, London, England, and obtained his Ph.D. in 1977. With typical enthusiasm, he constructed an ingenious apparatus for studying the effects of adaptation on perception, and the resulting thesis was of outstanding quality.

He then returned to Rochester, NY, to resume his career at Kodak, where he soon established a reputation as a most valuable consultant for other members of the staff.

He was the President of the AIC (Association Internationale de la Couleur) from 1978 to 1981, reviving it from a moribund state, and he took part in several committees of the CIE (Commission Internationale de l’Éclairage).

His main interests included optics, vision, photometry, colorimetry, color perception, color applications, and color photography. His leisure interests also included photography and travel.

Further Reading

1. Burnham, R.W., Haynes, R.M., Bartleson, C.J.: *Color: A Guide to Basic Facts and Concepts*. Wiley, New York (1963)
2. Grum, F. (ed.): *Optical Radiation Measurements*. Academic, New York (1984)
3. Grum, F., Becherer, R. J.: *Optical Radiation Measurements*. volume 1 Radiometry. Academic, New York (1979)
4. Grum, F., Becherer, R. J.: *Optical Radiation Measurements*. volume 2 Color Measurement. Academic, New York (1980)
5. Mielenz, K. D. (ed.): *Optical Radiation Measurements*. volume 3 Measurement of Photoluminescence. Academic, New York (1982)
6. Budde, W.: *Optical Radiation Measurements*. volume 4 Physical Detectors of Optical Radiation. Academic, New York (1983)
7. Grum, F., Becherer, R. J.: *Optical Radiation Measurements*. volume 5 Visual Measurements. Academic, New York (1984)

Basic Colors

► Primary Colors

Bayesian Approaches to Color Category Learning

Thomas L. Griffiths

Department of Psychology, University of California at Berkeley, Berkeley, CA, USA

Synonyms

[Ideal observer models of color category learning](#);
[Rational models of color category learning](#)

Definition

Bayesian approaches to color category learning formalize learning as a problem of Bayesian inference, requiring the learner to form generalizations that go beyond observed examples of members of a category. This formal framework can be used to make predictions about both individual judgments and how populations form color categories.

Color Category Learning

One of the challenges that children face as they acquire a language is discovering how words are used to refer to different colors. While human languages demonstrate variation in how they

partition the space of colors, there are also clear regularities in the kinds of systems of color categories that are used [1]. This raises two important questions: How might color categories be learned? And how might regularities in systems of color categories across languages be explained?

Learning color categories is an inductive problem, requiring learners to make an inference from labeled examples of colors to a full system of color categories. As in other domains of perception [2], an “ideal observer” model can be used to explore the optimal solution to this problem. Let h denote a hypothesis about a possible system of color categories and d the observed data – a set of labeled examples (such as “This color is *blue*, and this color is *yellow*”). If learners represent the degree of belief in the truth of each hypothesis with a probability, $P(h)$, then the ideal solution to the problem of updating these beliefs in light of the data d is provided by Bayes’ rule:

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h'} P(d|h')P(h')}$$

where $P(h|d)$ (known as the *posterior* probability, in contrast to the *prior* probability $P(h)$) indicates the degree of belief assigned to h after observing d and $P(d|h)$ (known as the *likelihood*) indicates the probability of seeing d if h were true.

The sum in the denominator of Bayes’ rule ranges over all possible hypotheses and ensures that $P(h|d)$ is a valid probability distribution, summing to 1. The key idea behind Bayes’ rule can be obtained by ignoring this constant and simply inspecting the numerator: The new beliefs of the learner result from combining the previous beliefs, captured in the prior distribution $P(h)$, with the probability of the observed data under each hypothesis, expressed by the likelihood $P(d|h)$. The prior distribution captures the expectations of the learner, but also indicates which hypotheses are easy or hard to learn. A hypothesis that has low prior probability requires stronger evidence (in the form of a higher likelihood) to end up with a high posterior probability and so will be harder to learn. The prior

distribution thus provides a way of encoding the perceptual or learning biases of the learner, favoring some hypotheses over others.

The Bayesian approach to modeling learning has proven successful in accounting for human behavior in a wide range of tasks [3]. In particular, Bayesian models have been used to account for how people learn new concepts and new words. Tenenbaum and Griffiths [4] presented an account of how people form generalizations from examples, such as inferring what other numbers might belong to a set when told that the set contains 2, 8, and 64. Under this account, hypotheses correspond to possible sets of numbers, and the likelihood is obtained by calculating how likely it is that the examples would be observed if they were sampled at random from this set. Xu and Tenenbaum [5] showed that a closely related model captured how children learned nouns corresponding to sets of objects, such as determining the appropriate referent of words corresponding to “Dalmatian” or “dog.” These results suggest that a Bayesian approach might also be fruitful for explaining the acquisition of terms for color categories.

Dowman [6] presented a model that took exactly this approach, providing a Bayesian account of color category learning. In this model, the space of colors is reduced to a one-dimensional ring of hues. Each color category then corresponds to an interval on this ring, picking out a set of adjacent colors. Labeled examples of color categories are assumed to be sampled from the categories at random, with a small probability of an error taking place. This makes it possible to calculate the probability of any observed set of labeled examples for each candidate interval from which they might be drawn, providing the likelihood $P(d|h)$. Bayes’ rule can then be used to compute a posterior distribution over possible intervals for each color category. The probability that a color that has not previously been labeled belongs to that category is then obtained by summing the probability of all intervals that contain that color under the posterior distribution. Dowman demonstrated that this model made reasonable inferences for simplified versions of the systems of color categories from real languages, such as Urdu.

The predictions that Dowman's model makes about learning of color categories have not been directly tested with human learners, but results in other domains and with other species provide support for this approach. As noted above, Xu and Tenenbaum [5] found that a very similar model accounted well for the generalizations that children made in learning novel words describing sets of objects. In addition, Jones, Osorio, and Baddeley [7] found that poultry chicks form generalizations about colors in a conditioning task that are consistent with a model based on that of Tenenbaum and Griffiths [4].

The results summarized so far indicate how Bayesian models might be used to explain learning of color categories. The same models also have the potential to provide insight into why regularities exist in the systems of color categories that appear across human languages. Dowman [6] explicitly had this goal in mind in defining his Bayesian learning model, which was used as a component in a simulation of the cultural transmission of systems of color categories. Dowman's aim was to investigate the consequences of cultural transmission of systems of color categories among a set of agents that used a realistic approximation to human learning. He found that cultural transmission by Bayesian agents produced systems of color categories with properties similar to those seen across human languages, providing a potential explanation for the source of those regularities.

Dowman's simulation of cultural transmission was an instance of a more general approach to exploring the origins of different kinds of structure in human languages, known as iterated learning [8]. In an iterated learning model, a set of agents each learn from some observed data and generate data that is observed by other agents. The simplest case is where the agents form a chain, with each agent learning from data generated by the previous agent and generating data that is provided to the next agent. Griffiths and Kalish [9] showed that when this form of iterated learning is carried out by Bayesian agents who all have the same prior, the hypotheses considered by those agents eventually converge to a distribution that matches the prior distribution. More

precisely, the probability that an agent selects a hypothesis h converges to the prior probability of that hypothesis $P(h)$ as the chain gets longer.

The results of Dowman [6] and Griffiths and Kalish [9] raise an interesting question: Can the regularities seen in systems of color categories across human languages be accounted for by cultural transmission producing convergence on a shared prior distribution? To explore this question, Xu, Dowman, and Griffiths [10] conducted an experiment in which human participants simulated cultural transmission by iterated learning. Each participant was given some examples of colors from novel categories and then asked to generalize to a larger set of colors. The generalization responses were then used to generate the examples that were seen by the next participant. Over time, the systems of color categories produced by the participants converged to forms that were consistent with the regularities seen across human languages. These results support the idea that cultural transmission and perceptual or learning biases of the kind that might be captured by a prior distribution in a Bayesian model may be sufficient to explain the origins of cross-linguistic regularities in systems of color categories.

Bayesian approaches to color category learning can be used to explore questions about how children might learn how their language partitions the space of colors and why regularities are seen in systems of color categories across languages. However, this research is still in its early stages, with many important questions remaining open. One fundamental question is how well Bayesian models can capture the generalizations that real human children make when learning color categories. Another is whether it is possible to characterize the human prior distribution over systems of color categories precisely and whether the structure of that distribution can capture cross-linguistic variation.

Cross-References

- [Berlin and Kay Theory](#)
- [Color Categorical Perception](#)
- [World Color Survey](#)

References

1. Kay, P., Maffi, L.: Color appearance and the emergence and evolution of basic color lexicons. *Am. Anthropol.* **101**, 743–760 (1999)
2. Kersten, D., Mamassian, P., Yuille, A.: Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**, 271–304 (2004)
3. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.: How to grow a mind: statistics, structure, and abstraction. *Science* **331**, 1279–1285 (2011)
4. Tenenbaum, J.B., Griffiths, T.L.: Generalization, similarity, and Bayesian inference. *Behav. Brain Sci.* **24**, 629–641 (2001)
5. Xu, F., Tenenbaum, J.B.: Word learning as Bayesian inference. *Psychol. Rev.* **114**, 245–272 (2007)
6. Dowman, M.: Explaining color term typology with an evolutionary model. *Cognit. Sci.* **31**, 99–132 (2007)
7. Jones, C.D., Osorio, D., Baddeley, R.J.: Colour categorization by domestic chicks. *Proc. R. Soc. Lond. B* **268**, 2077–2084 (2001)
8. Kirby, S.: Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Trans. Evol. Comput.* **5**, 102–110 (2001)
9. Griffiths, T.L., Kalish, M.L.: Language evolution by iterated learning with Bayesian agents. *Cognit. Sci.* **31**, 441–480 (2007)
10. Xu, J., Griffiths, T.L., Dowman, M.: Replicating color term universals through human iterated learning. In: Ohlsson, S., Catrambone, R. (eds.) *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pp. 352–357. Cognitive Science Society, Austin (2010)

Behnam's Disk

► [Fechner's Colors and Behnam's Top](#)

Berlin and Kay Theory

C. L. Hardin
Department of Philosophy, Syracuse University,
Syracuse, NY, USA

Definition

The Berlin-Kay theory of basic color terms maintains that the world's languages share all or part of

a common stock of color concepts and that terms for these concepts evolve in a constrained order.

Basic Color Terms

In 1969 Brent Berlin and Paul Kay advanced a theory of cross-cultural color concepts centered on the notion of a *basic color term* [1]. A basic color term (BCT) is a color word that is applicable to a wide class of objects (unlike *blonde*), is monolexicemic (unlike *light blue*), and is reliably used by most native speakers (unlike *chartreuse*). The languages of modern industrial societies have thousands of color words, but only a very slender stock of basic color terms. English has 11: *red, yellow, green, blue, black, white, gray, orange, brown, pink, and purple*. Slavic languages have 12, with separate basic terms for light blue and dark blue.

In unwritten and tribal languages the number of BCTs can be substantially smaller, perhaps as few as two or three, with denotations that span much larger regions of color space than the BCT denotations of major modern languages. Furthermore, reconstructions of the earlier vocabularies of modern languages show that they gain BCTs over time. These typically begin as terms referring to a narrow range of objects and properties, many of them noncolor properties, such as succulence, and gradually take on a more general and abstract meaning, with a pure color sense.

The development of English BCTs is typical [2]. The terms *red, black, white, gray, red, green, yellow, and brown* have roots that go back to Old English and before. All of them except for *red* had a variety of non-hue senses, taking on predominately hue sense during the Middle English period. *Purple* and *orange* were borrowed from Latin and Arabic with a hue sense from the beginning, and *pink* emerged sometime in the sixteenth century.

The Original Berlin-Kay Theory

These considerations invite two questions. Are there similarities in the ways that different

languages with the same number of BCTs carve up color space? Is the order in which languages acquire BCTs constrained? In their influential 1969 *Basic Color Terms* [1], Berlin and Kay answered both questions in the affirmative and proposed a sequence in which the number of BCTs in a language predicts what those terms would be. According to the scheme, a Stage I language will contain terms for black and white. In Stage II, these are joined by a word for red. In Stages III and IV, either a word for green appears, followed by a word for yellow, or else a word for yellow appears first, followed by a word for yellow. A word for blue arises in Stage V, followed by a word for brown in Stage VI. Thereafter, in Stage VII, words for pink, purple, orange, and gray appear in no particular order.

The World Color Survey and the Revised Berlin-Kay Theory

The 1969 sequence has been enshrined in many textbooks, but it has since been significantly modified and conceptually refined. The authors and their collaborators improved their methodology and greatly extended the scope of their samples with the 1976 *World Color Survey* (WCS) [3] published with analysis and interpretation in 2009, along with work by others, notably MacLaury's 1997 *Color and Cognition in Mesoamerica* [4]. For the WCS, Protestant missionaries from the Summer Institute of Linguistics collected data from 25 monolingual speakers of each of 110 unwritten minor and tribal languages from 45 different language families. Respondents were asked to name 320 Munsell color chips of maximum chroma and 8 value levels in addition to 10 achromatic chips, presented in random order. After two such naming sessions, they were presented with a miniature version of all 330 samples arrayed in standard order and asked to pick out focal (best) examples of each of the named colors. The resulting dataset [5] permitted much more sophisticated statistical analysis, which broadly – but not completely – confirmed Berlin and Kay's original conclusions. However, in the years between the appearance of *Basic Color*

Terms and the publication of the World Color Survey, the original sequence was replaced by a substantially more complex developmental scheme, as well as a reconceptualization of what that development consisted in.

The original sequence was understood as a succession of encodings of basic colors in which a language would acquire a term for red, and a bit later yellow or green, then blue and so on. It soon became clear that this wouldn't do, since in languages with fewer BCTs, the reference for each color term would encompass a correspondingly larger region of color space – a megacategory – so the addition of each new BCT should be understood as breaking up a megacategory into smaller categories. For instance, in a Stage I language with just two color categories, referring to one category as “white” and to the other as “black” is misleading, since the native category names extend across a broad range of hue samples. The more accurate gloss would be “warm/light” for the “white” category and “dark/cool” for the “black” category. In WCS notation, Stage I is represented as [W/R/Y, Bk/G/Bu]. In Stage II, the first megacategory breaks up, yielding [W, R/Y, Bk/G/Bu], and so on, culminating in Stage V, consisting of [W, R, Y, G, Bu, Bk]. The remaining BCTs appear in no fixed order.

A well-established basic megacategory in many North American and Mesoamerican Indian as well as some African languages is a “grue” term that covers the region that English speakers would separate into a blue and a green category. English, in turn, has its own megacategory. Modern Russian speakers break up our blue category with two BCTs: *goluboi*, light blue, and *sinji*, dark blue.

Constraints on Color-Category Formation

A fundamental question now arises. If the development of basic color categories is constrained in various ways, what are the bases of these constraints? One striking feature of the sequence is that as megacategories break up, categories whose foci lie close to the six elementary Hering colors get named before categories focused near the

binary colors such as orange. Furthermore, the focal choices for each of these categories tend to cluster around average unique hue choices. Studies have found the variability of focal choices among the speakers of a given language to be substantially greater than the variability of average focal choices across languages with a comparable number of BCTs. This suggests that the elementary colors are perceptually salient for all language users, making them more likely to be named [6].

Alternative Explanations

An alternative view, first advanced by Jameson and D'Andrade [7], starts with the observation that the outer surface of the Munsell solid that was used in the collection of the WCS data is lumpy. Regions of high chroma are “mountains” of greater perceptual prominence than “valleys” of lower chroma in which the samples are less readily discriminable from their neighbors. This salience makes them more readily nameable.

Kay and Regier [8] constructed a CIELAB representation of the WCS Munsell samples and used them to define a dispersion metric. This enabled them to determine individual and group centroids for every category of each WCS language. They found a strong tendency of category centroids to cluster across languages. Subsequently, they used the same CIELAB representation to formalize and test the ideas of Jameson and D'Andrade [9]. Defining similarity and dissimilarity measures, they derived a measure of the “well formedness” of a language. By this they mean the extent to which the language maximizes the perceptual similarity of colors within a category and minimizes it across categories. Rotating the data for each of the WCS languages around the lightness axis showed that in most cases the well formedness of the language was maximized at or very near 0° of rotation. They concluded that the irregularities in perceptual color space along with general principles of category formation could account for the organization of the WCS languages. Well formedness is not overly sensitive to the placement of category boundaries,

according to this account, and permits adjustment by linguistic convention.

Other explanations for the Berlin-Kay structure take different approaches. Vantage Theory, offered by Robert MacLaury, is psychological, giving a role to individual perceptual-cognitive differences. By contrast, a theory advanced by Sergej Yendrikhovskij [10] maintains that the structure of color-category systems originates in the statistics of the natural environment. He developed a clustering model that predicted the location, number and order of color categories and tested the model with a database of 630 natural images. The clustering was consistent with the 11 basic color categories of Berlin and Kay. These theories of universal constraints are not necessarily incompatible, since they focus on different aspects of color categorization.

Further Questions

Many questions concerning the evolution and structure of color categories remain to be answered. Why do red terms seem to be the first hue words to appear in color lexicons? Why do grue categories persist for so long in color-category evolution? How are we to understand the appearance of categories in languages from lands as far apart as New Guinea and North America that lump yellow with green instead of with red?

It is very likely that social factors such as trade and conquest drive the evolution of color vocabularies. It seems clear that culture plays a significant role in both the origins and the boundaries of color categories. Exactly what is that role? Can culture override perceptually based constraints? It appears that a proper understanding of even the denotations – let alone the connotations – of a language's color terms requires a proper grasp of the relative contributions of biological, cultural, and environmental factors [11].

Cross-References

- [Ancient Color Categories](#)
- [Color Categorical Perception](#)

- [Comparative Color Categories](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Effect of Color Terms on Color Perception](#)
- [Infant Color Categories](#)
- [Unique Hues](#)
- [Vantage Theory of Color](#)
- [World Color Survey](#)

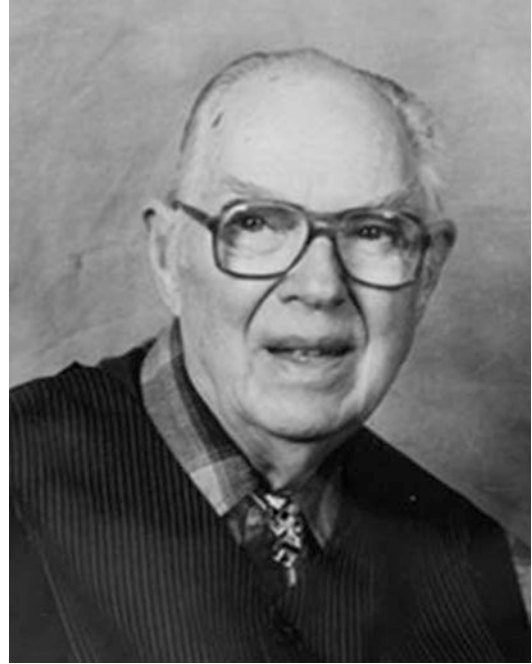
References

1. Berlin, B., Kay, P.: Basic Color Terms. University of California Press, Berkeley (1969)
2. Casson, R.W.: Color shift: evolution of English color terms from brightness to hue. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 224–239. Cambridge University Press, Cambridge (1997)
3. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009)
4. MacLaury, R.: *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. University of Texas Press, Austin (1997)
5. <http://www.icsi.berkeley.edu/wcs/data.html>
6. Webster, M.A., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R.L., Parmei, G., Dedrick, D. (eds.) *The Anthropology of Color*. John Benjamin, Amsterdam (2007)
7. Jameson, J., D'Andrade, R.G.: It's not really red, green, yellow, blue: an inquiry into perceptual color space. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 224–239. Cambridge University Press, Cambridge, UK (1997)
8. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**(15), 9085–9089 (2003)
9. Regier, T., Kay, P., Khetarpal, N.: Color naming reflects optimal partitions of color space. *Proc. Natl. Acad. Sci. U. S. A.* **104**(4), 1436–1441 (2007)
10. Yendrikhovskij, S.N.: A computational model of color categorization. *Color Res. Appl.* **26**, 235–238 (2001)
11. Belpaeme, T., Bleys, J.: The impact of statistical distribution of colours on colour category acquisition. *J. Cogn. Sci.* **10**(1), 1–20 (2009)

Billmeyer, Fred Wallace, Jr.

Ellen C. Carter

Color Research and Application and, Konica Minolta, Pennsville, NJ, USA



Fred Wallace Billmeyer, Jr., was an American chemist who contributed through industrial research and academia to the developing fields of polymer chemistry and color science during the second half of the twentieth century.

Billmeyer was born on August 24, 1919, in Chattanooga, Tennessee. In 1941, he received a B.S. in chemistry from the California Institute of Technology. Then, he received a Ph. D. in physical chemistry from Cornell University in 1945 after studying the measurement of molecular weights by light scattering under Nobel Laureate Peter Debye.

His professional career can be divided into two major parts, focusing first on industry and then on education. Upon graduation from Cornell, he joined the Plastics Department of E. I. du Pont de Nemours and Company, Wilmington, Delaware, where he remained for 20 years. There he

Bezold–Brücke Effect

- [Assimilation](#)

worked on various aspects of polymers. Instinctively an educator, he published *Textbook of Polymer Chemistry* (1957) and *Textbook of Polymer Science* (1962), which went through three editions and was also published in Japanese. These books helped to create the academic discipline of polymer science. He served as a visiting professor in chemical engineering at Massachusetts Institute of Technology for the 1960–1961 academic year. Billmeyer increasingly became involved with the growing field of color science. At the University of Delaware, he supervised his first graduate color science student, Joseph Atkins, whose 1964 Ph.D. thesis was on the Absorption and Scattering of Light in Turbid Media.

In 1964, he retired from DuPont to join academia full time as a professor of analytical chemistry at Rensselaer Polytechnic Institute (RPI), Troy, New York. At RPI, he taught and directed research in both polymer and color science and directed the Rensselaer Color Measurement Laboratory until he became professor emeritus in 1984. During his time at RPI, he not only taught and mentored undergraduate and graduate students but also, with Max Saltzman, initiated summer courses about color science for people from industry. His influence on the field of color science was extended throughout the United States by over 1000 people from diverse fields attending these summer programs.

Deepening his commitment to education, he undertook two other major projects while at RPI. In 1966, Professor Billmeyer continued his collaboration with Max Saltzman, publishing *Principles of Color Technology*. Their two editions of this textbook became widely acclaimed throughout the industrial world as the primary introduction to color science. A third edition entitled *Billmeyer and Saltzman's Principles of Color Technology* (published in 2000) by Roy Berns is still widely used. Also with Richard Kelly in 1975, he published *Entering Industry: A Guide for Young Professionals*.

The second project was the creation of an academic journal for color science. After supporting two unsuccessful attempts at journals of color engineering and science, in 1976, he initiated the journal *Color Research and Application*. Billmeyer

was the founding editor and served in that position until 1986. His approach of involving national color organizations, beginning with the Canadian Society for Color, the US Inter-Society Color Council, and the Colour Group (Great Britain) and later others throughout the world, has contributed to the success of this English-language journal published by John Wiley & Sons now in its 40th year of continuous publication.

Many of Billmeyer's scientific contributions throughout the years fall into the category of light scattering and the application of turbid medium theory to the formulation and shading of colored materials. Beginning with the basic theory of Beer's law, he programmed a digital computer to calculate the concentrations required to produce a given transparent color by mixing soluble dyes [1]. Next with Beasley and Atkins, he expanded two-flux Kubelka-Munk theory to a four-flux turbid medium theory [2], which became known as the BAB theory for colorant formulation. From there, he continued the development of the BAB to apply particularly to paint systems [3] and expanded that further to coatings with goniometric characteristics [4, 5]. He further developed practices for determining the optical properties needed for use in the turbid medium theory and multi-flux theory [6–9].

Billmeyer was a versatile and prolific communicator; his contributions were not limited to turbid medium theory. Two other areas that should be mentioned here are the characterization of fluorescent materials [10, 11] and use of color difference equations in manufacturing [12]. His involvement within the many professional organizations and collaboration with other scientists throughout the world influenced most areas of color science. In particular, his more than 350 published articles discussed polymer characterization and diverse aspects of color. His 13 books presented in clear, concise language the scientific principles necessary for the industrial use of polymers and color to enhance our lives.

Major Books

Billmeyer, F. W.: *Textbook of Polymer Chemistry*. Interscience, New York (1957) (in Russian 1958).

Billmeyer, F. W.: Textbook of Polymer Science, 3 edn. Wiley, New York (1962, 1971, 1884) (in Japanese 1969 & 1989; in Malaysian 1989).

Billmeyer, F. W., Saltzman, M.: Principles of Color Technology, 2 edn. Wiley, New York (1966, 1981) (in German Grundlagen der Farbtechnologie. Goettingen: Zeurich: Muster-Schmidt Verlag, 1993).

Billmeyer, F. W.: Synthetic Polymers. Double Day, Garden City (1972).

Collins, E. A., Bares, J., Billmeyer, F. W.: Experiments in Polymer Science. Wiley-Interscience, New York (1973).

Billmeyer, F. W., Kelley, R. N.: Entering Industry: A Guide for Young Professionals. Wiley, New York (1975).

References

1. Billmeyer, F.W., Beasley, J.K., Sheldon, J.A.: Formulation of transparent colors with a digital computer. *J. Opt. Soc. Am.* **50**, 70–72 (1960)
2. Beasley, J. K., Atkins, J. T., Billmeyer, F. W.: Scattering and absorption of light in turbid media electromagnetic scattering. In: Proceedings of the 2nd Interdisciplinary Conference on Electromagnetic Scattering, University of Massachusetts, Amherst, June 1965, pp. 765–785
3. Billmeyer, F.W., Abrams, R.L.: Predicting reflectance and color of paint films by Kubelka-Munk analysis I – turbid medium theory. *JPT* **45**(579), 23–38 (1973)
4. Billmeyer, F.W., Davidson, J.G.: Color and appearance of metalized paint films I. Characterization. *J. Paint Tech.* **46**(593), 31–37 (1974)
5. Billmeyer, F.W., Carter, E.C.: Color and appearance of metalized paint films II. Initial application of turbid-medium theory. *J. Paint Tech.* **48**(613), 53–60 (1976)
6. Billmeyer, F.W., Chassaing, P.G., Dubois, J.F.: Determining pigment optical properties for use in the mie and many-flux theories. *Color Res. Appl.* **5**, 108–112 (1980)
7. Billmeyer, F.W., Phillips, D.G.: Predicting reflectance and color of paint films by Kubelka-Munk analysis III effect of concentration errors on color for mixtures of one chromatic pigment with white. *J. Paint Tech.* **46**(592), 26–29 (1974)
8. Billmeyer, F.W., Phillips, D.G.: Predicting reflectance and color of paint films by Kubelka-Munk analysis IV Kubelka Munk scattering coefficient. *J. Paint Tech.* **48**(616), 30–36 (1976)
9. Billmeyer, F.W., Richards, L.W.: Scattering and absorption of radiation by lighting materials. *J. Color Appearance* **2**, 4–15 (1973)
10. Alman, D.H., Billmeyer, F.W.: Integrating-sphere errors in the colorimetry of fluorescent materials. *Color Res. Appl.* **1**, 114–145 (1976)
11. Alman, D.H., Billmeyer, F.W.: New method of colorimetric measurement of opaque fluorescent samples. *Color Res. Appl.* **2**, 19–25 (1977)
12. Morley, D. I., Munn Rich, R., Billmeyer, F. W. Small and moderate colour differences parts I and II. *J. Soc. Dyers Colour.* **91**, 229–242; 302–307 (1975)

Binocular Color Matching

David R. Simmons

School of Psychology, University of Glasgow,
Glasgow, Scotland, UK

Synonyms

[Binocular color perception](#); [Binocular color vision](#)

Definition

In the context of this article, “binocular color matching” refers to color vision with two eyes, focusing on the differences between when two eyes are used to view colors (i.e., binocular or dichoptic color) and when only one is used (i.e., monocular color), as well as phenomena (largely) unique to binocular vision such as fusion, rivalry, and stereopsis.

Introduction

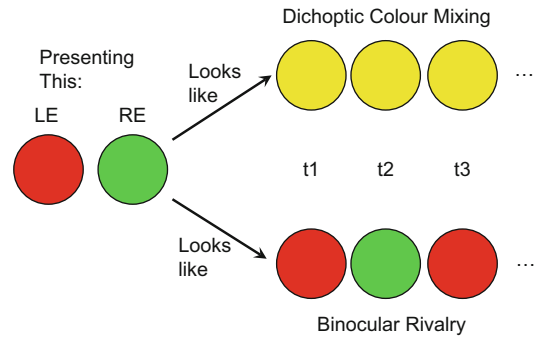
Most of us have two functioning eyes, and yet we still somehow see the visual world as a single entity, most of the time unaware that our brains are receiving and processing two separate and largely independent sets of visual signals from the outside world. In the context of color vision, this raises some interesting questions:

1. *Binocular color fusion*: How similar do the chromatic properties of visual signals in each eye have to be in order to be perceived as one?
2. *Binocular color rivalry*: How different do the chromatic properties of visual signals in each eye have to be in order to evoke binocular rivalry, which is an alternation in visibility of the visual information going to each eye?
3. *Color and stereopsis*: Stereopsis is the perception of depth that is obtained by analyzing the differences between the images going to each eye. To what extent is this perception of depth affected by the chromatic properties of visual information?
4. *Binocular color summation*: The processing of visual signals is affected by whether or not a scene is viewed with one or two eyes: to what extent do the chromatic properties of visual signals influence this?
5. *Binocular color appearance*: Do colors look different when viewed with one eye or two?

Note that the field of binocular vision and stereopsis has been thoroughly reviewed by Prof. Brian Rogers of the University of Oxford, UK, and Prof. Ian Howard, late of York University, Canada [11], and for all but the most recent studies, the interested reader should refer to them for further information.

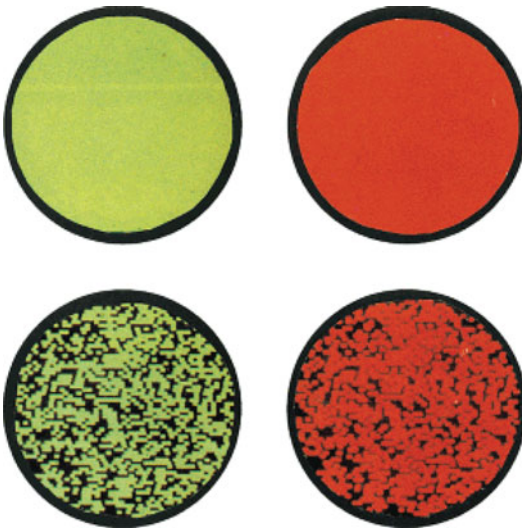
Binocular Color Fusion

The problem of binocular fusion, how the world is seen as one despite having two eyes, has fascinated philosophers and scientists for centuries (see [11] for a historical overview). It makes sense that color might have a role to play in this phenomenon. One might expect a feature in the visual world when viewed with one eye to have the same spectral reflectance properties and therefore presumably appear to be the same color, when viewed with the other eye. Consequently, interocular color similarity would presumably make a good visual cue for fusion. Curiously, however, most of the research effort has gone into investigating what happens when the colors presented to the two eyes are *different*. In



Binocular Color Matching, Fig. 1 Schematic diagram illustrating alternative percepts generated by dichoptic color mixing or binocular rivalry

particular, there is the controversial issue of *dichoptic color mixture*: when two lights with different chromatic contents are presented separately to each eye, do the lights combine in a similar way to when they are presented superimposed to the same eye or does *binocular rivalry* (an alternation in perception between the stimuli presented to each eye) result (see Fig. 1)? The first investigation of this phenomenon appears to have been in the early eighteenth century using differently colored silks viewed through an aperture [7]. Later, in the nineteenth century, the existence or not of dichoptic color mixture was a further source of controversy between Helmholtz and Hering, given the different predictions of the Young-Helmholtz and Hering theories of color vision. Arguably the most controversial aspect is whether or not “binocular yellow” can be obtained when red light is presented to one eye and green to the other. Recent studies have clarified that true dichoptic color mixture does appear to take place, so long as certain conditions are met. These are that the mixture is more stable with small and textured patches of light rather than large and homogeneous, with flickering rather than steady stimuli, and with patches of low luminance and saturation which are close in luminance and chromaticity rather than high luminance and saturation with different luminances and chromaticity ([11]; see also [10]). Some of these rules are nicely illustrated in a study by de Weert and Wade [5] (see Fig. 2). In a more recent study, the bold claim is



Binocular Color Matching, Fig. 2 Dichoptic color mixing: fusion of the *upper* two disks gives unstable binocular rivalry; fusion of the *lower* two textured disks produces stable dichoptic color mixing. This illustrates the role of texture in promoting dichoptic color mixing (Taken from de Weert and Wade [5], Reprinted from Vision Research, 28, Charles M.M. de Weert, Nicholas J. Wade, Compound binocular rivalry, 1031–1040, Copyright (1988), with permission from Elsevier)

made that “two eyes are worse than one,” based on data suggesting that monocularly visible features of different colors effectively disappear when dichoptically combined [1]. In an earlier study by Simmons [32], however, although detection thresholds for briefly presented isoluminant red-green gratings in antiphase between the eyes were higher than those for monocular presentation, they were not so high that the signals from each eye were effectively canceling each other out, as suggested by Anstis and Rogers [1]. Potentially, technical issues to do with the accurate registration of stimuli in each eye and the presence of barely detectable luminance artifacts complicate the interpretation of these experiments.

A different approach was taken by Malkoc and Kingdom [21], who, following on from an earlier study by Yoonessi and Kingdom [42], measured *dichoptic color difference thresholds (DCDTs)*, which are the thresholds for detecting a color difference between the two eyes, coinciding with the detection of a peculiar phenomenon called *binocular luster*. They found that these thresholds

were higher than those for detecting color differences between two stimuli when they were presented side by side, although lower than those required to provoke binocular rivalry (see below). Malkoc and Kingdom [21] also found that these thresholds were best predicted by the perceived color difference between the two stimuli, rather than any considerations based on cardinal or unique hue mechanisms. Kingdom and Libenson [14] specifically investigated the processing of interocular differences in saturation (or color contrast). They found that the appearance of the mixture obtained crucially depended on the relative amounts of luminance and chromatic contrast. With purely chromatic differences between the two eyes (i.e., lights with the same hue but different saturations), the more saturated/higher contrast stimulus dominated the percept, but the presence of a luminance pedestal forced the colors to blend and therefore reduced the saturation of the resultant. Kingdom and Libenson [14] argue from these results that the appearance of a dichoptic color mixture depends on whether or not the brain interprets the information from the two eyes as coming from the same object or not, a phenomenon which they term the “object commonality hypothesis.”

Binocular Color Rivalry

As rivalry is, to a certain extent, the obverse of fusion, most of the issues pertinent to this theme have already been discussed above. It is certainly true that the presentation of saturated red stimuli to one eye and saturated green to the other is often almost paradigmatic in studies where the aim is to evoke binocular rivalry (see, e.g., [2]; also, again, [5]). O’Shea and Williams [23] demonstrated that S-cone-isolating stimuli could induce binocular rivalry, suggesting that rivalry was not solely confined to luminance or red-green chromatic pathways. A detailed study of the wavelength sensitivity of binocular rivalry was performed by Sagawa [28]. Note that rivalry does not occur for briefly presented stimuli. In these situations the dichoptic stimuli tend to superimpose, although are still distinguishable from

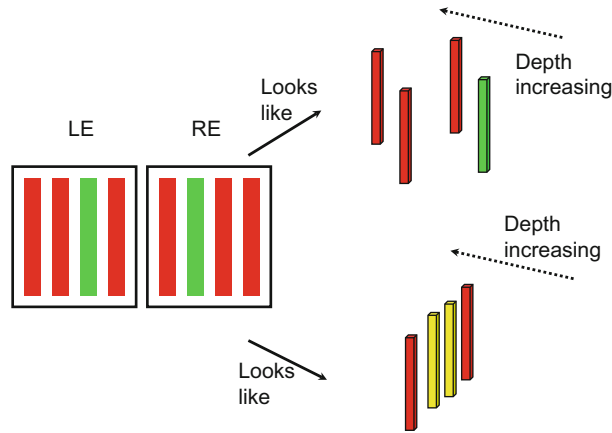
monocularly superimposed stimuli [11]. There is some evidence that the chromatic system is more affected by binocular rivalry suppression than the achromatic system [11, 24]. Mullen et al. [22] found that the visibility of chromatic grating stimuli presented to one eye was affected by the presence of luminance stimuli in the other, suggesting that, when stimuli differ between the two eyes, the suppression of one eye by the other is independent of whether the stimulus contains color or luminance contrast (which is not the case for monocular vision, or when the stimulus is the same in both eyes, when the suppression of one stimulus by another, masking, stimulus is more selective).

Color and Stereopsis

There are two issues which have dominated research on color and stereopsis. The first is whether or not there is actually a functioning chromatic stereopsis mechanism. The second is whether or not chromatic information helps to solve the stereo correspondence problem. In both cases the arguments have been similar to those in color motion perception. Historically, a good way of testing whether or not a purely chromatic mechanism exists has been to test performance at *isoluminance*, when the visual patterns in question are theoretically defined solely by chromatic contrast, without any luminance contrast being present. The earliest study of stereopsis at isoluminance suggested that stereoscopic depth perception was very weak and possibly nonexistent at red-green isoluminance [20]. Subsequent studies provided conflicting results, depending on the precise stimulus characteristics [11]. Around the early 1990s, the extreme views on this issue were represented by two studies: Livingstone and Hubel [19] and Scharff and Geisler [29]. Livingstone and Hubel [19] claimed that all previous demonstrations of intact stereoscopic depth perception were artifactual, and down to the technical difficulties associated with removing all luminance information from the stimulus. Their view that stereopsis was essentially “color blind” was consistent with their theories on the parallel processing of visual information by

magnocellular- and parvocellular-mediated visual pathways, with stereopsis mediated by the “color-blind” magnocellular stream. Scharff and Geisler [29], on the other hand, claimed that their careful calibration of the stimulus, accounting for all potential luminance artifacts, not only demonstrated that stereopsis was possible at isoluminance but that it was as good as it could be, given the limits on obtainable color contrast due to the overlap in spectral sensitivities of long- and medium-wavelength-sensitive cones. In a series of studies published between 1994 and 2002, Simmons and Kingdom updated this view of the status of stereopsis at isoluminance to demonstrate that there exists a rudimentary chromatic stereopsis mechanism which is less contrast sensitive and has a more limited disparity range, poorer stereoacuity, and poorer ability to encode a stereoscopically defined shape than its luminance counterpart ([15, 17, 33–35]; reviewed in [16]). While this view was challenged in a study by Krauskopf and Forte [43], who argued, similarly to Livingstone and Hubel [19], for a complete disappearance of stereopsis at isoluminance, their data could be partially explained by the presence of high spatial frequency luminance artifacts in their stimuli. Having said that, it is very difficult to control for these artifacts, especially given the high contrast sensitivity of luminance-based stereopsis [31, 37]. While most of the above studies were carried out at red-green isoluminance, Grinberg and Williams [9] demonstrated that stereopsis is also possible at “blue-yellow” isoluminance.

Can chromatic information help to solve the stereo correspondence problem? The correspondence problem in stereopsis refers to the ability of the human brain to work out which feature in one eye’s view matches with which feature in the other eye’s view (see Fig. 3). As stated above, it makes sense that chromatic information could help in solving this problem, and, sure enough, a number of studies have demonstrated that this is indeed the case (e.g., [6, 12, 18]). Simmons and Kingdom [36] argued that these results might be explicable in terms of the interactions between different independent stereopsis mechanisms, some of which are sensitive to chromatic contrast (see also [38]). Wardle and Gillam [41] have argued that



Binocular Color Matching, Fig. 3 Schematic illustration of the potential role of color in the stereo correspondence problem. If the images illustrated are presented to the left and right eyes, respectively, if color promotes stereo correspondence, then a *green bar* will appear in front and a

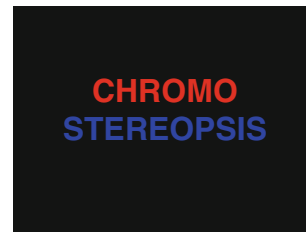
red bar behind the fixation plane (*upper right*). If color has no role, then four bars will be seen in the frontoparallel plane (*lower right*). The bars colored *yellow* might also appear rivalrous, depending on the precise stimulus conditions

color information is also important in the so-called da Vinci stereopsis, which is the perception of depth obtained, under certain viewing conditions, from visual regions only visible to one eye.

There are some complicated dependences of stereoscopic sensitivity on stimulus dynamics at isoluminance which were investigated by Tyler and Cavanagh [39]. Note that the phenomenon of *chromostereopsis*, most often experienced as a saturated red target appearing in front of a saturated blue one (when viewed with two eyes) despite the stimuli being physically in the same depth plane, is thought to be largely due to chromatic aberrations in the eye inducing a relative difference (i.e., disparity) in the retinal locations of the red and blue objects in each eye [11, 25]. Figure 4, especially if presented on a data projector and viewed in a dark room, normally induces a strong sensation of this illusory depth percept.

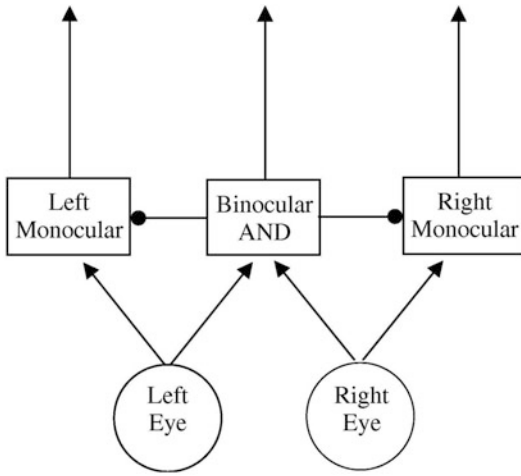
Binocular Color Summation

To what extent does viewing a chromatic stimulus with two eyes affect its visibility or the performance obtained with it? The scientific literature presents a somewhat confusing picture, with Anstis and Rogers [1] claiming that “two eyes



Binocular Color Matching, Fig. 4 Stimulus for chromostereopsis. If viewed on the full screen (especially if presented on a data projector), most people will see the *red* “CHROMO” look slightly in front of the *blue* “STEREOPSIS”

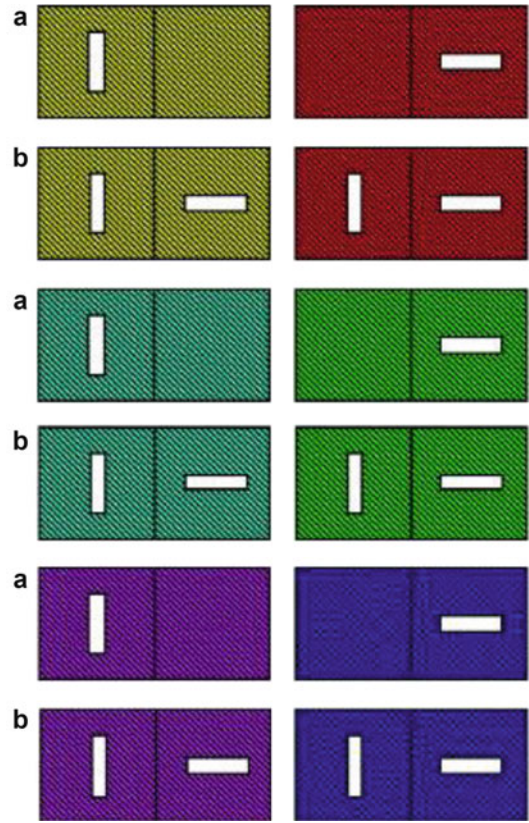
are worse than one,” Tyler and Cavanagh [39] claiming that “two eyes” are “as sensitive as one,” and Simmons [32] finding evidence for binocular contrast summation close to “full summation,” when contrast is effectively linearly added between the eyes, and so a binocular stimulus is twice as detectable as when presented to both eyes rather than one (this would be “two eyes are twice as good as one”!). The conditions under which dichoptic color mixing is obtained have already been outlined above. It therefore suffices to say that the extent of binocular summation for a given chromatic stimulus will depend on the task and the precise chromatic and spatiotemporal parameters of the stimulus.



Binocular Color Matching, Fig. 5 Model from Shimono et al. [30] which illustrates the inhibitory action of their proposed “purely chromatic” binocular mechanism on monocular mechanisms (Taken from Shimono et al. [30], Reprinted from Vision Research, 49, Shimono, K., Shioiri, S., Yaguchi, H., Psychophysical evidence for a purely binocular color system, 202–210, Copyright (2009), with permission from Elsevier)

Binocular Color Appearance

Those of us who have typical binocular vision will have a region of binocular overlap in the middle of the visual field flanked on either side by regions that are only visible in one or the other eye. However, we are usually not aware that colors change in the center versus the periphery of our visual field, although changes are detectable under controlled laboratory conditions [3, 26]. A whole range of dichoptic mechanisms have been proposed to account for these apparent compensations between the two eyes. Shimono et al. [30] suggested a model which takes into account previous results on interocular transfer of chromatic adaptation and chromatic induction effects (e.g., [4, 8, 40]) by proposing a “purely binocular” chromatic mechanism, which responds only to simultaneous binocular presentation of similar colors to each eye but which also inhibits the responses of “purely monocular” color mechanisms (see Fig. 5). This model would also seem to be reasonable in the light of physiological data collected from monkey visual cortex



Binocular Color Matching, Fig. 6 Binocular color induction in monocular patches. Sets of multicolored stereograms are shown for three different color combinations. (a) The white, monocular patches gradually fill with distinct pastel colors during binocular viewing of the stereograms. The buildup of the effect typically takes a few seconds. Dependent on the viewer, the full development of the colors may even take minutes. The color effects are most compelling when the stereograms are viewed printed out and in bright light (see original publication). When one eye is closed, the illusion gradually becomes weaker and disappears. This buildup and decay of the illusion after opening or closing one eye indicates that the color appearances are induced by a binocular mechanism. (b) During binocular viewing, the white, binocular patches do not change color. During monocular viewing after a period of binocular viewing, however, induced colors are apparent in the patches (Taken from Erkelens and Van Ee [8], Reprinted from Vision Research, 42, Erkelens, C.J., Van Ee, R., Multi-coloured stereograms unveil two binocular colour mechanisms in human vision, 1103–1112, Copyright (2002), with permission from Elsevier)

[27]. However, more recent results, such as those of Kingdom and Libenson [14] and Mullen et al. [22] (discussed above), complicate this simple picture.

Another aspect of binocular color appearance that has recently received attention is how presenting different colors to the two eyes can be an easy way to simulate surface glossiness in stereoscopic visual displays, without the need to create a detailed reflectance model [13].

Conclusion

The field of binocular color vision has proven surprisingly controversial over the years, with the existence of true dichoptic color mixing, the contribution of color to stereopsis, and the presence or not of a purely binocular color mechanism being key battlegrounds. The current consensus of the field appears to be that dichoptic color mixing can happen, that chromatic mechanisms contribute to stereopsis, and that there is a purely binocular color mechanism, but demonstrating these things requires careful stimulus control and experimental technique. Added to this there are some curious phenomena such as being able to make stimuli which are easily detectable when viewed monocularly disappear when viewed binocularly [1] and the interocular transfer of chromatic induction [8] (Fig. 6). It seems that we still have some more work to do before we can say that we fully understand the complex interactions between binocular color mechanisms.

Cross-References

- [Chromostereopsis](#)
- [Color Processing, Cortical](#)
- [Color Vision, Opponent Theory](#)
- [Cone Fundamentals](#)
- [Ganglion Cells](#)
- [Helmholtz, Hermann Ludwig von](#)
- [Hering, Karl Ewald Konstantin](#)
- [Magno-, Parvo-, Koniocellular Pathways](#)
- [Simultaneous Color Contrast](#)
- [Stereo and Anaglyph Images](#)

References

1. Anstis, S., Rogers, B.: Binocular fusion of luminance, color, motion and flicker – two eyes are worse than one. *Vision Res.* **53**, 47–53 (2012)
2. Bhardwaj, R., O’Shea, R.P.: Temporal analysis of image-rivalry suppression. *PLoS One* **7**(9), e45407 (2012). doi:10.1371/journal.pone.0045407
3. Bompas, A., Powell, G., Sumner, P.: Systematic biases in adult color perception persist despite lifelong information sufficient to calibrate them. *J. Vis.* **13**, 1–19 (2013)
4. Coltheart, M.: Color-specificity and monocularity in the visual cortex. *Vision Res.* **3**, 2595–2598 (1973)
5. De Weert, C.M.M., Wade, N.J.: Compound binocular rivalry. *Vision Res.* **28**, 1031–1040 (1988)
6. Den Ouden, H.E.M., Van Ee, R.R., de Haan, E.H.E.: Colour helps to solve the binocular matching problem. *J. Physiol.* **567**, 665–671 (2005)
7. Desagulier, J.T.: A plain and easy experiment to confirm Sir Isaac Newton’s doctrine of the different refrangibility of the rays of light. *Philos. Trans. R. Soc. Lond. A* **29**, 448–452 (1716)
8. Erkelens, C.J., Van Ee, R.: Multi-coloured stereograms unveil two binocular colour mechanisms in human vision. *Vision Res.* **42**, 1103–1112 (2002)
9. Grinberg, D.L., Williams, D.R.: Stereopsis with chromatic signals from the blue-sensitive mechanism. *Vision Res.* **25**, 531–537 (1985)
10. Hovis, J.K.: Review of dichoptic color mixing. *Optom. Vis. Sci.* **66**, 181–190 (1989)
11. Howard, I.P., Rogers, B.J.: *Perceiving in Depth: volume 2 Stereoscopic Vision*. Oxford University Press, Oxford, UK (2012)
12. Jordan, J.R., Geisler, W.S., Bovik, A.C.: Color as a source of information in the stereo correspondence process. *Vision Res.* **30**, 1955–1970 (1990)
13. Jung, W.-S., Moon, Y.-G., Park, J.-H., Song, J.-K.: Glossiness representation using binocular color difference. *Opt. Lett.* **38**, 2584–2587 (2013)
14. Kingdom, F.A.A., Libenson, L.: Dichoptic color saturation mixture: binocular luminance contrast promotes perceptual averaging. *J. Vis.* **15**(5), Article no. 2. (2015)
15. Kingdom, F.A.A., Simmons, D.R.: Stereoacuity and colour contrast. *Vision Res.* **36**, 1311–1319 (1996)
16. Kingdom, F.A.A., Simmons, D.R.: The relationship between colour vision and stereoscopic depth perception (invited paper). *J. Soc. 3D Broadcast. Imaging* **1**, 10–19 (2000)
17. Kingdom, F.A.A., Simmons, D.R., Rainville, S.R.: On the apparent collapse of stereopsis in random-dot stereograms at isoluminance. *Vision Res.* **39**, 2127–2141 (1999)
18. Kovacs, I., Julesz, B.: Depth, motion, and static-flow perception at metaisoluminant color contrast. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10390–10394 (1992)

19. Livingstone, M.S., Hubel, D.H.: Psychophysical evidence for separate channels for the perception of form, colour, movement, and depth. *J. Neurosci.* **7**, 3416–3468 (1987)
20. Lu, C., Fender, D.H.: The interaction of color and luminance in stereoscopic vision. *Invest. Ophthalmol.* **11**, 482–489 (1972)
21. Malkoc, G., Kingdom, F.A.A.: Dichoptic difference thresholds for chromatic stimuli. *Vision Res.* **62**, 75–83 (2012)
22. Mullen, K., Kim, Y.J., Gheiratmond, M.: Contrast normalization in colour vision: the effect of luminance contrast on colour contrast detection. *Sci. Rep.* **4**, 7350 (2014). doi:10.1038/srep07350
23. O'Shea, R.P., Williams, D.R.: Binocular rivalry with isoluminant stimuli visible only via short-wavelength-sensitive cones. *Vision Res.* **36**, 1561–1571 (1996)
24. Ooi, T.L., Loop, M.S.: Visual suppression and its effect upon color and luminance sensitivity. *Vision Res.* **34**, 2997–3003 (1994)
25. Ozolinsh, M., Muizniece, K.: Color difference threshold of chromostereopsis induced by flat display emission. *Front. Psychol.* **6**, Article: 337 (2015)
26. Parry, N.R.A., Panorgias, A., McKeefry, D.J., Murray, I.J.: Real-world stimuli show perceived hue shifts in the peripheral visual field. *J. Opt. Soc. Am. A* **29**, 96–101 (2012)
27. Peirce, J.W., Solomon, S.G., Forte, J.D., Lennie, P.: Cortical representation of color is binocular. *J. Vis.* **8**, 1–10 (2008)
28. Sagawa, K.: Minimum light intensity required for color rivalry. *Vision Res.* **21**, 1467–1474 (1981)
29. Scharff, L.V., Geisler, W.S.: Stereopsis at isoluminance in the absence of chromatic aberrations. *J. Opt. Soc. Am. A* **9**, 868–876 (1992)
30. Shimono, K., Shioiri, S., Yaguchi, H.: Psychophysical evidence for a purely binocular color system. *Vision Res.* **49**, 202–210 (2009)
31. Simmons, D.R.: The minimum contrast requirements for stereopsis. *Perception* **27**, 1333–1343 (1998)
32. Simmons, D.R.: The binocular combination of chromatic contrast. *Perception* **34**, 1035–1042 (2005)
33. Simmons, D.R., Kingdom, F.A.A.: Contrast thresholds for stereoscopic depth identification with isoluminant and isochromatic stimuli. *Vision Res.* **34**, 2971–2982 (1994)
34. Simmons, D.R., Kingdom, F.A.A.: Differences between stereopsis with isoluminant and isochromatic stimuli. *J. Opt. Soc. Am. A* **12**, 2094–2104 (1995)
35. Simmons, D.R., Kingdom, F.A.A.: On the independence of chromatic and achromatic stereopsis mechanisms. *Vision Res.* **37**, 1271–1280 (1997)
36. Simmons, D.R., Kingdom, F.A.A.: Interactions between chromatic- and luminance-contrast-sensitive stereopsis mechanisms. *Vision Res.* **42**, 1525–1546 (2002)
37. Simmons, D.R.: Stereopsis at red-green isoluminance: chasing the luminance artifacts. [Abstract]. *J. Vis.* **2**(10), 69a (2002). doi:10.1167/2.10.69. <http://journalofvision.org/2/10/69/>
38. Stuart, G.W., Edwards, M., Cook, M.L.: Color inputs to random-dot stereopsis. *Perception* **21**, 717–729 (1992)
39. Tyler, C.W., Cavanagh, P.: Purely chromatic perception of motion in depth: two eyes as sensitive as one. *Percept. Psychophys.* **49**, 53–61 (1991)
40. Vimal, R.L.P., Shevell, S.K.: A central binocular mechanism affects chromatic adaptation. *Vision Res.* **27**, 429–439 (1987)
41. Wardle, S.G., Gillam, B.J.: Color constrains depth in da Vinci stereopsis for camouflage but not occlusion. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1525–1540 (2013)
42. Yoonessi, A., Kingdom, F.A.A.: Dichoptic difference thresholds for uniform color changes applied to natural scenes. *J. Vis.* **9**(2), 1–12 (2009) (article no. 3)
43. Krauskopf, J., Forte, J.D.: Influence of chromaticity on vernier and stereo acuity. *J. Vis.* **2**(9), 645–652 (2002)

Binocular Color Perception

► Binocular Color Matching

Binocular Color Vision

► Binocular Color Matching

Blackbody and Blackbody Radiation

Joanne Zwinkels

National Research Council Canada, Ottawa, ON, Canada

Synonyms

Blackbody (specific terms) – Black body, Full radiator, Planckian radiator, Natural object

Blackbody (alt. terms) – Ideal black surface, Ideal emitter, Ideal thermal radiator

Blackbody radiation – Black-body radiation, Cavity radiation

Blackbody locus – Planckian locus

Definition

A *blackbody* is an ideal thermal radiator that absorbs completely all incident radiation whatever the wavelength, the direction of incidence, or the polarization [1]. This radiator has, for any wavelength and any direction, the maximum spectral concentration of radiance for a thermal radiator in thermal equilibrium at a given temperature.

Blackbody radiation is radiant energy emitted by an ideal black surface (blackbody) whose spectral power distribution is only governed by its own temperature.

Blackbody color is the temperature of a true blackbody emitter which to the human eye is a close match to the color of an incandescent object, which is not a blackbody.

Blackbody locus is the locus of points in a chromaticity diagram that represents chromaticities of the radiation of Planckian radiators at different temperatures.

Description

All bodies at nonzero temperature absorb and emit radiation. If a body is at thermal equilibrium, it is because the amount of radiation that is emitted per second is equal to the amount absorbed. A blackbody (or ideal Planckian or black radiator) is a body that can completely absorb radiation at any wavelength and completely emit radiation at any wavelength so that it emits more radiant power at each wavelength than any other object at the same temperature. A blackbody is not only an ideal emitter but also a diffuse emitter, which means that the emitted radiation is isotropic or independent of direction [2]. It gets its name because of the fact that, at room temperature, it appears visibly black.

In the late nineteenth century, Kirchhoff [3] and other experimental physicists found that an ideal blackbody, in thermodynamic equilibrium, emitted radiant energy whose spectrum was continuous and a function only of its wavelength and

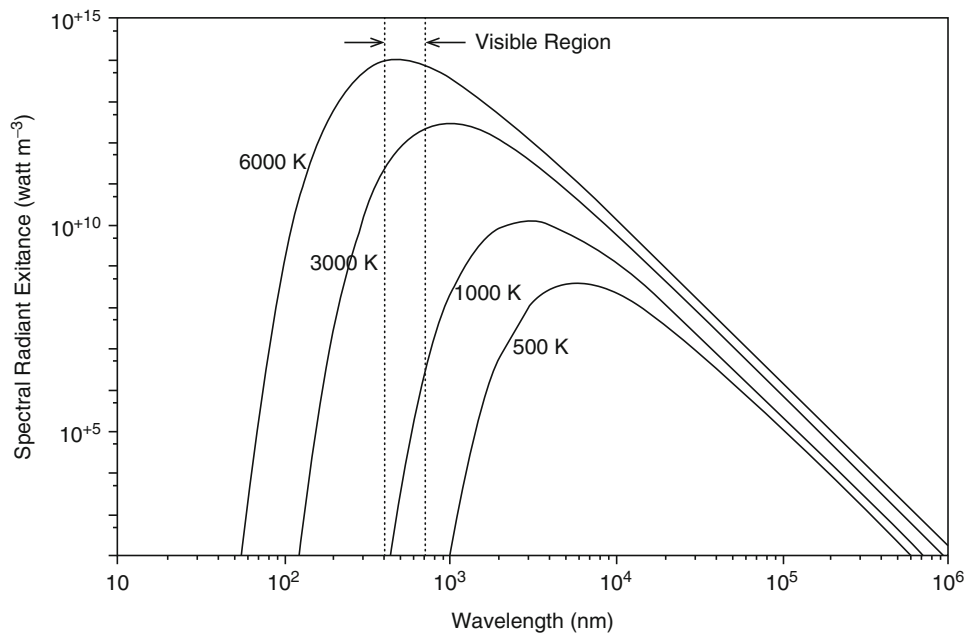
temperature, independent of the nature of the blackbody. However, this behavior could not be explained using thermodynamics or classical statistical mechanics. It was not until Einstein introduced the idea that light could be quantized that Planck formulated his blackbody radiation law based upon this principle [4]. This is generally known as *Planck's law* and gives the amount of power emitted by a blackbody in equilibrium at a temperature T according to:

$$M_e(\lambda, T) = \frac{c_1}{\lambda^5} \left(\frac{1}{e^{c_2/\lambda T} - 1} \right)$$

where $M_e(\lambda, T)$ is the spectral radiant exitance (power per unit area per unit wavelength interval), λ is the wavelength in meters, and T is the absolute temperature in kelvins. The constants c_1 and c_2 are known as Planck's radiation constants and have the values: $c_1 = 2\pi hc^2 = 3.7415 \times 10^{-16} \text{ W m}^2$; $c_2 = hc/k = 1.4388 \times 10^{-2} \text{ m K}$, where h is Planck's constant, c is the speed of light in vacuum, and k is Boltzmann's constant.

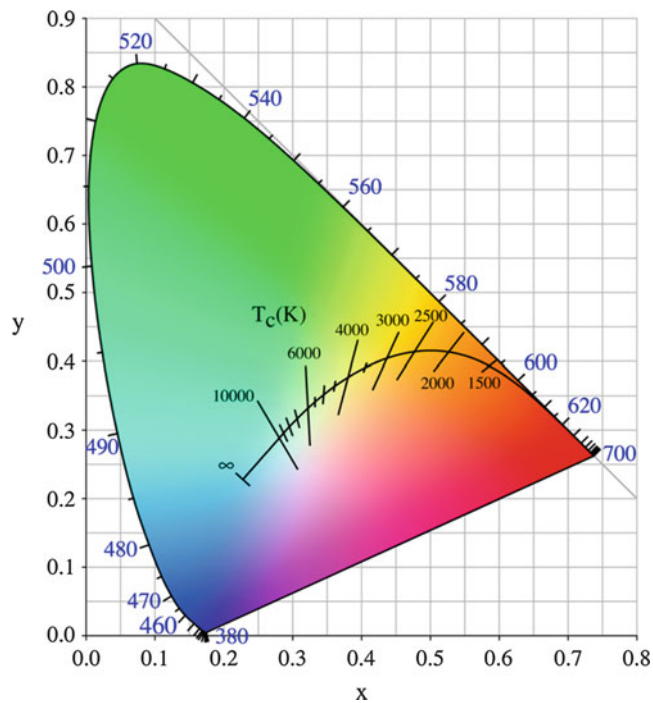
The spectral power distributions of some blackbody radiators at different temperatures are shown in Fig. 1. It can be seen that the peak of the blackbody radiation curve shifts to shorter wavelength and higher energy with increasing temperature. At room temperature (300 K), the peak of the curve is at about 10 μm , and the emission is largely thermal in the infrared region. As discussed above, this room-temperature blackbody that has essentially no emission in the visible region appears black. However, as the temperature of the blackbody source rises through the range shown, it increasingly emits in the visible region, and its color changes continuously from black to red through orange and yellow to white and finally a bluish white.

The color coordinates of these blackbody sources at different temperatures can be plotted on a chromaticity diagram. This is shown in Fig. 2, which shows that as the temperature of the blackbody radiator increases, its CIE chromaticity coordinates (xy) move along a curved line in the CIE 1931 chromaticity diagram. This curve is known as the *blackbody* or *Planckian locus*. For



Blackbody and Blackbody Radiation, Fig. 1 Blackbody spectral exitance curves for several temperatures

Blackbody and Blackbody Radiation, Fig. 2 CIE 1931 chromaticity diagram showing the locus of blackbody radiators calculated for Planckian temperatures from 1500 K to infinity; the Planckian temperature that provides the smallest chromaticity difference between an artificial source and this blackbody locus is its correlated color temperature, T_c



white light sources for general lighting applications, they are designed so that their CIE chromaticities ($u'v'$) are located near this Planckian locus, when plotted on a Uniform Color Scale (UCS) diagram ($u'v'$).

For colorimetric applications, the International Commission on Illumination (CIE) has defined a standard illuminant A, which is representative of illumination under incandescent lighting [5] and whose values of relative spectral power distribution are that of a Planckian radiator at a temperature of approximately 2856 K. It is important to note that while the original CIE standard illuminant A was defined in 1931 in terms of a relative spectral power distribution of a Planckian radiator at a temperature of 2848 K, the Planckian radiation constants, c_1 and c_2 , have changed several times since then (in 1948, 1968, and 1990) with changes in the International Temperature Scale (ITS), causing a corresponding change in the associated temperature. For this reason, the CIE Standard which currently defines CIE standard illuminant A is no longer a function of temperature [5]. However, for an artificial source that is representative of CIE standard illuminant A, the concept of *correlated color temperature* is used, which is the temperature of the Planckian radiator having a chromaticity along the Planckian locus nearest the chromaticity calculated for the spectral distribution of the test source. There are also many artificial white light sources whose spectral power distribution is very different from a blackbody source, such as fluorescent lamps and sources based on light-emitting diodes. However, the concept of correlated color temperature is still used to describe the colorimetric properties of these types of white light sources.

For astronomical applications, the electromagnetic radiation from stars and planets is sometimes characterized in terms of an *effective temperature*, the temperature of a blackbody that would emit the same total flux of electromagnetic energy. However, stars are observed at a depth inside the atmosphere (the photosphere), so it is more typical to describe a star's visual surface by its *photospheric temperature* (see entry on ► [Apparent Magnitude, Astronomy](#)).

There are several other important properties of blackbody sources that are given by the following three laws [6]:

Stefan-Boltzmann Law: This law states that the total radiant exitance, M of a blackbody, determined by integrating the spectral radiant exitance curve, $M_e(\lambda)$, for all wavelengths, λ , is proportional to the fourth power of its absolute temperature:

$$M = \sigma T^4$$

where σ is Stefan's constant ($=0.56686 \times 10^{-7} \text{ Wm}^2\text{K}^{-4}$). No surface at thermal equilibrium can emit more than this amount of energy.

Kirchhoff's Law of Thermal Radiation: This law states that, at each wavelength, the emissivity of the blackbody, $\varepsilon(\lambda)$, at thermal equilibrium is equal to its absorptivity, $\alpha(\lambda)$:

$$\varepsilon(\lambda) = \alpha(\lambda)$$

Wien's Displacement Law: This law states that the curves of the spectral power distribution of the blackbody radiation are peaked at a wavelength, λ_{peak} , given by:

$$\lambda_{\text{peak}} = \frac{2898}{T}$$

This can be simply shown by differentiating the Planckian spectral power distribution with respect to wavelength and solving the derivative when it equals zero.

Examples

Since the last century, blackbody radiation has been of great interest for use as a primary (calculable) source because the properties are universal and independent of the particular material substance. A good laboratory implementation of a blackbody is a small hole in the wall of an otherwise opaque cavity [7]. The light that enters the cavity through the hole will undergo so many reflections before it can escape from the hole that the light will be effectively absorbed and the hole

will appear black. Until 1979, a blackbody like this, heated to the temperature of melting platinum, was the basis of the definition and realization of the SI base unit for photometry, the candela [8].

The earliest types of artificial light were based on heating an object until it was red hot or hotter and was able to emit visible light. These types of artificial light sources are referred to as *incandescent*. Examples of incandescent objects are fire, candles, carbon arcs, and tungsten-filament lamps. They emit radiation with approximately the same spectral power distribution as a blackbody but with an intensity reduced by a factor called the *emissivity*, ε . If the emissivity does not vary with wavelength, the source is called a *gray body* or *nonselective radiator*. An incandescent tungsten-filament lamp is close to being a gray body with an emissivity of approximately 0.40 [7]. The temperature of a blackbody that is closest in color to this gray body source is often used as an index of its color. This is referred to as the *correlated color temperature*, T_c . An alternative index is the *distribution temperature*, which is the temperature of a blackbody whose relative spectral power distribution is closest to that of the source in question. However, all three terms should be used with caution as they do not give a full picture of the actual spectral distribution [9]. Typical tungsten-filament lamps have correlated color temperatures and distribution temperatures in the range 2500–3000 K.

However, these incandescent lamps emit not only visible light but a significant amount of infrared radiation which makes them inefficient as light sources compared with more modern lighting technologies, such as solid-state lighting which can be designed to emit optical radiation only in the visible range.

There are many examples of everyday materials that behave like blackbodies but emit significant thermal radiation. The human body is one example of a blackbody radiator which emits at roughly 300 K. Another obvious example is the Sun, whose surface is roughly 6000 K. The Earth is also considered to be a blackbody radiator, but it is less than ideal.

Cross-References

- [Apparent Magnitude, Astronomy](#)
- [CIE Chromaticity Coordinates \(xyY\)](#)
- [Spectral Power Distribution](#)
- [USC Diagrams; Uniform Chromaticity Scales; Yu'v'](#)

References

1. CIE S017/E:2011, ILV-international lighting vocabulary, CIE Central Bureau, Vienna
2. Mahmoud, M.: §2.1 Blackbody radiation. In: Engineering Thermofluids: Thermodynamics, Fluid Mechanics, and Heat Transfer, p. 568. Springer, Berlin (2005). ISBN 3-540-22292-8
3. Kirchho, G.U.: "ber das Verha"ltniszwischen dem Emissionsverm"ogen und dem Absorptionsverm"ogen der Korper fur Warme und Licht. *Poggendorff Annalen der Physik und Chemie*, 1860, v. 109, 275–301. (English translation by F. Guthrie: Kirchho G. On the relation between the radiating and the absorbing powers of different bodies for light and heat. *Phil. Mag.*, ser. 4, v. 20, 1–21 (1860))
4. Planck, M.: Distribution of energy in the spectrum. *Ann. Phys.* **4**, 553–563 (1901)
5. ISO 11664-2:2007/CIE, S014-2:2006 Colorimetry – Part 2: CIE standard illuminants for colorimetry
6. Williamson, S.J., Cummins, H.Z.: *Light and Color in Nature and Art*. Wiley, New York (1983)
7. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
8. Zwinkels, J.C., Ikonen, E., Fox, N.P., Ulm, G., Rastello, M.L.: Photometry, radiometry and "the candela" evolution in the classical and quantum world. *Metrologia* **47**, R15–R32 (2010)
9. Robertson, A.R.: Computation of correlated color temperature and distribution temperature. *J. Opt. Soc. Am.* **58**, 1528–1535 (1968)

Black-Light Lamps

- [Ultraviolet Radiator](#)

Blending

- [Compositing and Chroma Keying](#)

Blind Spot

Jonathan Aboshiha
The UCL Institute of Ophthalmology and
Moorfields Eye Hospital, London, UK

Synonyms

[Physiologic scotoma](#); [Punctum cecum](#)

Definition

The **blind spot** is the name given to the scotomatous area of each eye's visual field that lacks visual input, due to the photoreceptor-free region of the retina where the optic nerve exits the eye.

Overview

The phenomenon of the blind spot was first described in 1668 by the French Roman Catholic priest and scientist Edme Mariotte, who used small circles of white paper to locate the region of visual space in which they disappeared [1]. Prior to this discovery, the role of the optic disc was not clear, and scientists such as Leonardo da Vinci (1452–1519) had thought that the visual image actually fell only on the head of the optic nerve itself [2]. The vertically oval area of the blind spot subtends about $5 \times 7^\circ$ of visual angle and corresponds to the projection in visual space of the optic nerve head, which has average vertical and horizontal disc diameters of 1.88 and 1.77 mm, respectively [3]. The physiological blind spot is centered about 15° temporally from the point of fixation, due to the nasal anatomical location of the optic nerve head in relation to the fovea. One is normally not aware of the area of the visual field which corresponds to the blind spot in binocular vision, due to the fellow eye's retinal and cortical representation of the corresponding area of the visual field. Even in monocular vision, this region is perceptually filled in at a cortical level

using the brightness, color, and texture detected in the visual field adjacent to the blind spot [4]. This filling in of the scotomatous area of visual field caused by the optic nerve head has also been shown to be an active process involving spatial integration, wherein the pattern and color of not just the area immediately adjacent to the scotoma but also from the more remote surround are used to perceptually fill in the scotomatous region [5].

The size of the blind spot strongly correlates with the area of the optic disc, as well as other features such as the region around the nerve head where the retina has wasted away (i.e. parapapillary chorioretinal atrophy) [6] and the surface topography of the disc, with the more prominent nasal part of the disc appearing less "blind" than the shallow temporal part, probably due to a higher degree of light scattering by the more prominent nasal disc region [7]. In certain pathological conditions, most of which affect the optic nerve head, such as inflammation (i.e. optic neuritis) or optic disc swelling due to raised intracranial pressure, the area of the blind spot can become enlarged, and this can be detected clinically using automated visual field tests or, in more gross abnormalities, confrontational estimation of the blind spot size using a small target such as red hat pin in direct comparison with a normal observer's blind spot. Such pathological conditions of the optic nerve are often associated with defects in color vision, most classically that of red desaturation. These defects may be further evident on testing using pseudoisochromatic plates, typically demonstrating more severe red-green deficits with milder blue-yellow losses [8]. Only the vertebrate eye has a physiological blind spot, on account of light having to pass through the anterior retinal nerve fiber layer (which forms the optic nerve head as it exits the eye) in order to reach the photoreceptor layer beneath; cephalopods, such as the squid and octopus, have retinæ wherein the nerve fiber layer lies behind the photoreceptor layer, and so have no physiological blind spot [9].

Cross-References

► [Pseudoisochromatic Plates](#)

References

1. Grzybowski, A., Aydin, P.: Edme Mariotte (1620–1684): Pioneer of neurophysiology. *Surv. Ophthalmol.* **52**(4), 443–451 (2007)
2. Arrington, G.E.: *A History of Ophthalmology*. MD Publications, New York (1959)
3. Quigley, H.A., Brown, A.E., Morrison, J.D., Drance, S. M.: The size and shape of the optic disc in normal human eyes. *Arch. Ophthalmol.* **108**(1), 51–57 (1990)
4. Spillmann, L., Otte, T., Hamburger, K., Magnussen, S.: Perceptual filling-in from the edge of the blind spot. *Vision Res.* **46**(25), 4252–4257 (2006)
5. Ramachandran, V.S., Gregory, R.L.: Perceptual filling in of artificially induced scotomas in human vision. *Nature* **350**(6320), 699–702 (1991)
6. Jonas, J., Gusek, G., Fernández, M.: Correlation of the blind spot size to the area of the optic disk and parapapillary atrophy. *Am. J. Ophthalmol.* **111**(5), 559–565 (1991)
7. Meyer, J.H., Guhlmann, M., Funk, J.: Blind spot size depends on the optic disc topography: a study using SLO controlled scotometry and the Heidelberg retina tomograph. *Br. J. Ophthalmol.* **81**(5), 355–359 (1997)
8. Yates, J., Diamantopoulos, I., Daumann, F.: Acquired (transient and permanent) color vision disorders. In: Menu, J. (ed.) *Operational Color Vision in the Modern Aviation Environment*. North Atlantic Treaty Organization Research and Technology Organization A, Neuilly-Sur-Seine Cedex (2001). 1
9. Miller, R.F.: The physiology and morphology of the vertebrate retina. In: Ryan, S.J. (ed.) *Retina*, vol. 1, 2nd edn, pp. 58–71. Mosby, St. Louis (1994)

“Blue Blindness”

- [Tritanopia](#)

“Blue-Seven Phenomenon”

- [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)

Blue-Yellow Deficiency

- [Tritanopia](#)

C

Candela Distribution

► [Light Distribution](#)

Canonical Color

► [Memory Color](#)

Car Lighting

► [Automotive Lighting](#)

Carbon Arc Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Arc lamps](#)

Definition

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

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

Carbon Arc Electric Lamps

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

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

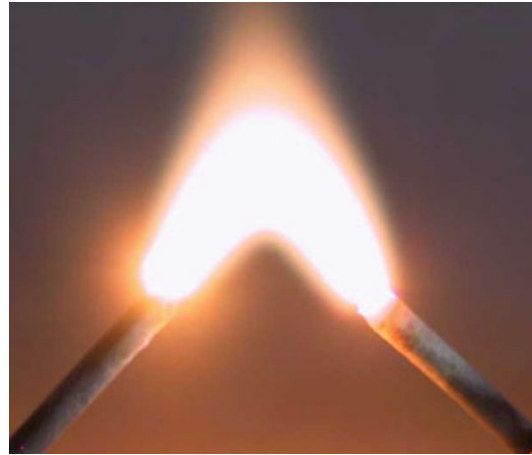
replaced carbon arc lighting installations. Up to the 1950s, extremely high-intensity arc lamps were still used in search lights, in film studios, and in cinema projectors, until the short-arc gas discharge xenon lamp took over. Today, electrical arcs are used not for lighting but for industrial purposes,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

Working Principle

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

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

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



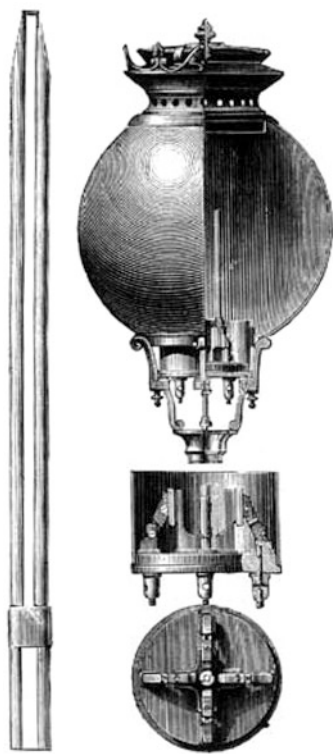
Carbon Arc Lamp, Fig. 1 Because of the rise of heated air, the arc rises in the form of an arch (Photograph: Achgro: Creative Commons 3.0 unported)

Materials and Construction

Electrodes

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

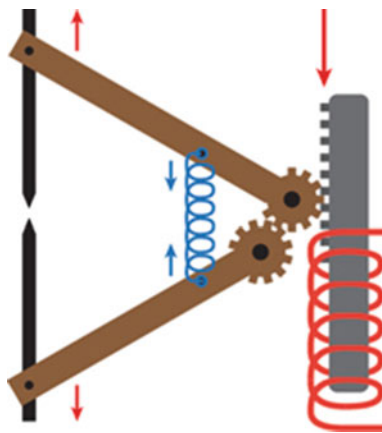
Jablochkoff's Parallel Electrodes Around 1880 the Russ Paul Jablochkoff introduced a whole new concept of electrodes that did away with the need for continuous regulation of the distance between the rods. The "Jablochkoff electric candle," as it is usually called, consists of two parallel rods of carbon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



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

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

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



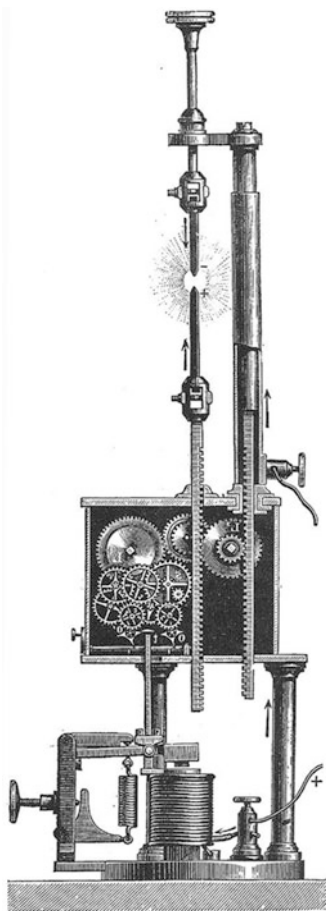
Carbon Arc Lamp, Fig. 3 Principle of a self-regulating mechanism making use of the force of a spring (blue) and that of an electromagnet (red)

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

Electrodes Regulator

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

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



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

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

Lantern

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

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

Properties

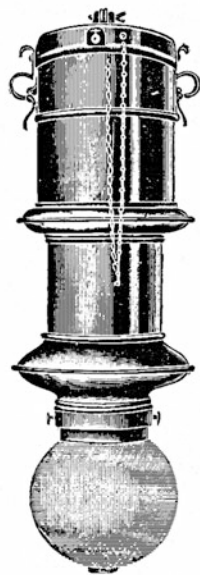
Open carbon arc lamps have a light output of up to some 4,000 lm (500 W versions) at a luminous efficacy of some 4–8 lm/W. Enclosed lamps have a 10–15 % lower output and efficacy. Flame arc lamps, using carbon rods with additives, have a light output up to 15,000 lm (500 W versions) at efficacies between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

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

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

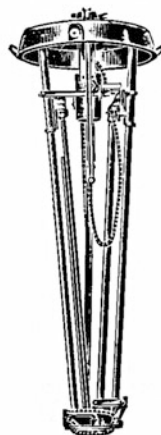
Carbon Arc Lamp,
Fig. 5 Carbon arc street-lighting lantern with self-regulating gap distance between the arcs [4]

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



No. 4.

Innenansicht der
Lampentype G. u. W. 60



No. 5.

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



No. 6.

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

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



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

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

Cross-References

► [Xenon Lamp](#)

References

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

CAT

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

Categorical Perception

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

Chevreul, Michel-Eugène

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

Definition

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

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

Chevreul's Life and Work

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

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



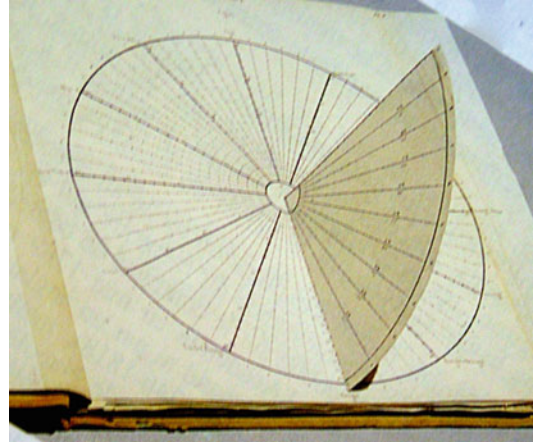
Chevreul, Michel-Eugène, Fig. 1 Michel-Eugène Chevreul (1786–1889) at the approximate age of 50 (lithograph by Nicolas-Eustache Maurin, 1836, engraving by Conrad Cook)

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

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

Means of Naming and Defining Colors

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



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

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

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



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

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

Dyes

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

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

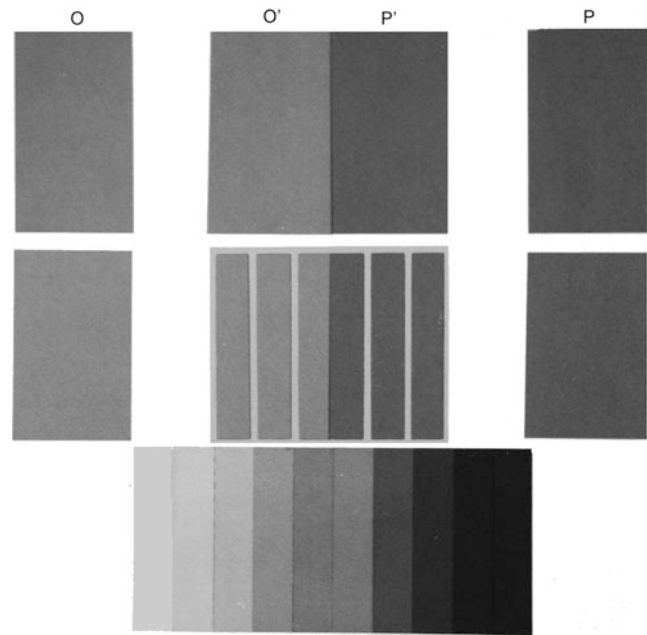
Although Chevreul's work on color covers many aspects, his most important contribution to color is his law of simultaneous contrast of colors, as expounded in his book translated into English

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

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

Chevreul, Michel-Eugène,

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chevreul's Influence on Artists and Artisans

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

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

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

Cross-References

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

References

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

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

Chiasma Opticum

- [Optic Chiasm, Chiasmal Syndrome](#)

Chromatic Adaptation

► Color Constancy

Chromatic Contrast

► Color Contrast

Chromatic Contrast Sensitivity

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

Definition

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

Conceptual Clarifications

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

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

Achromatic Contrast and Isochromatic Images

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

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

Spatial and Temporal Contrast

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

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

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

Color Contrast and Color Edges

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

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

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

Detection, Discrimination, and Contrast Sensitivity

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

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

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

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

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

Spatial Resolution and Chromatic Acuity

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

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

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

Method

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

Contrast Indices

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

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

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

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

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

Color Differences

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

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

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

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

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

Spatial Frequency

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

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

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

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

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

Temporal Frequency

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

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

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

Chromatic Contrast Sensitivity Functions

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

Spatial Contrast Sensitivity Functions

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

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

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

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

Temporal Contrast Sensitivity Functions

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

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

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

Development

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

Eye Movements

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

Theories on Chromatic Contrast Sensitivity

Physiological Basis

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

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

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

Functional Role in Ecology

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

Cross-References

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

References

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

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

Chromatic Image Statistics

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

Chromatic Processing

► [Color Processing](#), [Cortical](#)

Chromostereopsis

Akiyoshi Kitaoka

Department of Psychology, Ritsumeikan University, Kyoto, Japan

Synonyms

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

Definition

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

Overview

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

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



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



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

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

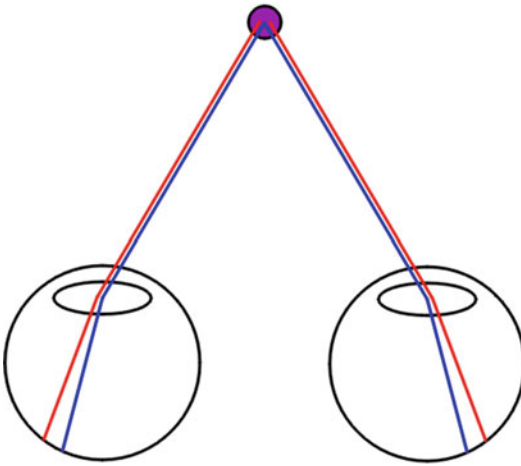
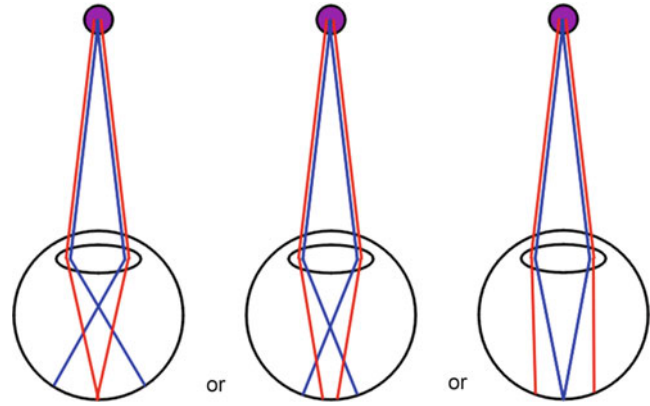
Both Donders and Bruecke attributed chromostereopsis to accommodative feeling, which was translated to the perception of distance [4]. Blue or short-wavelength light has a larger refractive index than the red or long-wavelength one. This makes the color rays run in a different way as shown in Fig. 3.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

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

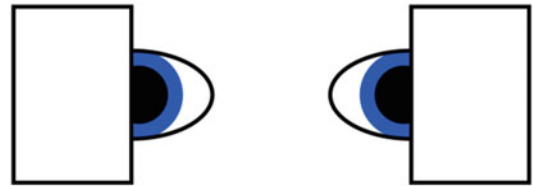
According to Vos [4], Einthoven discovered in 1885 that chromostereopsis is enhanced or reversed by using a simple method as shown in Fig. 5 [11, 12]. Covering the temporal parts of both pupils forces observers to see blue in front of red (upper image of Fig. 5). On the other hand, covering the nasal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

Chromostereopsis,

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



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



Appearance: *blue* in front



Appearance: *red* in front

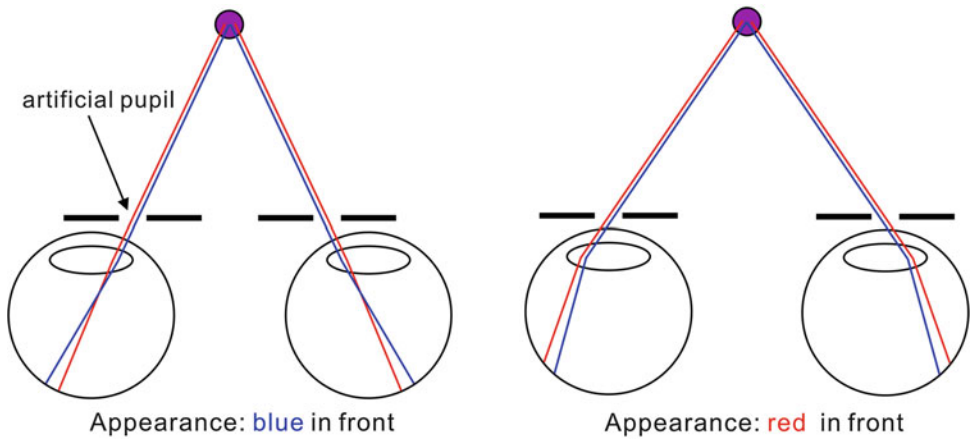
Chromostereopsis, Fig. 5 Eindhoven's covering method. *Blue* in front of *red* is generated or enhanced by covering the temporal parts of both pupils (*upper image*), while *red* in front of *blue* is produced or enhanced by covering the nasal parts of both pupils (*lower image*)

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

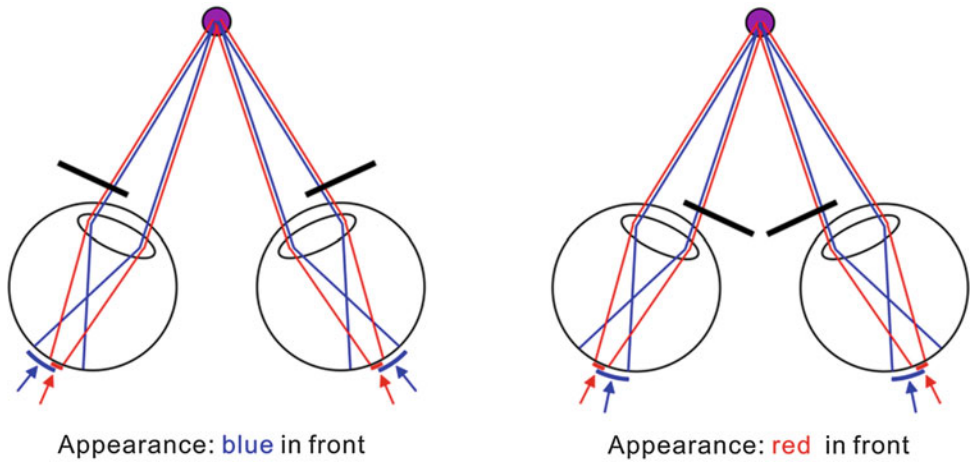
Subsequently, Eindhoven's covering method was simplified to the method using artificial pupils. Nasally placed artificial pupils gave blue in front of red (upper image of Fig. 6), while temporally placed ones rendered red in front of blue (lower image of Fig. 6) [15–23]. Vos [4, 16, 18] attributed chromostereopsis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

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

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



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

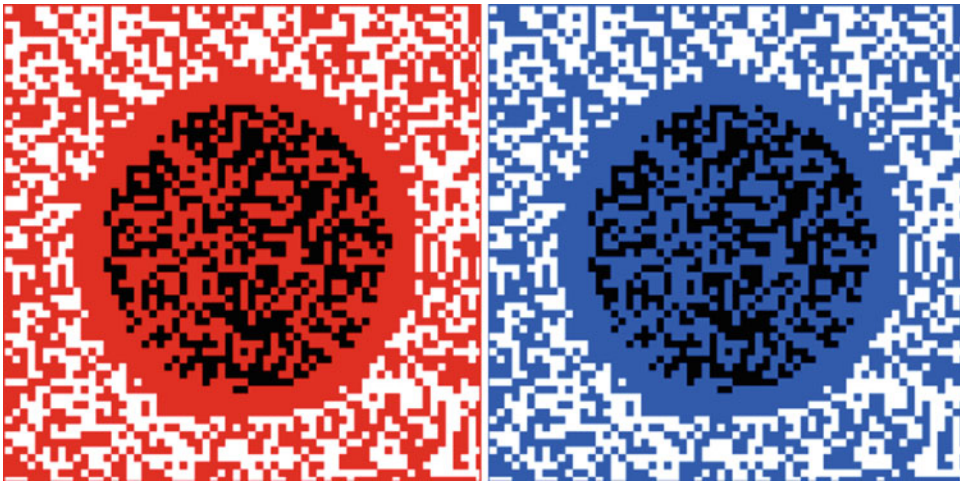


Chromostereopsis, Fig. 7 The center-of-gravity model. This model takes longitudinal chromatic aberration into account, in which the perceived position of a color is judged to be at the center of gravity of its diffused light

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

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

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



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

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

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

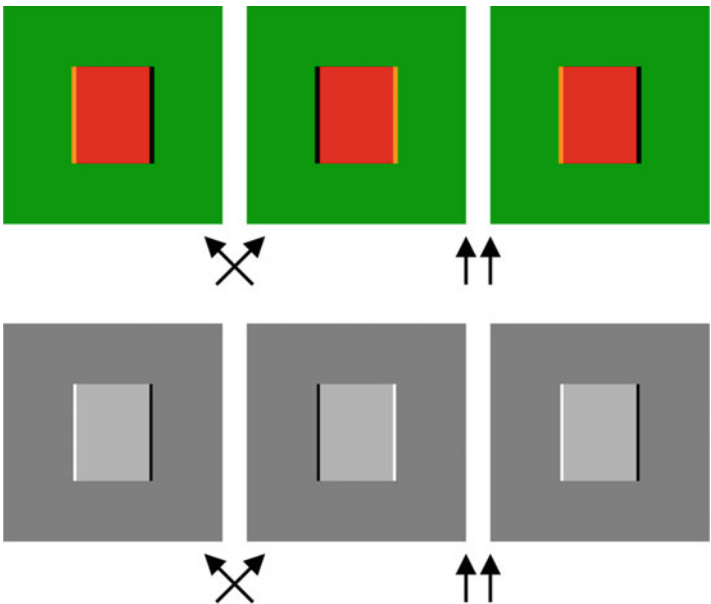
In 1928, Verhoeff reported that the perceived depth order between red and blue is reversed by changing the background from black to white (Fig. 2) [9, 28–30]. One account is that red surrounded by white lacks blue, while blue surrounded by white lacks red, suggesting that there are virtually blue and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

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



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

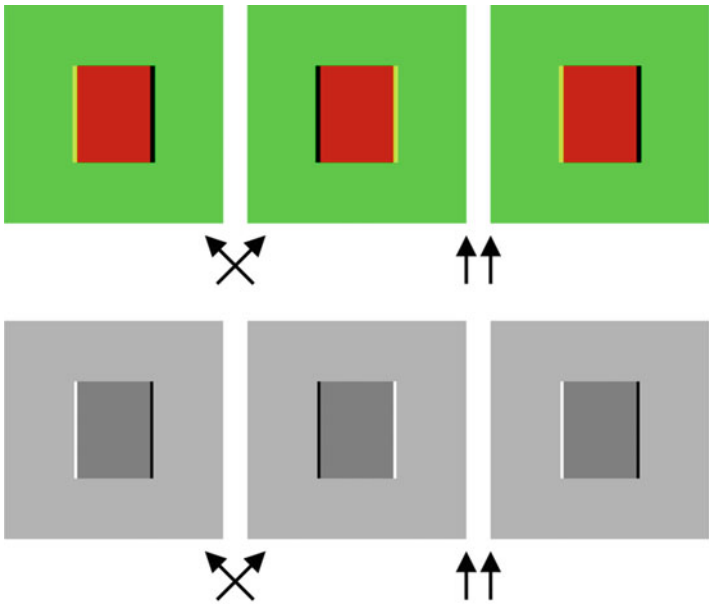
correspond to the one which Gregory and Heard showed in 1983 as shown in Fig. 10 [34, 35], though Faubert did not mention it. If the luminance order is reversed between the two colors, the apparent depth is reversed as shown in Fig. 11.

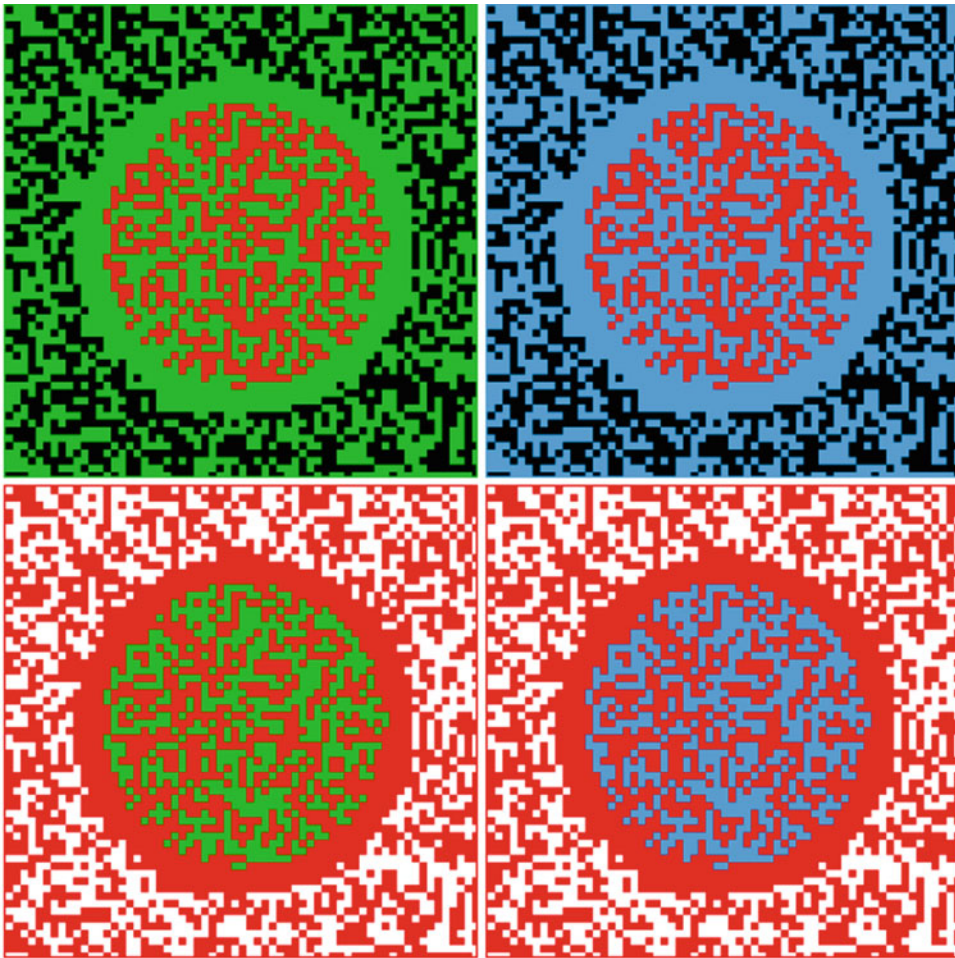


Chromostereopsis, Fig. 10 Faubert’s luminance-profile-based idea [31]. Color fringes or luminance profiles of a *red* object in front of *green* such as that in Fig. 9 produced by transverse chromatic aberration are depicted. It is supposed that mirror images are rendered to each eye. This binocular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

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

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





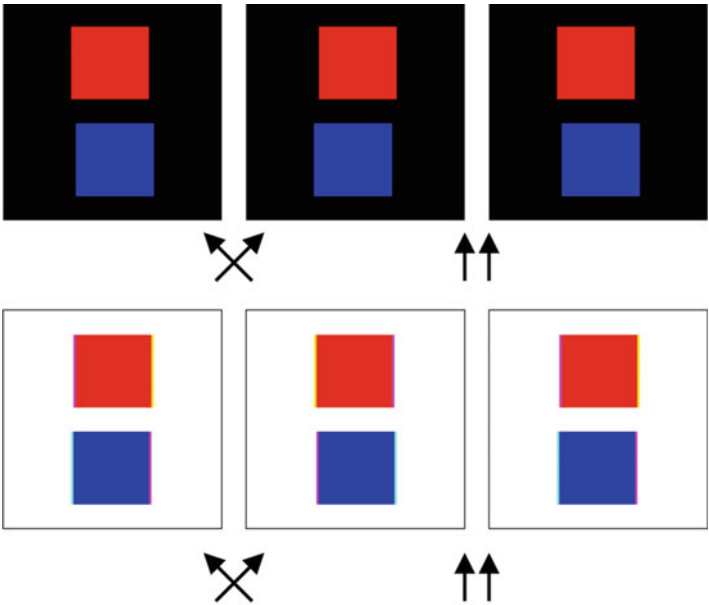
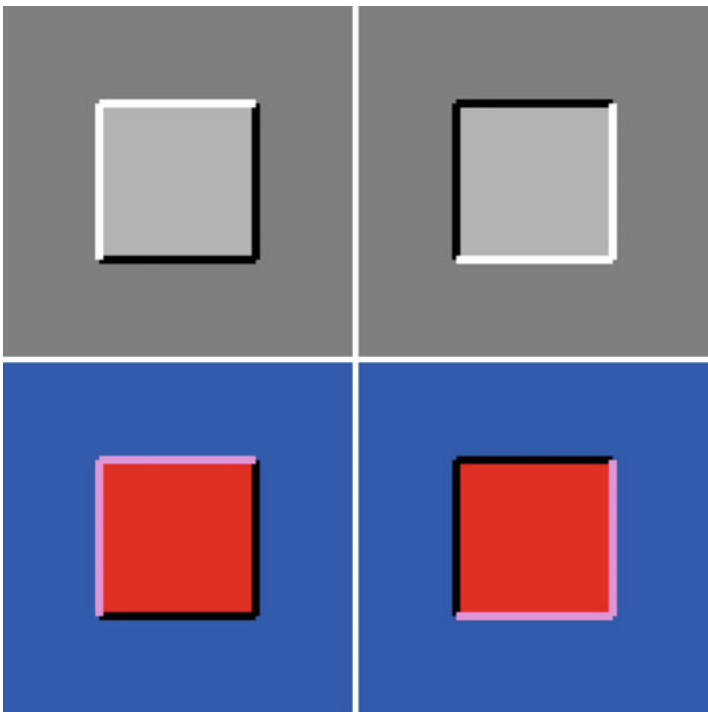
Chromostereopsis, Fig. 12 Isoluminant colors make observers perceive rivaldepth with luster. Note that precise isoluminance is realized depending on displays and individuals

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

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

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

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

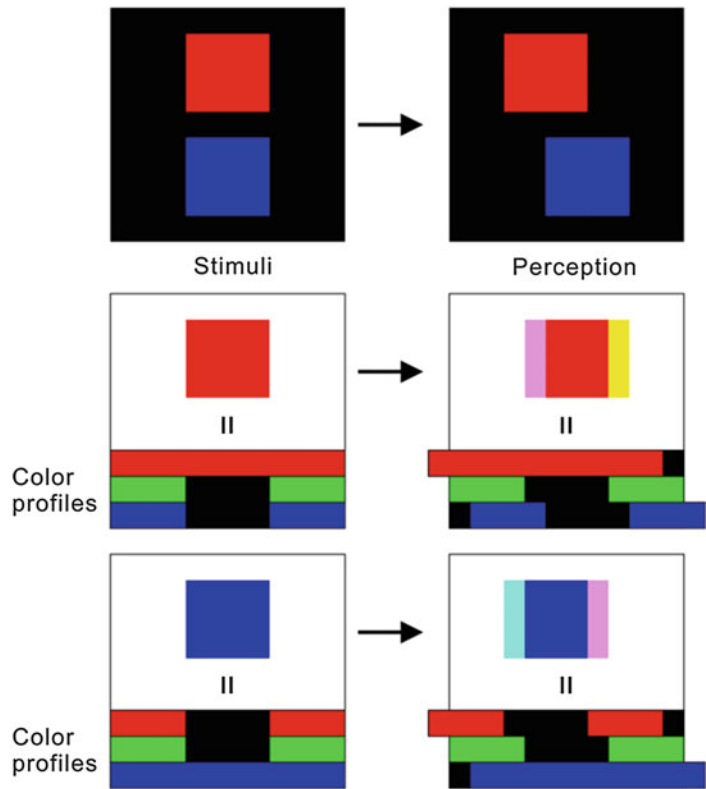


Chromostereopsis, Fig. 14 Faubert’s model can be extended to explain the color reversal. If *red* and *blue* objects are vertically aligned with the *black* background but appear to be shifted in position by transverse chromatic aberration as shown in the *upper panel*, the expected shifts with the *white*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

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

Chromostereopsis,

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

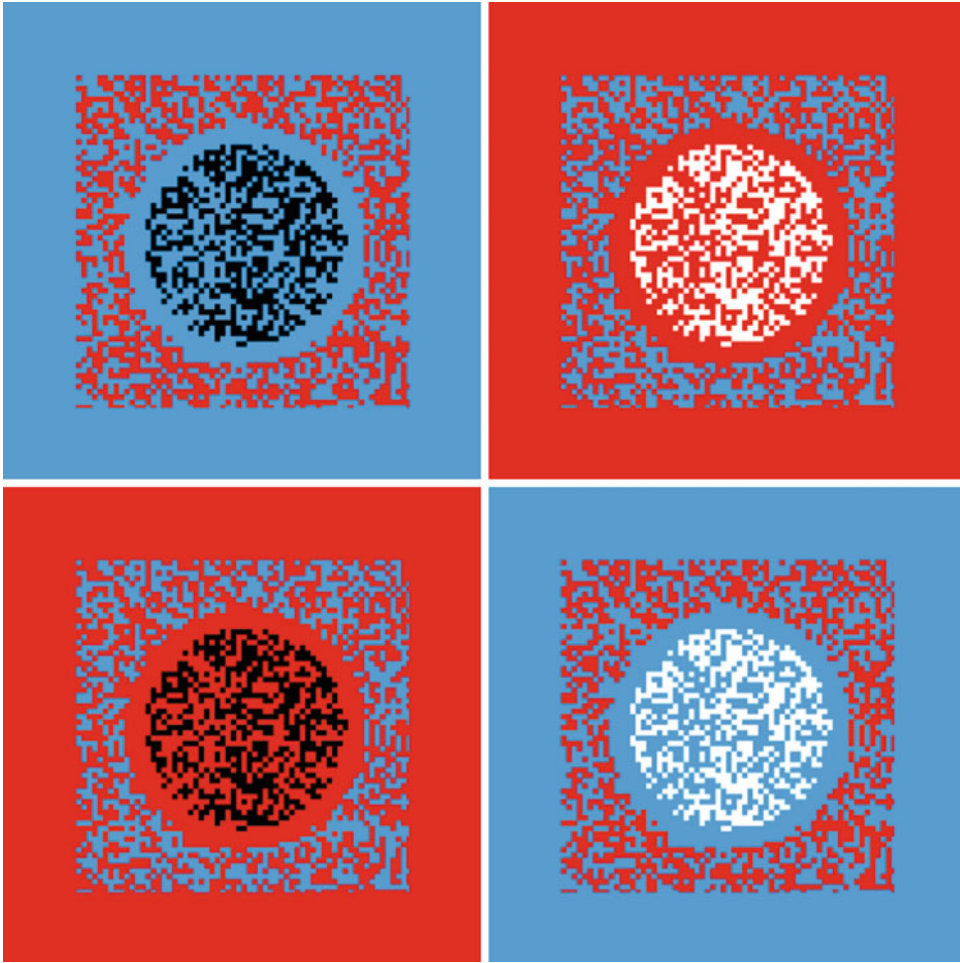


Chromostereopsis, Fig. 16 The *inset* appears to be in front of the surround. High-luminance and/or high-contrast objects appear to be closer than low-luminance and/or low-contrast ones

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

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

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



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

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

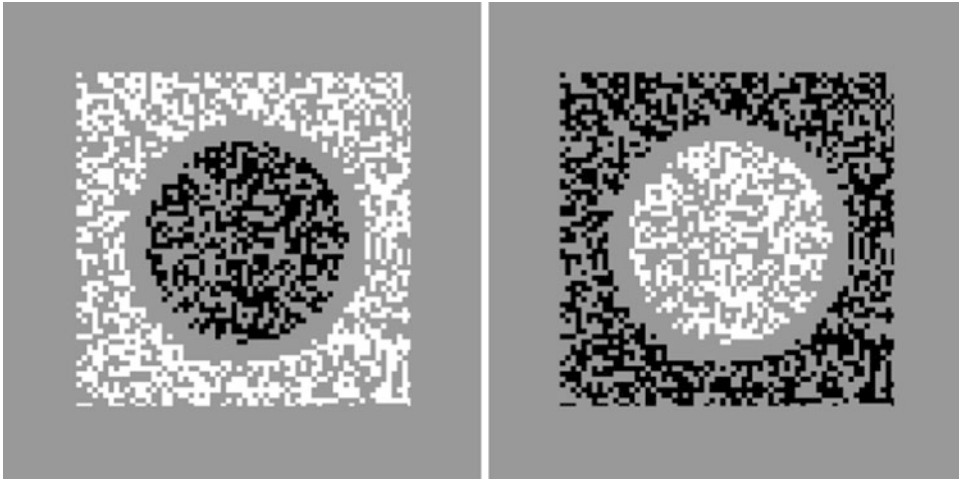
It is thought that some part of the effect is due to luminance differences or contrast differences (Fig. 16), with bright objects appearing closer than dim ones [6] or high-contrast objects appearing closer than low-contrast ones [38]. Saturation also affects chromostereopsis [11, 13]. Desaturation decreased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

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

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

Summary

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



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

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

Cross-References

► Stereo and Anaglyph Images

References

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

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

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

János Schanda
Veszprém, Hungary

Synonyms

CIE Color-Matching Functions

Definition

CIE color-matching functions

CIE 1931 standard colorimetric observer

CIE 1964 standard colorimetric observer

Functions

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

Overview

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

Chromotherapy

► Color Psychology

János Schanda: deceased

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

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

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

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

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

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

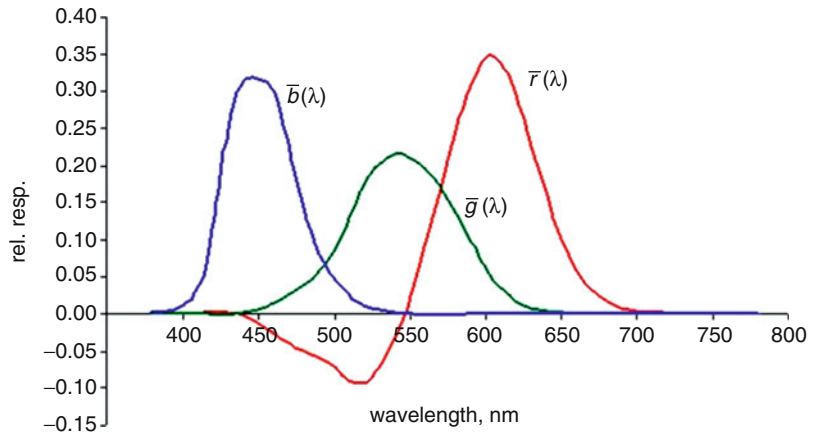
Using $1 \text{ cd}\cdot\text{m}^{-2}$ monochromatic red light of 700 nm as $[R]$, $4.5907 \text{ cd}\cdot\text{m}^{-2}$ monochromatic green light of 546.1 nm as $[G]$ and $0.0601 \text{ cd}\cdot\text{m}^{-2}$ monochromatic blue light of 435.8 nm as $[B]$ for the units of the matching stimuli, one gets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

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

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

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

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



the $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, $\bar{b}(\lambda)$

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

The transformation matrix has the following form:

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

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

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

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

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

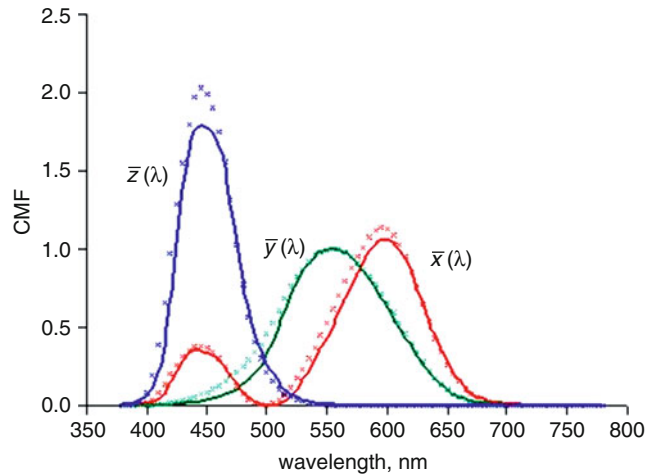
Short History of the CIE Colorimetric Observer

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

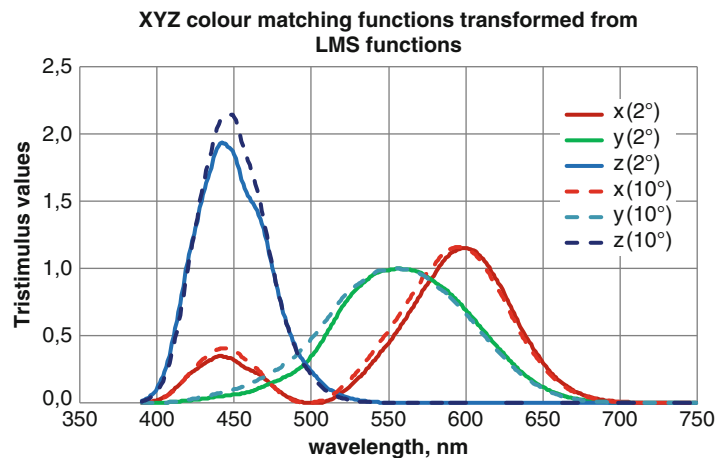
Photometry and colorimetry were further developed in the USA that lead to the so-called OSA excitation curves.

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

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



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



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

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

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

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

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

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

Cross-References

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

References

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

CIE 1976 L*a*b*

- CIELAB

CIE 1976 L*u*v* Color Space

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

CIE 1994 (ΔL^* ΔC^*_{ab} ΔH^*_{ab})

- CIE94, History, Use, and Performance

CIE 2000 Color-Difference Equation

- CIEDE2000, History, Use, and Performance

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

Ming Ronnier Luo

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

School of Design, University of Leeds, Leeds, UK

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

Synonyms

CAT

Definition

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

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

Overview

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

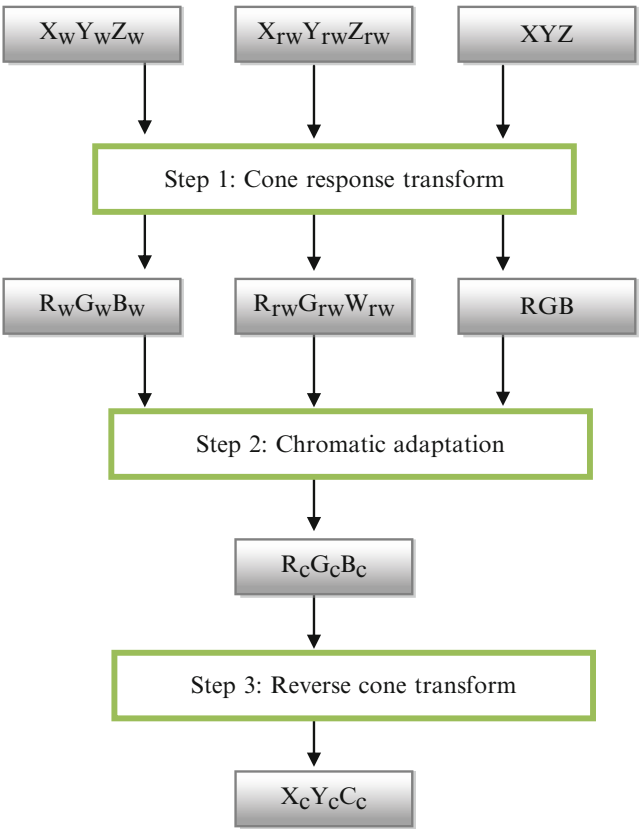
appearance model, which is capable of predicting color appearance under different viewing conditions such as illuminants, levels of luminance, background colors, and media (e.g., reflection, transmissive and self-luminous display). The CIE [4] recommended CIECAM02 for the application of the color management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

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

Step 1 Cone response transform

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

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



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

Step 2 Chromatic adaptation mechanism

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

Step 3 Reverse cone transformation

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

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

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

von Kries Chromatic Adaptation Transform

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

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

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

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

where

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

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

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

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

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

Step 2 Calculate the α, β , and γ von Kries coefficients and the R_c, G_c , and B_c values using Eqs. 1 and 2.

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

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

CIELAB

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

CMCCAT97 Chromatic Adaptation Transformation

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

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

Step 1 Calculation of R , G , B , R_w , G_w , and B_w and R_w , G_w , and B_w using Eq. 5:

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

where

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

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

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

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

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

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

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

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

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

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

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

CAT02 Transform

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

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

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

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

where

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

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

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

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

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

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

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

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

Experimental Datasets Investigated by CIE TC1-52

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

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

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

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

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

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

Evaluation of CATs

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

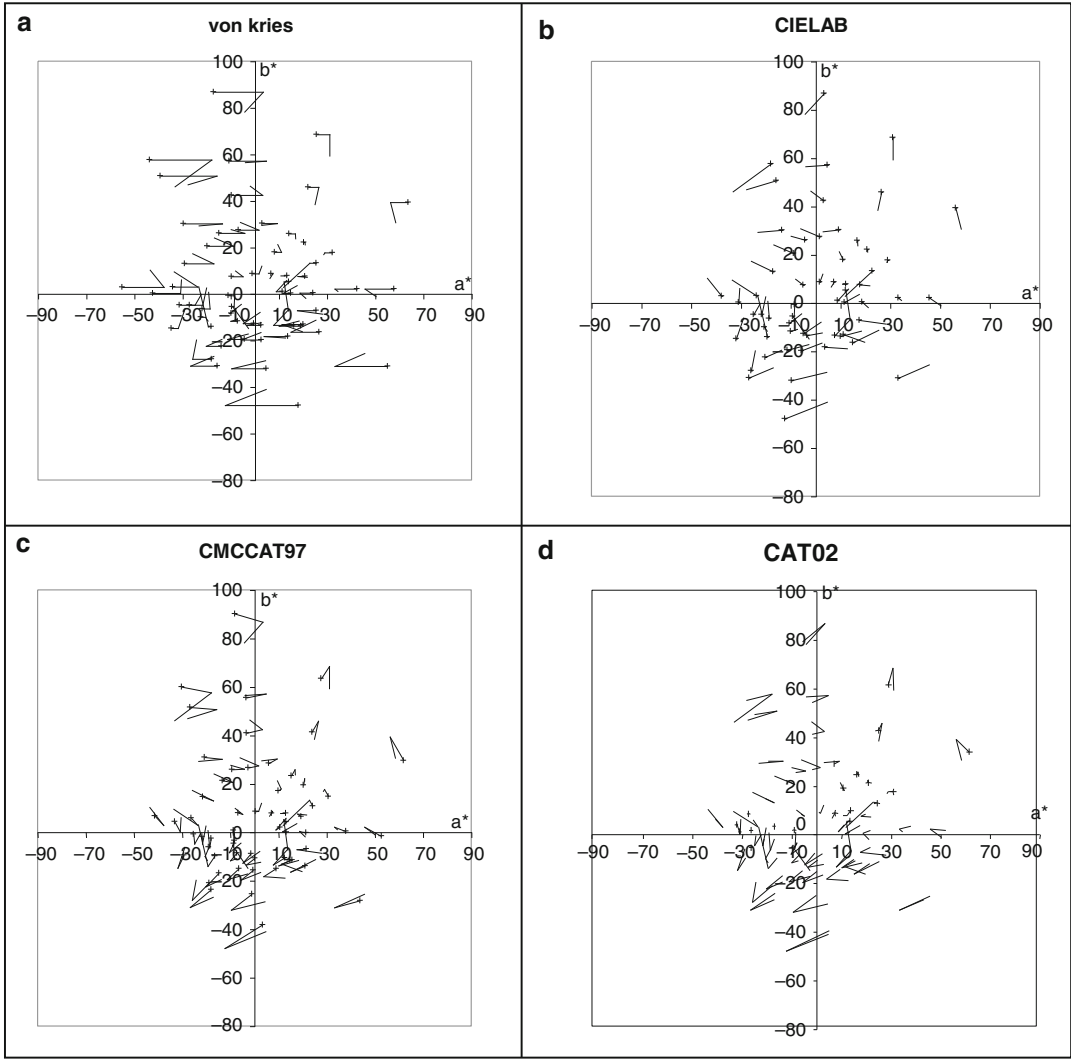
It was found [17] that the typical observer variation for studying chromatic adaptation was about 4 CMC (1:1) units. Hence, if a CAT has an error of prediction equal to or less than 4 units, it may be considered to be satisfactory. As shown in Table 2, the most accurate transform for each dataset (the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

account the number of corresponding pairs in each dataset.

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

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

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



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

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

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

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

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

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

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

Future Directions

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

Cross-References

- CIECAM02
- CIELAB

References

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

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

CIE Chromaticity Coordinates (xyY)

Stephen Westland

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

Definition

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

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

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

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

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

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

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

Chromaticity Coordinates

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

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

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

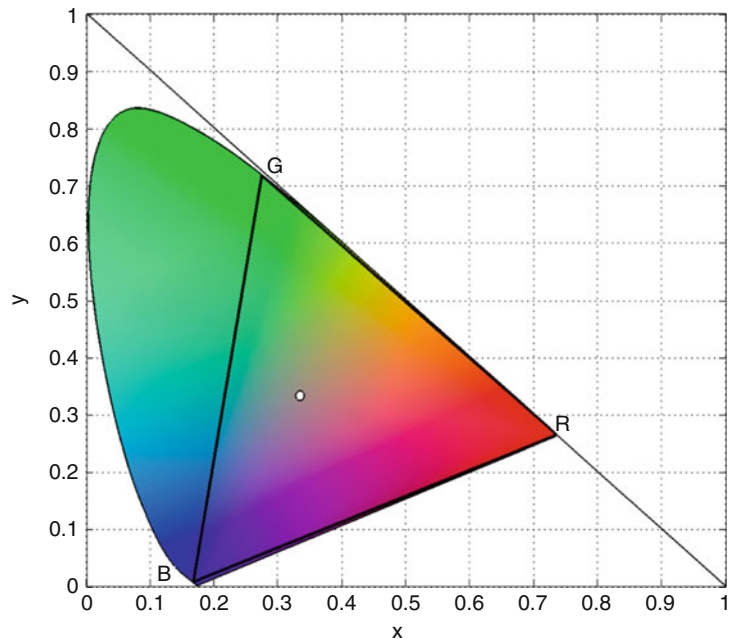
Inherent in the way that the chromaticity coordinates are calculated (Eq. 1) is the constraint that $x + y + z = 1$, and this means that there are only really two degrees of freedom since $z = 1 - x - y$. Since there are only two *free* variables, it is possible to construct a 2D diagram referred to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

Properties of the Chromaticity Diagram

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

CIE Chromaticity Coordinates (xyY),

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



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

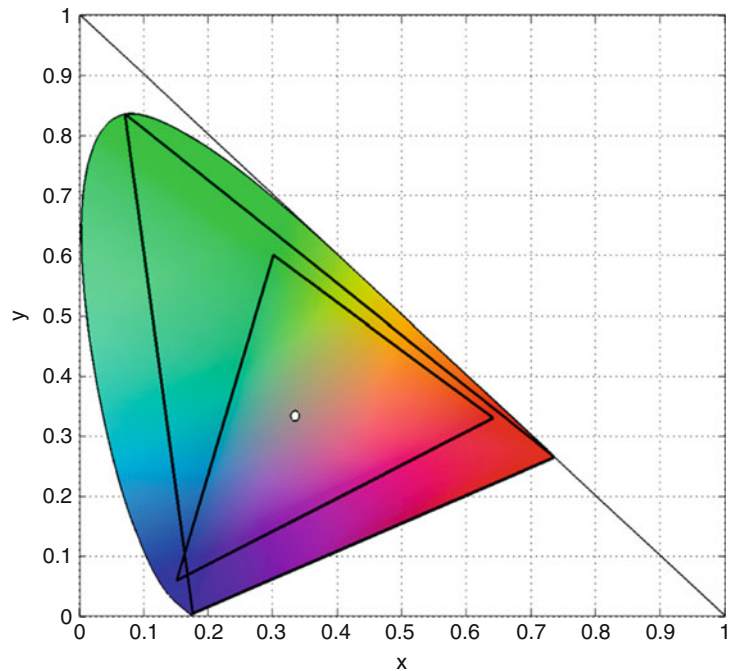
If two lights are represented in the chromaticity diagram by two points, then chromaticities of the additive mixtures of the two lights will be represented by the straight line that joins the two points. Thus, in Fig. 1, all mixtures of the *R* and *G* primaries would lie on the straight line joining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

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

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

CIE Chromaticity Coordinates (xyY),

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



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

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

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

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

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

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

Further Considerations and Future Directions

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

Cross-References

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

References

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

CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength

János Schanda
Veszprém, Hungary

Definitions

Chromaticity Diagram

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

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

Purity (of a Color Stimulus)

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

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

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

Excitation Purity [p_e]

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

color stimulus considered, leading to the following expressions:

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

or

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

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

Unit: 1

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

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

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

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

Colorimetric Purity [p_c]

Quantity defined by the relation

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

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

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

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

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

Dominant Wavelength (of a Color Stimulus) [λ_d]

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

Unit: nm

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

Overview

x, y Chromaticity Diagram

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

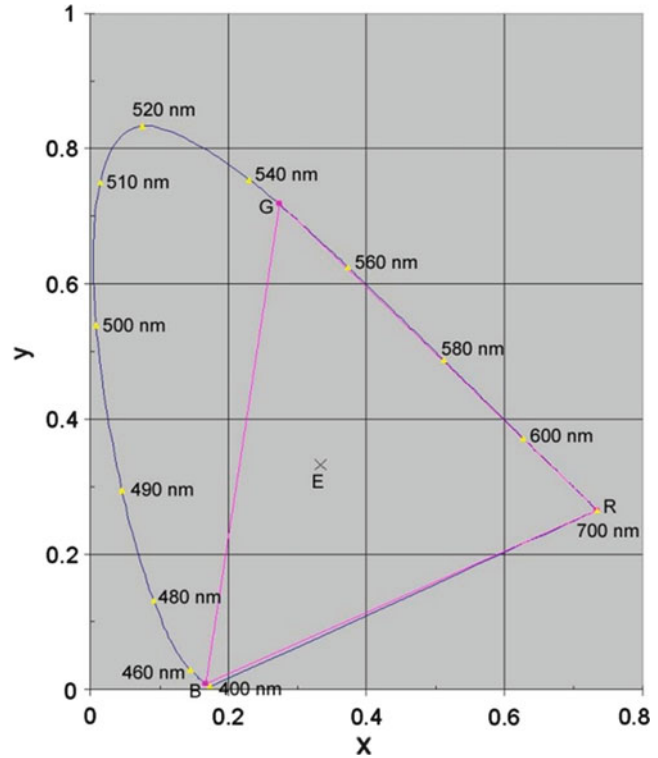
$$x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}. \quad (1)$$

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

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

CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength, Fig. 1 CIE 1931

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



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

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

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

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

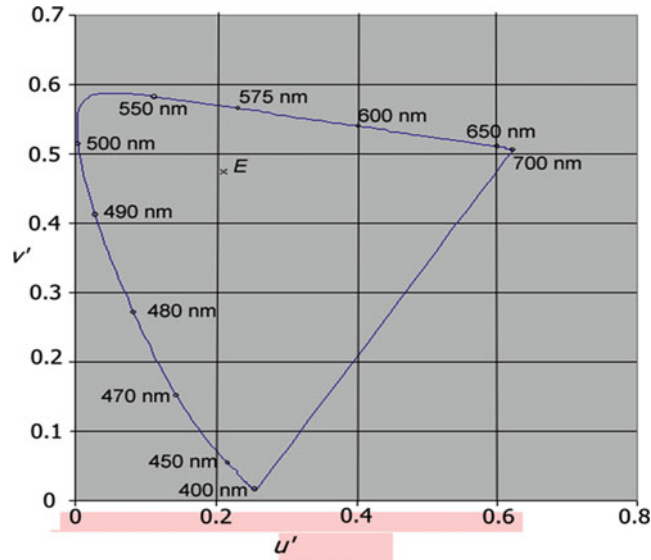
With these coordinates, the chromaticity diagram has the form as shown in Fig. 2. Comparing this diagram with the x , y diagram, it becomes obvious how nonuniform the x , y diagram is (see details in entry “► CIE u' , v' Uniform Chromaticity Scale Diagram and CIELUV Color Space”).

An equivalent transformation starting from the tristimulus values is

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

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

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



Dominant Wavelength and Purity

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

Dominant and Complementary Wavelength

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

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

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

Excitation Purity

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

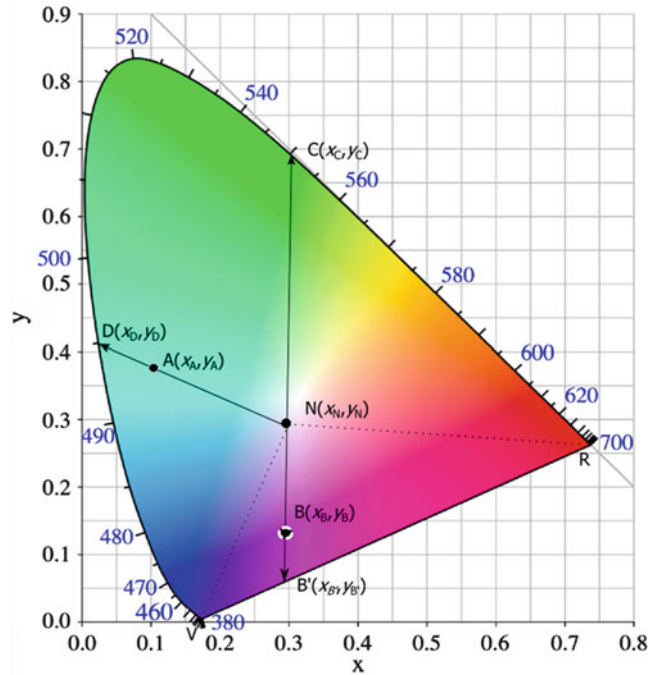
Excitation purity of chromaticity point A :

$$p_{e,A} = \frac{x_A - x_N}{x_D - x_N} \quad (4)$$

Excitation purity of chromaticity point B :

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

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



Colorimetric Purity

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

$$p_c = \frac{L_D}{L_N + L_D}. \quad (6)$$

Summary

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

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

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

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

$$u = u', \quad v = \frac{2}{3}v'. \quad (7)$$

For further details, see Eq. 2.

Cross-References

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

References

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

CIE Color Appearance Model 2002

- [CIECAM02](#)

CIE Color-Matching Functions

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

CIE Color-Rendering Index

János Schanda
Veszprém, Hungary

Definition

Color rendering (of a light source)

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

Color rendering index

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

CIE 1974 special color rendering index [R_i]

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

CIE 1974 general color rendering index [R_a]

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

Overview

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

CIE Color Rendering Index

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

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

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

Transformation from color differences (ΔE_i) to special CRI-s (R_i) is by

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

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

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

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

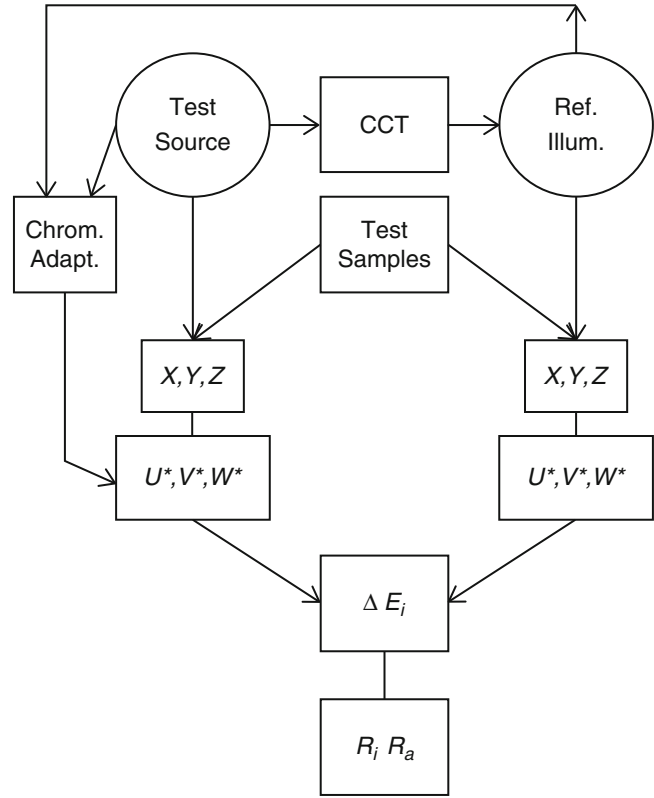
Colorimetry used in the calculation of above CRI was the best correlate for color appearance at the time of its elaboration. Since the 1960s, the description of color appearance progressed considerably; new color appearance models have been developed [7]. During the past 30 years, a large number of papers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

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

Color Fidelity, Preference, and Discrimination

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

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



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

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

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

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

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

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

Further Color Quality Proposals

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

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

References

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

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

CIE Cone Fundamentals

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

CIE Fundamental Color Matching Functions

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

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

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

Definition

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

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

Overview

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

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

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

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

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

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

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

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

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

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

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

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

$$L^*_t = (1 - p_C)L^*_{o_0} + p_C L^*_s \quad (1)$$

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

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

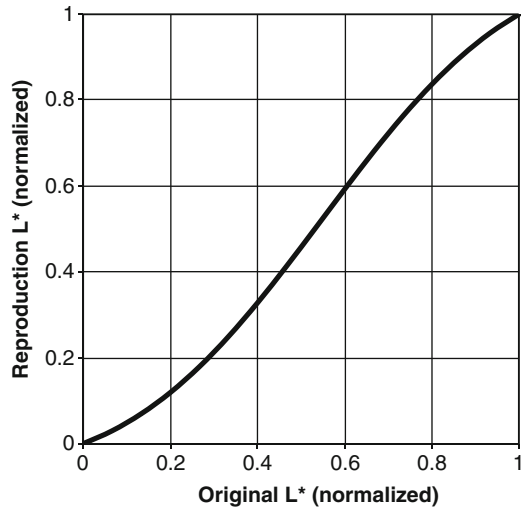
and where L^*_s is the result of the original's L^*_o being mapped using the following sigmoidal function (Fig. 2), having x_0 and Σ parameters empirically derived for different levels of reproduction medium minimum L^* (e.g., for a minimum reproduction L^* of 15, $x_0 = 58.2$ and $\Sigma = 35$):

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

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

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

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



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

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

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

$$d_r = \begin{cases} d_o; & d_o \leq 0.9d_{gr} \\ 0.9d_{gr} + (d_o - 0.9d_{gr})0.1d_{gr}/(d_{go} - 0.9d_{gr}); & d_o > 0.9 \times d_{gr} \end{cases} \quad (6)$$

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

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

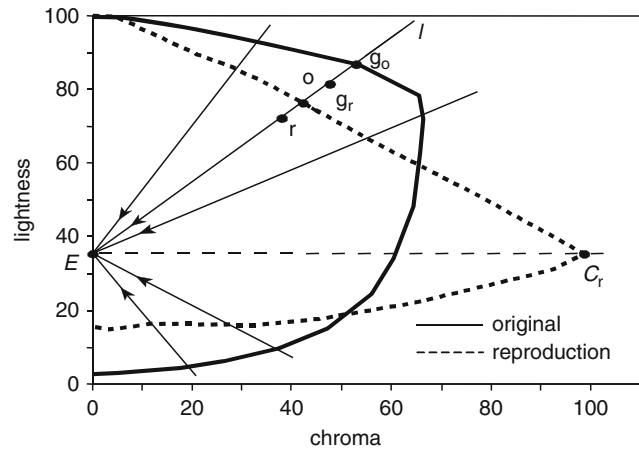
In terms of color spaces in which gamut mapping is to be performed, the *Guidelines* again only require the reporting of whichever space was used. A recommendation is made for

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

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

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

Fig. 3 L^* and C^* mapping
toward the L^* of the cusp



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

Future Directions

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

References

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

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

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

CIE Guidelines for Mixed Mode Illumination: Summary and Related Work

Suchitra Sueeprasan

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

Synonyms

Mixed adaptation condition; Mixed chromatic adaptation

Definition

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

Background

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

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

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

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

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

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

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

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

Recommended Model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Implementation

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

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

Cross-References

- [Adaptation](#)
- [CIECAM02](#)

References

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

CIE L*a*b*

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

CIE Method of Assessing Daylight Simulators

Robert Hirschler

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

Definition

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

Development of the CIE Methods of Assessing Daylight Simulators

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

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

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

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

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

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

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

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

Chromaticity Limits of Daylight Simulators

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

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

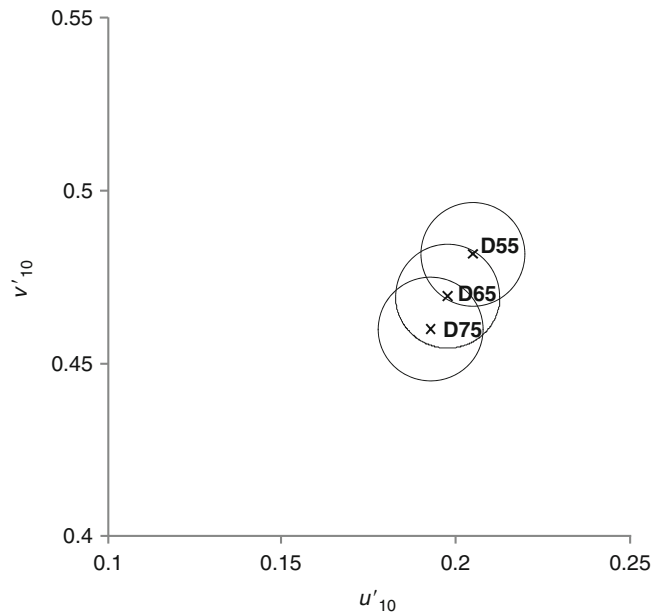
Visible Range Evaluation of Daylight Simulators

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

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

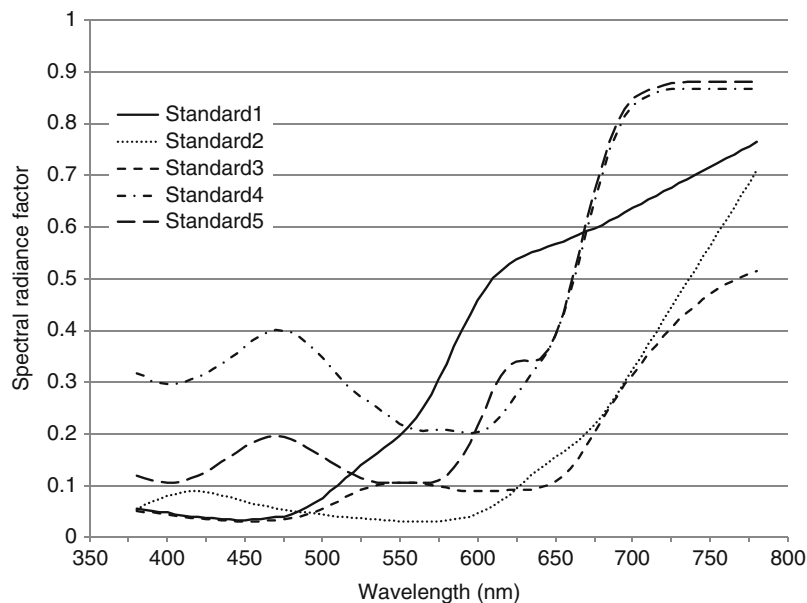
CIE Method of Assessing Daylight Simulators,

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



CIE Method of Assessing Daylight Simulators,

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



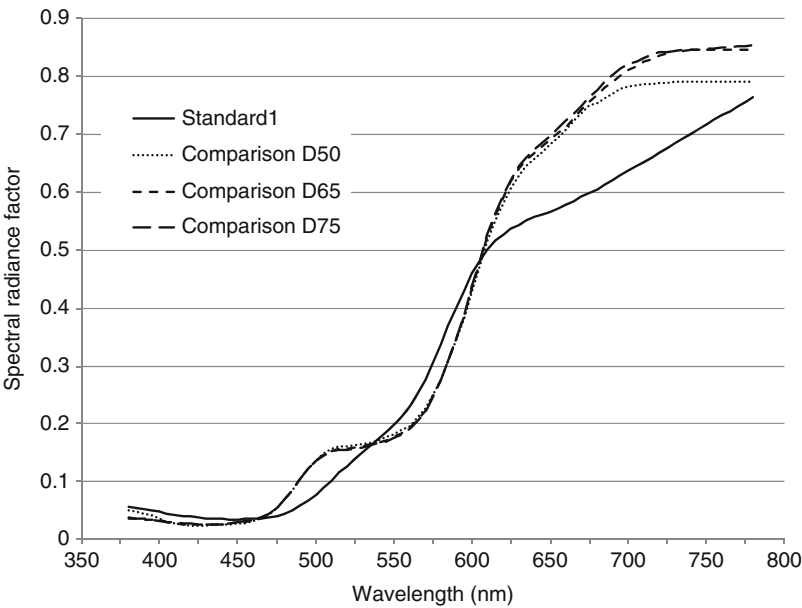
assessment based on the tables published in the CIE Standard [1].

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

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

CIE Method of Assessing Daylight Simulators,

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



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

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

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

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

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

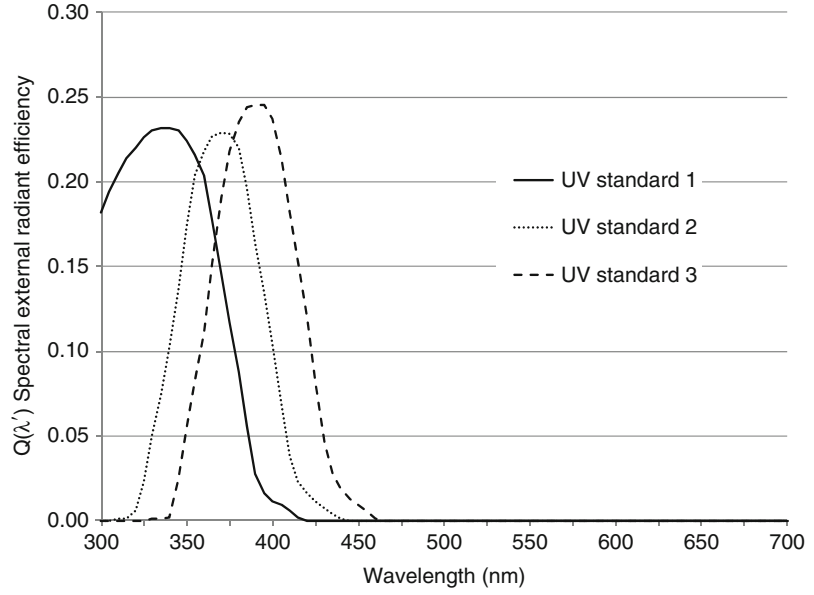
Ultraviolet Range Evaluation of Daylight Simulators

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

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

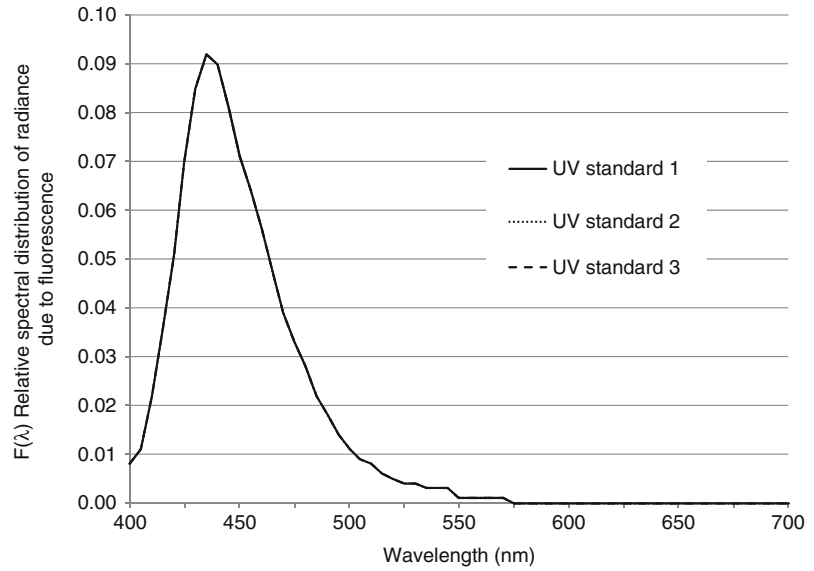
CIE Method of Assessing Daylight Simulators,

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



CIE Method of Assessing Daylight Simulators,

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



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

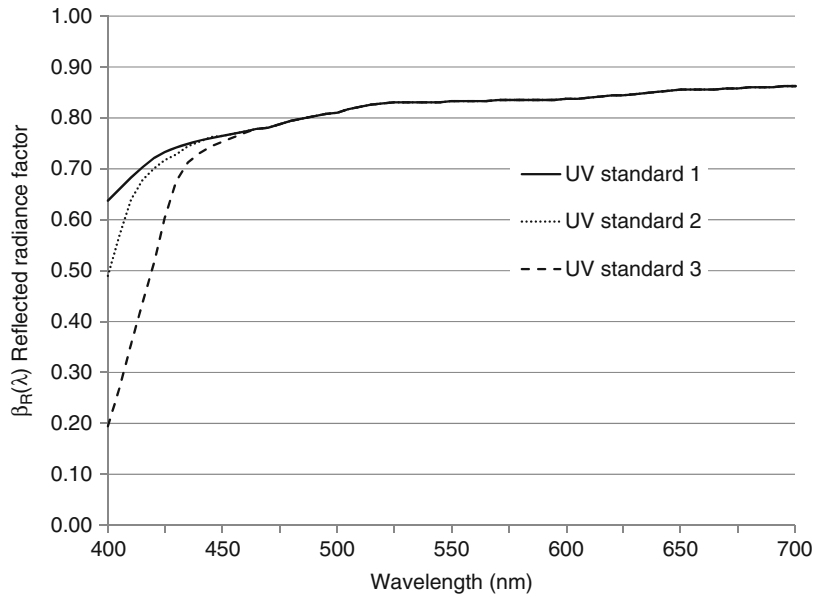
$$N = \sum_{300}^{460} S_n(\lambda') \cdot Q(\lambda') \cdot \Delta\lambda' \quad (1)$$

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

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

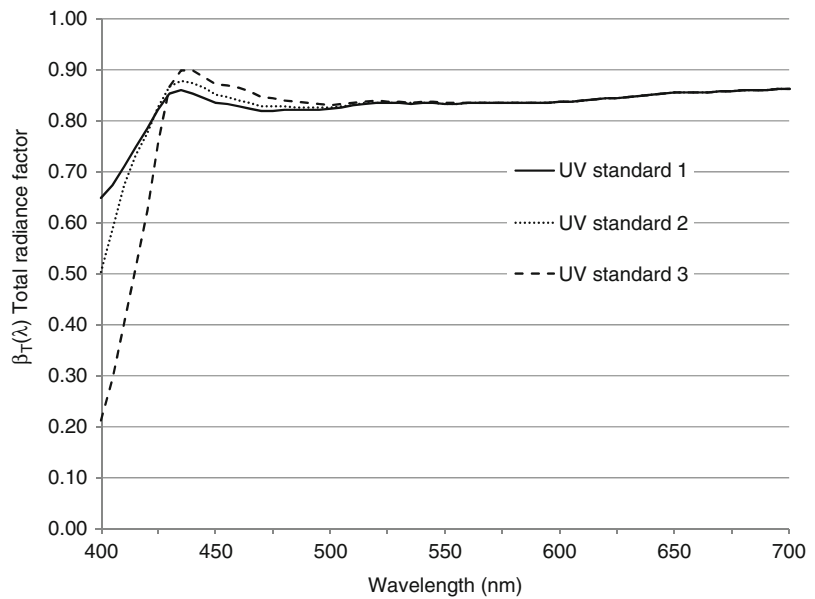
CIE Method of Assessing Daylight Simulators,

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



CIE Method of Assessing Daylight Simulators,

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



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

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

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

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

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

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

$$\beta_T(\lambda) = \beta_R(\lambda) + \beta_F(\lambda) \quad (3)$$

as illustrated in Fig. 7.

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

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

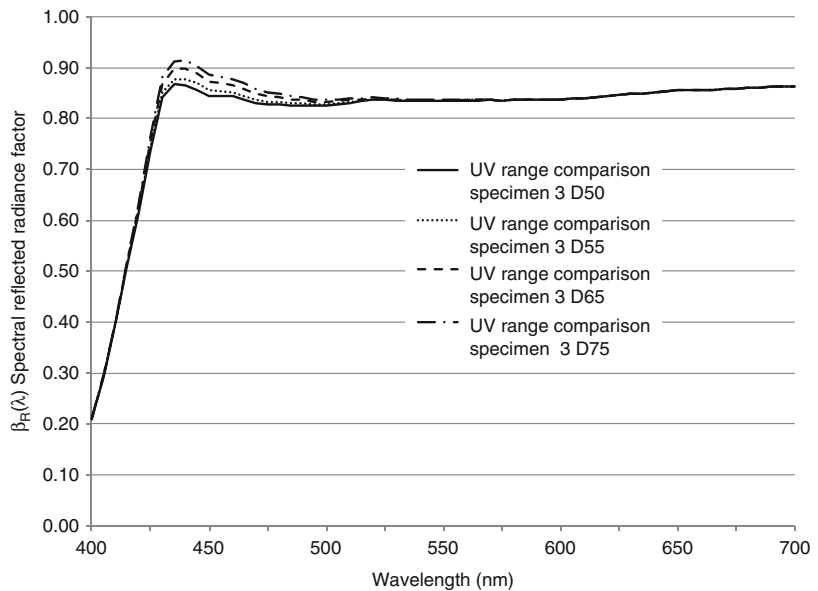
The UV standard and the UV comparison specimens are isomeric, i.e., they are practically identical when illuminated by the reference illuminants. Thus, the UV standard 3 curve for D65 from Fig. 7 is the same as the comparison specimen 3 curve for D65 from Fig. 8. In Fig. 9, we see both curves (dotted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

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

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

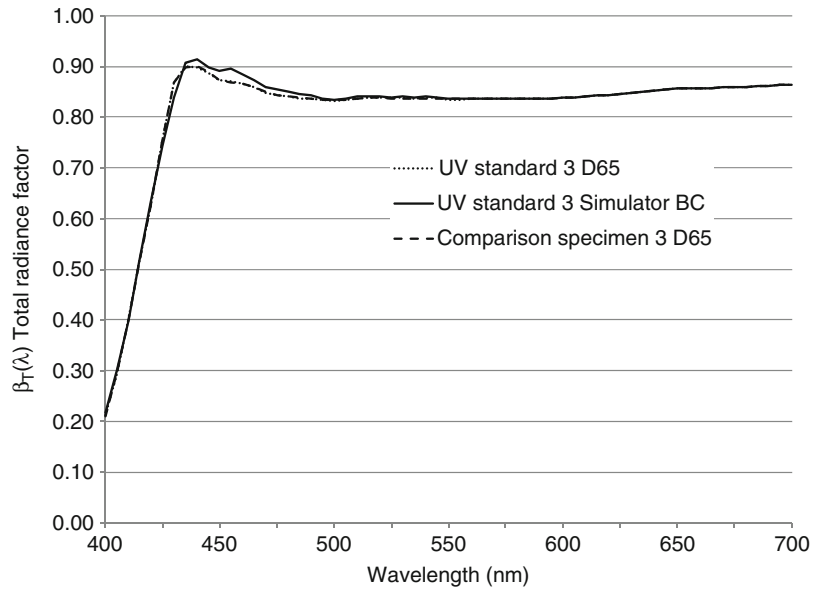
- Determine the relative spectral irradiance of the light source $S(\lambda)$ in the 300–700 nm range
- Take $Q(\lambda')$ for each of the three standards from the data illustrated in Fig. 4
- Calculate N for each of the three standards from Eq. 1
- Take the $F(\lambda)$ values for each of the three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

CIE Method of Assessing Daylight Simulators,
Fig. 8 Spectral reflected radiance factor of the first comparison specimen for the four CIE illuminants



CIE Method of Assessing Daylight Simulators,

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



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

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

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

Practical Application of the CIE Method

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

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

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

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

References

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

CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams

Andrew Stockman

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

Synonyms

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

Definition

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

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

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

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

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

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

Overview

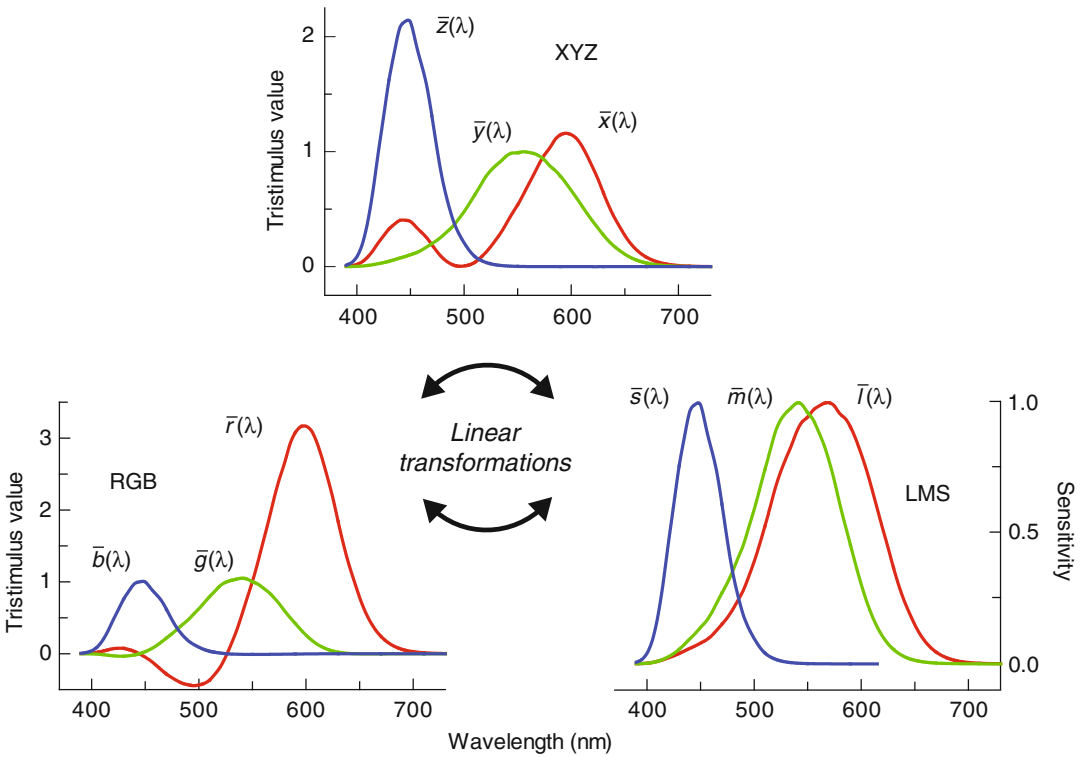
A consequence of trichromacy is that the color of any light can be specified as the intensities of the three primary lights that match it. The bottom left-hand panel of Fig. 1 shows $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs for **RGB** (red-green-blue) primaries of 645, 526,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

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

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

Choice of “Physiologically Relevant” RGB CMFs

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



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

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

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

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

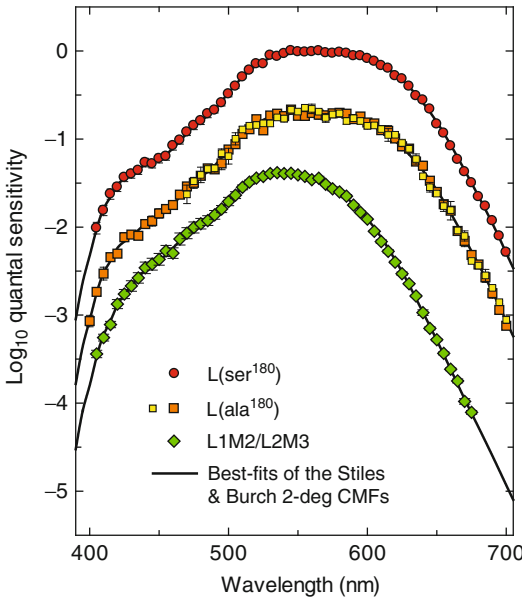
The most secure and comprehensive set of directly measured color matching data are the large-field, centrally viewed 10-deg CMFs of Stiles and Burch [3]. They were measured in 49 subjects from approximately 390–730 nm (and in nine subjects from 730 to 830 nm). Consequently, the 10-deg CMFs of Stiles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

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

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

Spectral Sensitivity Measurements

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



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

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

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

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

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

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

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

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

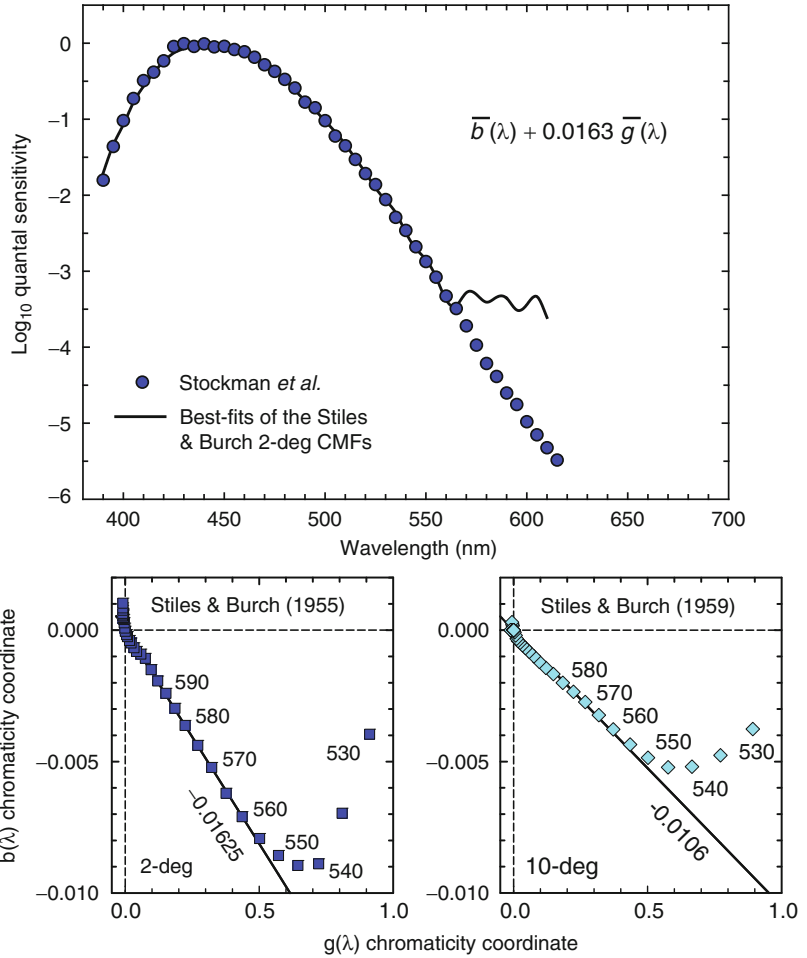
$$\bar{s}_R \bar{r}(\lambda) + \bar{s}_G \bar{g}(\lambda) + \bar{s}_B \bar{b}(\lambda) = \bar{s}(\lambda),$$

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

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

CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams, Fig. 3

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



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

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

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

L- and M-cone Fundamentals

The four M- and L-cone unknowns in Eq. 3, \bar{l}_R/\bar{l}_B , \bar{l}_G/\bar{l}_B , \bar{m}_R/\bar{m}_B , and \bar{m}_G/\bar{m}_B , can be estimated by fitting CMFs to the cone spectral sensitivity data shown in Fig. 2. However, since the cone spectral sensitivity data are defined for 2-deg viewing conditions and the CMFs for 10-deg, we employed an intermediate step of fitting the 2-deg data to the Stiles and Burch [8] 2-deg CMFs. Figure 2 shows the linear combinations of the Stiles and Burch 2-deg CMFs that best fit the mean L(ser¹⁸⁰) deuteranope data (red circles), L(ala¹⁸⁰) deuteranope data (yellow and orange squares), and L1M2/L2M3 protanope data

(green diamonds) of Sharpe et al. [11]. An overall population mean for the L-cone spectral sensitivity function was derived by averaging the L (ser¹⁸⁰) and L(ala¹⁸⁰) fits after weighting them in ratio of 62 L(ser¹⁸⁰) to 38 L(ala¹⁸⁰), which is the ratio believed to correspond to normal population incidences (see Table 1 of Reference 2).

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

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

S-Cone Fundamental

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

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

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

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

LMS Transformation Matrix

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

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

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

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

XYZ Transformation Matrix

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

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

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

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

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

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

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

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

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

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

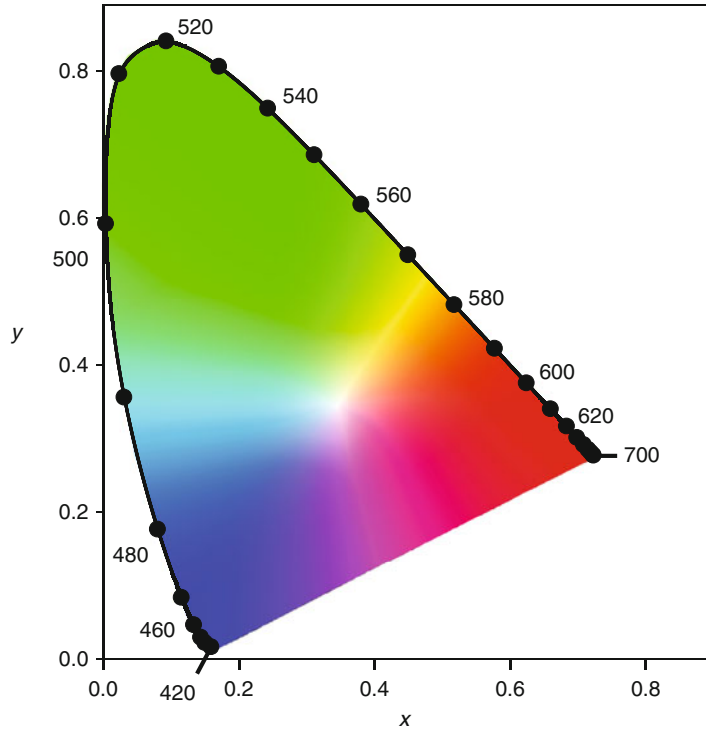
$$x(\lambda) = \frac{\bar{x}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \tag{7}$$

and

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

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

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



Cross-References

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

References

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

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

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

CIE Special Metamerism Index: Change in Observer

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

Synonyms

CIE Standard Deviate Observer

Definition

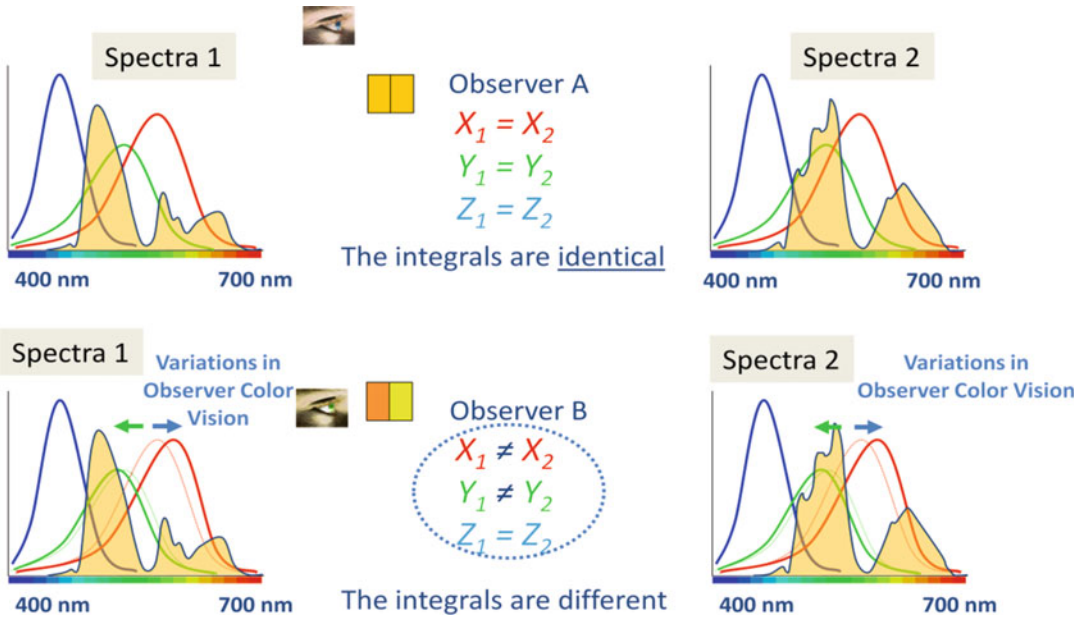
Special Metamerism Index: Change in Observer refers to a method for evaluating the average of and the range of color mismatches for metameric color pairs when test observers with normal color vision are substituted for a reference observer (i.e., a CIE standard colorimetric observer). This method was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

Overview

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

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

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



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

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

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

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

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

Standard Deviate Observer

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

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

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

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

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

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

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

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

Procedure for Computing CIE Observer Metamerism Index

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

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

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

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

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

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

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

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

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

standard deviate observer are obtained by using Eq. 4.

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

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

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

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

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

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

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

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

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

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

Deriving 95 % Confidence Ellipse

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Effect of Age on Color Mismatches

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

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

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

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

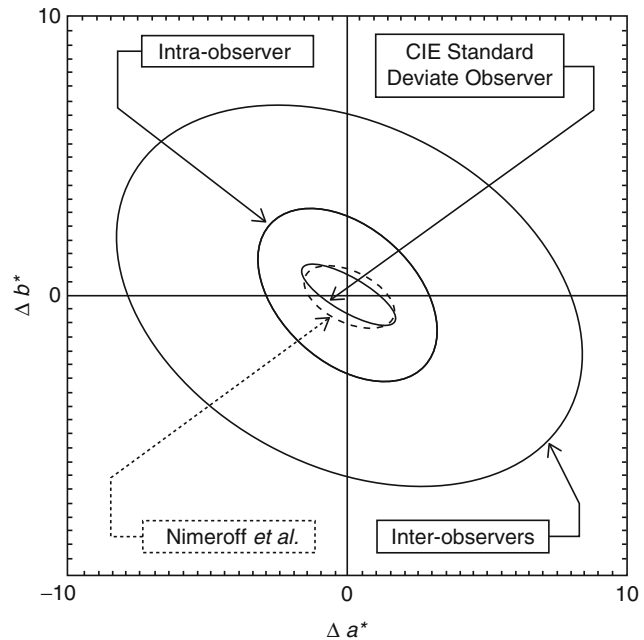
Evaluation of the CIE Observer Metamerism Index

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

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

CIE Special Metamerism Index: Change in Observer,

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



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

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

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

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

Future Directions

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

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

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

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

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

Cross-References

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

References

1. CIE Publication 80: Special metamerism index: Change in observer. Central Bureau of the CIE, Vienna (1989)
2. Nayatani, Y., Takahama, K., Sobagaki, H.: A proposal of new standard deviate observers. *Color Res. Appl.* **8**(1), 47–56 (1983)
3. Takahama, K., Sobagaki, H., Nayatani, Y.: Prediction of observer variation in estimating colorimetric values. *Color Res. Appl.* **10**(2), 106–117 (1985)
4. Nimeroff, I., Rosenblatt, J.R., Dannemiller, M.C.: Variability of spectral tristimulus values. *J. Opt. Soc. Am. A* **52**, 685–691 (1962)

CIE Standard Deviate Observer

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

CIE Standard Illuminant A

- [CIE Standard Illuminants and Sources](#)

CIE Standard Illuminant D65

- [CIE Standard Illuminants and Sources](#)

CIE Standard Illuminants and Sources

János Schanda
Veszprém, Hungary

Synonyms

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

Definitions

CIE Standard Illuminants

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

Note 1: These illuminants are intended to represent:

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

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

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

CIE Standard Sources

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

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

Overview

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

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

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

CIE Standard Illuminant A

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

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

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

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

CIE Standard Illuminant D65

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

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

Short History

CIE Standard Illuminant A

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

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

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

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

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

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

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

Daylight Illuminants

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

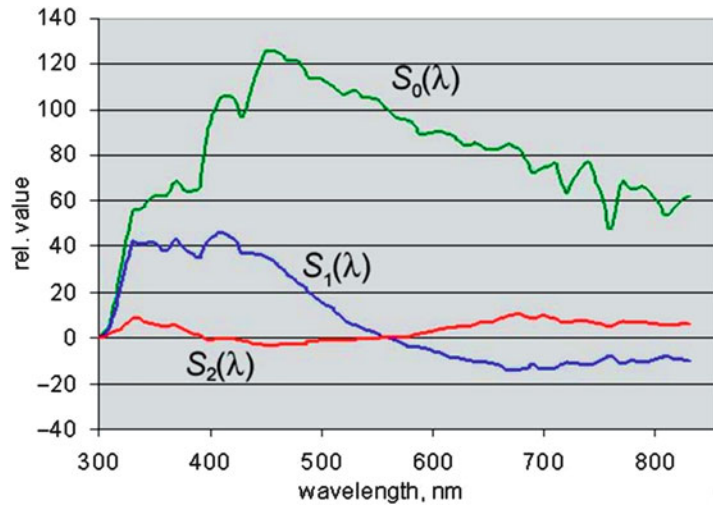
With the introduction of optical brighteners, it became necessary to define the SPD also in the near ultraviolet. The CIE accepted a recommendation by Judd and coworkers [7] to describe phases of daylight [8]. These authors found that although daylight is highly variable, the chromaticities of different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

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

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

CIE Standard Illuminants and Sources,

Fig. 1 Characteristic vectors used to reconstitute phases of daylight



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

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

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

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

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

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

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

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

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

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

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

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

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

Further CIE Illuminants

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

CIE Standard Source

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

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

Cross-References

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

References

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

CIE Standard Source A

- [CIE Standard Illuminants and Sources](#)

CIE Tristimulus Values

Ming Ronnier Luo

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

School of Design, University of Leeds,
Leeds, UK

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

Synonyms

XYZ

Definition

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

Background

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

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

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

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

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

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

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

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

Cross-References

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

References

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

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

János Schanda
Veszprém, Hungary

Synonyms

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

Definitions

Uniform Chromaticity Scale Diagram (UCS Diagram)

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

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

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

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

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

Equivalent definition:

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

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

where

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

Uniform Color Space

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

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

$$u^* = 13 L^* (u' - u'_n) \quad (7)$$

$$v^* = 13 L^* (v' - v'_n) \quad (8)$$

where

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

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

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

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

Correlates of Lightness, Saturation, Chroma, and Hue

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

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

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

Color Differences

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

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

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

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

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

Overview

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

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

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

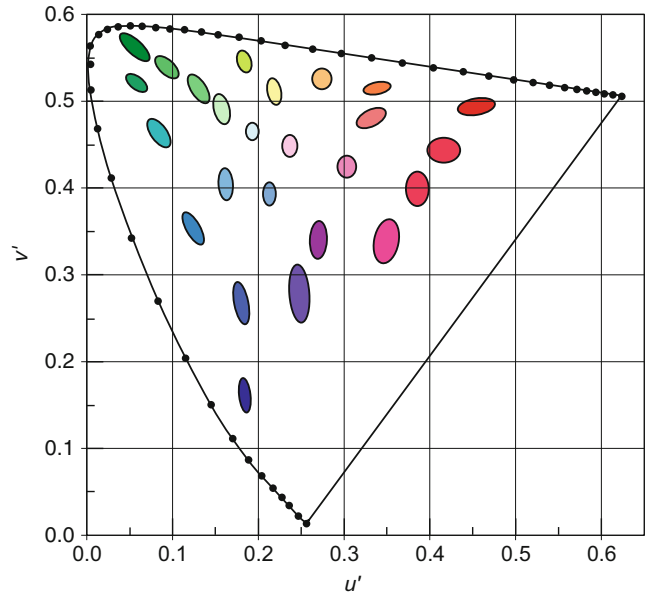
thus, $u' = u$, and $v' = 3v/2$ [5].

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

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

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

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



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

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

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

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

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

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

Cross-References

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

References

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

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

CIE Whiteness

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

Definition

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

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

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

$$W_{10} = Y + 800 (x_{n,10} - x_{10}) + 1,700 (y_{n,10} - y_{10}), \quad (2)$$

where x_{10}, y_{10} and $x_{n,10}, y_{n,10}$ are the chromaticity coordinates of the sample and the illuminant, respectively.

The tint coefficient T_w or $T_{w,10}$, given by the following formulae, is zero for a sample without reddishness or greenishness. For the CIE 1931 standard colorimetric observer,

$$T_w = 1,000 (x_n - x) - 650 (y_n - y), \quad (3)$$

and for the CIE 1964 supplementary standard colorimetric observer,

$$T_{w,10} = 900 (x_{n,10} - x_{10}) - 650 (y_{n,10} - y_{10}), \quad (4)$$

where x_{10}, y_{10} and $x_{n,10}, y_{n,10}$ are the chromaticity coordinates of the sample and the illuminant, respectively. The tint coefficient allows two samples with the sample whiteness index, W or W_{10} , but with different hue to be distinguished. The more positive the value of T_w or $T_{w,10}$, the greater is the indicated greenishness, and the more negative, the greater the reddishness [2].

The higher the value of W or W_{10} , the higher the whiteness of the sample. However, the CIE standard includes the following restriction:

The application of the formulae is restricted to samples that are called 'white' commercially, that do not differ much in color and fluorescence, and that are measured on the same instrument at nearly the same time; within these restrictions, the formulae provide relative, but not absolute, evaluations of whiteness that are adequate for commercial use, when employing measuring instruments having suitable modern and commercially available facilities. The following boundaries are proposed for the application of whiteness formulae:

$$40 < W \text{ or } W_{10} < 5Y - 280; \quad -3 < T_w \text{ or } T_{w,10} < +3, \quad (5)$$

where Y is the tristimulus value of the sample.

Development of the CIE Whiteness Formula

Whiteness is a commercially important property of color appearance in a number of industries, most notably those of textiles and paper. According to Wyszecki and Stiles, white is the attribute of a visual sensation according to which a given stimulus appears to be void of any hue and grayness [3]. Alternatively, the percept of whiteness is caused by a combination of high lightness

and lack of yellowness [4]. MacAdam published the first instrumental method for the assessment of whiteness in 1934 [5], and by the 1960s, over 100 different methods to predict perceptual whiteness had been proposed. Although color is a three-dimensional percept, whiteness has traditionally been measured using a univariant metric. Hayhurst and Smith suggest that the reason for this is that the color of white textile or paper is often influenced by the quantity of a specific (single) impurity in it [4]; the white color therefore falls on a line in three-dimensional color space, albeit one that is curved. However, especially when different products are being compared, it is important to have a metric that is at least dependent upon all three dimensions of color space even if the metric is condensed to a single number, as is the case with the CIE whiteness method. Many of the early whiteness formulae considered only one or two of the three dimensions of color and therefore had limited general applicability.

The foundations for a modern and effective whiteness formula began with Ganz [6] who proposed a new generic formula for whiteness W_G in 1972 with the following form:

$$W_G = DY + Px + Qy + C, \quad (6)$$

where Y and x , y are the Y tristimulus value (sometimes called luminous reflectance) and chromaticity coordinates, respectively, of the sample, and D , P , Q , and C are coefficients that could be varied to adjust for various parameters including the illuminant, the observer, and the hue preference. The coefficient of D was set to unity and $C = P_{x,n} + Q_{y,n}$. Ganz suggested three pairs of values for P and Q each corresponding to a different hue preference. The formula was linear in chromaticity space because experience in producing white scales had shown that the chromaticities of uniformly spaced samples turn out to be approximately equidistant in this space [7].

A round-robin test [7], organized by a technical committee (TC-1.3) of the CIE in the 1970s, showed that there are individuals with extreme hue preference for whiteness. This suggests that a single whiteness formula cannot reproduce the

whiteness preferences of a sizable portion of the population and this is why Ganz originally suggested three pairs of values for P and Q . The CIE adopted a form of Eq. 6 with values of P and Q corresponding to neutral hue preference, and the CIE equation is normally written in the forms of Eqs. 1 and 2. The use of unity for the coefficient D has the implication that the whiteness of a perfect reflecting diffuser would be 100. Samples containing fluorescent whitening agents may have W or $W_{10} \gg 100$. The difference that is just perceptible to an experienced visual assessor is about three CIE whiteness units [4].

Ganz and Griesser developed a generic formula for the instrumental evaluation of tint [8]. Lines of equal tint run approximately parallel to the line of dominant wavelength 470 nm (467.6 and 464.7 nm, respectively, for the 1931 and 1964 CIE standard observers) in the chromaticity diagram. Figure 1 shows the line of dominant wavelength 470 nm and also two lines of iso-whiteness according to the CIE whiteness equation. The CIE adopted the Ganz-Griesser tint formula as part of the CIE whiteness standard.

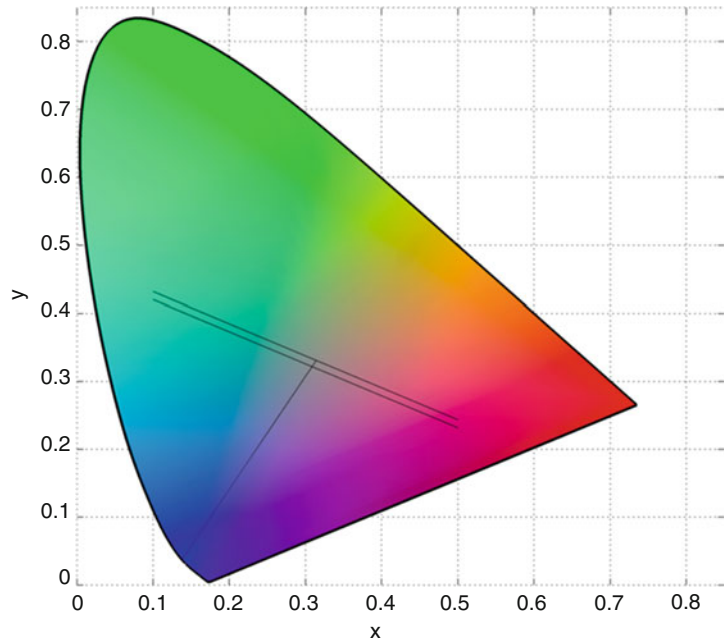
The whiteness formula should be used only for the comparison of samples that are commercially white and do not differ much in color or fluorescence. The samples should be measured on the same instrument at the same time. The light source in the spectrophotometer used to measure the reflectance factors should match, as closely as possible, the illuminant D65. The requirement for instruments that could deliver consistent light output (including in the near UV) that closely matches illuminant D65 led to developments in instrument-calibration techniques toward the end of the twentieth century.

Further Considerations and Future Directions

Since the introduction of the CIE whiteness equation, the approximately uniform color space, CIELAB, has been developed. This led some practitioners to want to calculate CIE whiteness in CIELAB space directly. Ganz and Pauli [9] developed the following equations that

CIE Whiteness,**Fig. 1** CIE 1964

chromaticity diagram showing the line at dominant wavelength of 465 nm and two lines of iso-whiteness. The *upper* iso-whiteness line shows the locus of points with $W_{10} = 100$, passing through the white point; each point on this line has the same whiteness but a different tint. The *lower* iso-whiteness line shows the locus of points with $W_{10} = 120$. Whiteness generally increases from the *yellowish* part of the diagram to the *bluish* part



approximate the CIE whiteness and tint equations in the CIE (1976) $L^*a^*b^*$ space for the 1964 standard observer:

$$W_{10} = 2.41L^* - 4.45b^*[1 - 0.0090(L^* - 96)] - 141.4, \quad (7)$$

and

$$T_{w,10} = -1.58a^* - 0.38b^*, \quad (8)$$

where L^* , a^* , and b^* are the CIELAB coordinates of the sample. The upper limit $W_{10} < 5Y - 280$ was transformed by a linear approximation yielding $W_{10} < 10.6 L^* - 852$. The other limits $W_{10} > 40$ and $-3 < T_{w,10} < 3$ remained unchanged.

The CIE whiteness formula has been found to correlate with visual assessments for many white samples where the tint is similar and where the level of fluorescence is comparable [10]. However, the level of agreement is much less when there are different tints and levels of fluorescence [10]. Observers frequently give the highest whiteness estimation to fluorescent-whitened samples with a hint of red, blue, or green. This is called “preferred white.” Uchida considered data from

26 observers who evaluated the whiteness of 49 fluorescent-whitened samples and devised a new whiteness formula W_u where

$$W_u = W_{10} - 2(T_{w,10})^2 \quad (9)$$

The Uchida equation [10] was recommended for samples where $40 < W_{10} < 5Y - 275$ and was found to give better agreement with the visual results than the CIE equation. In a recent study [11], 22 observers were asked to rank 20 samples (with low to high CIE whiteness indices). Results showed a significant consistency between the variations in the ordering decisions of the observers for the white samples with low CIE whiteness index but a high level of disagreement between the observers for the whiter samples.

Recently the generic form of the CIE whiteness equation (Eq. 2) with modified coefficients has been shown to be able to predict the whiteness of human teeth even though many of the samples would not be considered to be commercially white in the textile, plastic, and paper industries in which the CIE whiteness equation has traditionally been applied [12].

Cross-References

- [Anchoring Theory of Lightness](#)
- [Blackbody and Blackbody Radiation](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)

References

1. CIE, Colorimetry Second Edition: The Evaluation of Whiteness, pp. 36–38. CIE Publication 15.2 (1986)
2. Hunt, R.W.G.: The Reproduction of Colour. Chichester, UK: Wiley (2004)
3. Wyszecki, G., Stiles, W.S.: Color Science: Concepts and Methods, Quantitative Data and Formulae. Hoboken, USA: Wiley (1982)
4. Hayhurst, R., Smith, K.J.: Instrumental evaluation of whiteness. *J. Soc. Dye. Colour.* **111**, 263–266 (1995)
5. MacAdam, D.L.: The specification of whiteness. *J. Opt. Soc. Am.* **24**, 188–191 (1934)
6. Ganz, E.: Whiteness measurement. *J. Color Appearance* **1**, 33–41 (1972)
7. Ganz, E.: Whiteness formulas: a selection. *Appl. Optics* **18**(7), 1073–1078 (1979)
8. Ganz, E., Griesser, R.: Whiteness: assessment of tint. *Appl. Optics* **20**(8), 1395–1396 (1981)
9. Ganz, E., Pauli, H.K.A.: Whiteness and tint formulas of the Commission Internationale de l’Eclairage: approximations in the $L^*a^*b^*$ color space. *Appl. Optics* **34**(16), 2998–2999 (1995)
10. Uchida, H.: A new whiteness formula. *Color Res. Appl.* **23**(4), 202–209 (1998)
11. Jafari, R., Amirshahi, S.H.: Variation in the decisions of observers. *Color. Technol.* **124**(2), 127–131 (2008)
12. Luo, W., Westland, S., Ellwood, R., Pretty, I., Cheung, V.: Development of a whiteness index for dentistry. *J. Dent.* **37**, e21–e26 (2009)

CIE94, History, Use, and Performance

Manuel Melgosa

Optics Department, University of Granada,
Granada, Spain

Synonyms

CIE 1994 (ΔL^* ΔC^*_{ab} ΔH^*_{ab}); ΔE^*_{94} ; Hue difference, delta H

Definition

The CIE94 color-difference formula [1] was developed by the CIE Technical Committee 1–29 “Industrial Color Difference Evaluation” chaired by Dr. D.H. Alman (USA). CIE94 is a CIELAB-based color-difference formula where CIELAB lightness, chroma, and hue differences are appropriately weighted by very simple “weighting functions” correcting CIELAB’s lack of visual uniformity, as well as by “parametric factors” accounting for the influence of illuminating/viewing conditions in color-difference evaluation. CIE94 may be considered a simplified version of the CMC color-difference formula [2], as well as a predecessor of currently CIE-recommended color-difference formula, the joint ISO/CIE standard CIEDE2000 [3, 4].

Overview

In 1976, CIE recommended the CIELUV and CIELAB color spaces with their corresponding color-difference formulas, defined as Euclidean distances in such spaces. These two color spaces are only approximately uniform, and it was indicated that the use of different weighting for lightness, chroma, and hue differences may be necessary in different practical applications [5]. CIELAB has been widely accepted in many industrial applications [6], and in the past recent years, several CIELAB-based color-difference equations have been proposed to improve its performance. Among these formulas, we can mention, in chronological order, JPC79 [7], CMC [2], BFD [8], and CIE94. Nowadays, it is generally agreed that most of these formulas significantly improved CIELAB [9].

Assuming that subindices 1 and 2 indicate the two samples in a color pair, the next equations define the CIE94 color-difference formula, designated as $\Delta E^*_{94}(k_L:k_C:k_H)$:

$$\Delta E_{94}^*(k_L : k_C : k_H) = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H S_H}\right)^2} \quad (1)$$

where the CIELAB lightness, chroma, and hue differences are given by

$$\Delta L^* = L_{1,1}^* - L_{2,1}^* \quad (2)$$

$$\Delta C_{ab}^* = C_{ab,1}^* - C_{ab,2}^* \quad (3)$$

$$\Delta H_{ab}^* = 2 \sqrt{C_{ab,1}^* C_{ab,2}^*} \sin\left(\frac{\Delta h_{ab}}{2}\right) \quad (4)$$

$$\Delta h_{ab} = h_{ab,1} - h_{ab,2}. \quad (5)$$

The “weighting functions” for lightness (S_L), chroma (S_C), and hue (S_H) are defined as

$$S_L = 1 \quad (6)$$

$$S_C = 1 + 0.045 \left(C_{ab,1}^* C_{ab,2}^*\right)^{1/2} \quad (7)$$

$$S_H = 1 + 0.015 \left(C_{ab,1}^* C_{ab,2}^*\right)^{1/2}, \quad (8)$$

and the “parametric factors” k_L , k_C , k_H are set as 1.0 (i.e., do not affect the total color difference computation) under the next illuminating/viewing conditions, usually known as “reference conditions”:

Illumination: D65 source
 Illuminance: 1000 lx
 Observer: Normal color vision
 Background field: Uniform, neutral gray with $L^* = 50$
 Viewing mode: Object
 Sample size: Greater than 4 degrees
 Sample separation: Direct edge contact
 Sample color-difference magnitude: Lower than 5.0 CIELAB units
 Sample structure: Homogeneous (without texture)

In the textile industry, it is common practice to set the lightness parametric factor k_L to 2. Although the experimental conditions leading

to this parametric correction to lightness difference are not yet well understood, it introduces important improvements in the performance of CIE94, which in this case must be designated as $\Delta E_{94}^*(2:1:1)$ or CIE94 (2:1:1).

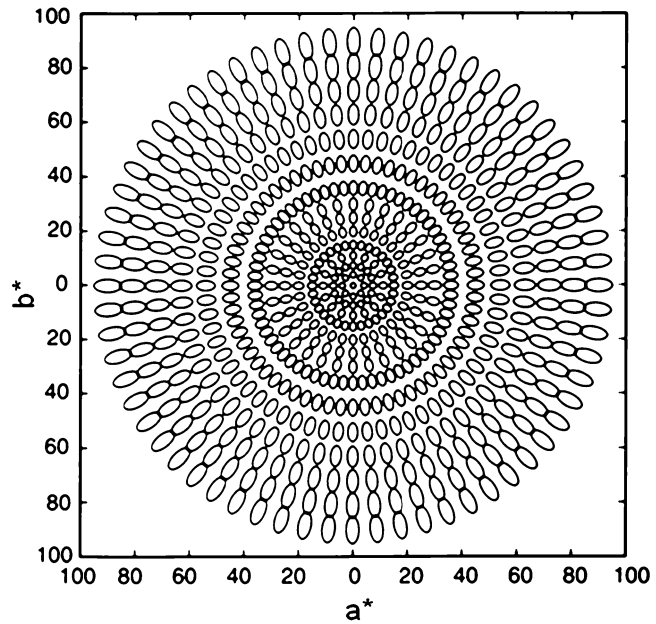
For a constant ΔE_{94}^* value, Eq. 1 approximately represents [10] an ellipsoid with semiaxis lengths given by the denominators $k_L S_L$, $k_C S_C$, $k_H S_H$. Under reference conditions, the semiaxis lengths S_L , S_C , S_H are often denominated lightness, chroma, and hue tolerances, respectively. Equation 6 indicates that lightness tolerance is the same for all color centers. However, Eq. 7 indicates that chroma tolerances increase with chroma; that is, human sensitivities to chroma differences are smaller for color centers with higher chroma, as earlier pointed out by McDonald [11]. Similarly, Eq. 8 also indicates that hue tolerances/sensitivities increase/decrease with chroma. In summary, assuming small color differences [10], the loci of constant CIE94 differences in CIELAB color space can be represented by ellipsoids with constant lightness semiaxes, which sections in the a^* , b^* plane are ellipses with their major axes pointing to the origin (i.e., ellipses radially oriented). The major and minor semiaxes of these a^* , b^* ellipses linearly increase with the chroma of the ellipse centers (Fig. 1).

Development

The development of CIE94 [12, 13] began with a selection of experimental visual datasets meeting the next main conditions: statistical significance (i.e., to represent a population average with its corresponding uncertainty), well-documented experimental conditions, and use of object color specimens. CIE TC 1–29 decided to use only three experimental datasets: Witt [14, 15], Luo and Rigg [16], and RIT-DuPont [17, 18]. The main goal was to use these datasets to find the best

CIE94, History, Use, and Performance,

Fig. 1 CIE94 color-tolerance ellipses (or contours of approximately constant CIE94 units) in the a^* , b^* plane (Figure from R. S. Berns [13], p. 121, graph courtesy of S. Quan)



weighting functions S_L , S_C , S_H correcting CIELAB. The analyses also considered the characteristics of the previous CMC color-difference formula, which was an ISO standard in textiles [19] and was successfully employed in different industries.

As indicated by Eqs. 7 and 8, it was found that simple linear chroma functions described the main trends in the experimental datasets analyzed. Both the lightness dependence of lightness tolerances and the hue angle dependence of hue tolerances proposed by the CMC color-difference formula were found not robust trends in the experimental datasets analyzed [13], and therefore they were disregarded in CIE94. It is possible that these two corrections proposed by CMC were influenced by some specific parametric factors. Anyway, it can be said that CIE94 was a conservative approach incorporating only the main corrections to CIELAB. Currently, the CIEDE2000 color-difference formula has incorporated additional corrections to CIELAB. For example, CIEDE2000 proposes a V-shaped function for the S_L function, which is different to both the lightness function proposed by CMC and the simple $S_L = 1$ adopted by CIE94.

CIE94 was the first CIE-recommended color-difference formula incorporating the influence of

illuminating/viewing experimental conditions in color-difference evaluations through the use of the so-called parametric factors k_L , k_C , k_H (see Eq. 1).

The work carried out by CIE TC 1–29 finished with the proposal of CIE94 and some guidelines for coordinated future work on industrial color-difference evaluation [20]. These guidelines updated those earlier given by Robertson [21], proposing a new set of 17 color centers to be studied in future research, considering the effects of changes from the “reference conditions,” and suggesting the development of a database of color-difference visual responses.

Performance

From the combined dataset employed at CIEDE2000 development [22], the performance of different color-difference formulas has been tested using the *STRESS* index [23]. Low *STRESS* values (always in the range 0–100) indicate better color-difference formula performance. *STRESS* values for CIELAB, CIE94, and CIEDE2000 are 43.9, 32.1, and 27.5, respectively [24]. As we can see, the improvement achieved by CIE94 upon CIELAB was considerably higher than the one

achieved by CIEDE2000 with respect to CIE94. From *STRESS* values, it can be also concluded that in CIE94 the S_C function (which was also adopted in CIEDE2000) is a much more important correction to CIELAB than the S_H function [24]. Besides CIE94 being proposed for object colors, it has been reported that this formula also performed satisfactorily for self-luminous color datasets [25].

Curiously, Eqs. 7 and 8 involve the geometrical mean of the CIELAB chroma of the two samples in the color pair, in place of the more simple arithmetical mean proposed in other color-difference formulas, for example, CIEDE2000. Strictly speaking, this implies that the locus of constant CIE94 differences with respect to a given color center is not an ellipsoid/ellipse, although deviations from ellipsoidal/elliptical contours can be considered negligible in most practical situations. CIE94 color differences computed using the geometrical and arithmetical means of the CIELAB chroma of the samples in the color pair are slightly different, in particular for colors with very low chroma [26].

In comparison with other recent formulas like JPC79, CMC, BFD, or CIEDE2000, it can be said that CIE94 was relevant because it was a very simple and versatile color-difference formula accounting for the main robust trends found in reliable color-difference visual datasets. CIE94 just proposes to use simple corrections to CIELAB provided by the weighting functions S_L , S_C , S_H plus consideration of the influence of the illuminating/viewing conditions using the parametric factors k_L , k_C , k_H . After the CIE94 proposal, CIE TC 1–47 continued further work leading to the last CIE-recommended color-difference formula, CIEDE2000, which significantly improved CIE94 for the experimental datasets used in its development [22].

Cross-References

- CIEDE2000, History, Use, and Performance
- CIELAB

References

1. CIE Publication 116: Industrial Colour-Difference Evaluation. CIE Central Bureau, Vienna (1995)
2. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. *J. Soc. Dye. Colour.* **100**, 128–132 (1984)
3. CIE Publication 142: Improvement to Industrial Colour-Difference Evaluation. CIE Central Bureau, Vienna (2001)
4. CIE S 014-6/E:2013: Colorimetry – Part 6: CIEDE2000 Colour-Difference Formula. CIE Central Bureau, Vienna (2013)
5. CIE Publication 15.2: Colorimetry, 2nd edn, Note 9, p. 33. CIE Central Bureau, Vienna (1986)
6. Kuehni, R.G.: Industrial color-difference: progress and problems. *Color. Res. Appl.* **15**, 261–265 (1990)
7. McDonald, R.: Industrial pass/fail colour matching. Part III: development of a pass/fail formula for use with instrumental measurement of colour difference. *J. Soc. Dye. Colour.* **96**, 486–497 (1980)
8. Luo, M.R., Rigg, B.: BFD(l:c) colour-difference formula. Part 1 – development of the formula. *J. Soc. Dye. Colour.* **103**, 86–94 (1987)
9. Melgosa, M.: Testing CIELAB-based color-difference formulas. *Color. Res. Appl.* **25**, 49–55 (2000)
10. Brill, M.H.: Suggested modification of CMC formula for acceptability. *Color Res. Appl.* **17**, 402–404 (1992)
11. McDonald, R.: The effect of non-uniformity in the ANLAB color space on the interpretation of visual colour differences. *J. Soc. Dye. Colour.* **90**, 189–198 (1974)
12. Berns, R. S.: The mathematical development of CIE TC 1–29 proposed color difference equation: CIELCH. In: Nemcsics, A., Schanda, J. (eds.) AIC (Association Internationale de la Couleur). 1993. AIC Color 93, Proceedings of the 7th Congress, Budapest, 13–18 June 1993, 3 vols. Hungarian National Colour Committee, Budapest. (1993)
13. Berns, R.S.: Billmeyer and Saltzman's Principles of Color Technology, pp. 120–121. Wiley, New York (2000)
14. Witt, K., Döring, G.: Parametric variations in threshold color-difference ellipsoid for green painted samples. *Color. Res. Appl.* **8**, 153–163 (1983)
15. Witt, K.: Three-dimensional threshold color-difference perceptibility in painted samples: variability of observers in four CIE color regions. *Color. Res. Appl.* **12**, 128–134 (1987)
16. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colors. *Color. Res. Appl.* **11**, 25–42 (1986)
17. Alman, D.H., Berns, R.S., Snyder, G.D., Larsen, W. A.: Performance testing of color-difference metrics using a color tolerance dataset. *Color. Res. Appl.* **14**, 139–151 (1989)
18. Berns, R.S., Alman, D.H., Reniff, L., Snyder, G.D., Balonon-Rosen, M.R.: Visual determination of

- suprathreshold color-difference tolerances using probit analysis. *Color. Res. Appl.* **16**, 297–316 (1991)
19. ISO 105-J03: Textiles: Test for Colour Fastness – Part 3: Calculation of Colour Differences. ISO, Geneva (2009)
 20. Witt, K.: CIE guidelines for coordinated future work on industrial colour-difference evaluation. *Color. Res. Appl.* **20**, 399–403 (1995)
 21. Robertson, A.R.: CIE guidelines for coordinated research on color-difference evaluation. *Color. Res. Appl.* **3**, 149–151 (1978)
 22. Luo, M.R., Cui, G., Rigg, B.: The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color. Res. Appl.* **26**, 340–350 (2001)
 23. García, P.A., Huertas, R., Melgosa, M., Cui, G.: Measurement of the relationship between perceived and computed color differences. *J. Opt. Soc. Am. A* **24**, 1823–1829 (2007)
 24. Melgosa, M., Huertas, R., Berns, R.S.: Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. *J. Opt. Soc. Am. A* **25**, 1828–1834 (2008)
 25. Melgosa, M., Hita, E., Pérez, M.M., El Moraghi, A.: Sensitivity differences in chroma, hue, and lightness from several classical threshold datasets. *Color. Res. Appl.* **20**, 220–225 (1995)
 26. Melgosa, M., Huertas, R., Hita, E., Benítez, J. M.: Diferencias de color CIE94 con/sin un estímulo de referencia. In: *Proceedings VI Congreso Nacional del Color* (ISBN 84-699-9187-6), Area de Nutrición y Bromatología, pp. 117–118. Universidad de Sevilla (2002)

CIECAM02

Nathan Moroney
Hewlett-Packard Laboratories, Palo Alto, CA,
USA

Synonyms

[CIE color appearance model 2002](#)

Definition

CIECAM02 is a color appearance model that provides a viewing condition specific method for transforming between tristimulus values and perceptual attribute correlates. This model was first published [1] in 2002 by Division 8 of the

International Commission on Illumination (CIE). CIECAM02 was developed for use in color management systems and was based on the previously published CIECAM97s color appearance model [2, 3]. The model provides a number of parameters for defining a viewing condition and also inverse equations for transforming perceptual attribute correlates back to tristimulus values for a given set of viewing conditions. In this way, CIECAM02 can be used to transform perceptual attribute correlates, such as lightness, chroma, and hue, across different viewing conditions.

Overview

A color appearance model [4, 5] transforms between colorimetry, which specifies if stimuli match, and perceptual attribute correlates, which are scales of lightness, chroma, and hue. To do so, CIECAM02 provides a set of viewing condition parameters in order to model specific color appearance phenomena, such as chromatic adaptation and simultaneous contrast. The resulting perceptual attribute correlates can then be used in research and engineering applications requiring a viewing condition independent color representation, such as color calibration of color printers.

The viewing conditions for CIECAM02 consist of the background, the adapting field, the surround, the white point, and the luminance of the adapting field. Figure 1 shows an example of a color stimulus, a background, and an adapting field. Note that this example is of a real-world situation of viewing a reflectance print. The central stimulus is a 2° region that corresponds to a portion of the print the size of a thumb viewed at arm's length. The background is the 10° region surrounding the stimulus and is roughly the size of a fist viewed at arm's length. The adapting field is everything else in the field of view.

The luminance of the adapting field can be measured directly with an illuminance meter. To incorporate a gray-world assumption and convert to luminance, this is then divided by 5π . The surround setting for the model is categorical and follows roughly the specific application. Dark surrounds are those with no ambient illumination



CIECAM02, Fig. 1 An example of a colored stimulus, background, and adapting field. The stimulus subtends 2° or roughly the size of a thumb viewed at arm's length. The background subtends 10° or approximately the size of a fist viewed at arm's length. The adapting field is everything

else in the field of view. The stimulus is shown as the *central smaller black circle*. The background is shown as the *larger black circle*. The adapting field is everything outside of the *larger black circle*

or viewing film projected in a darkened room. Dim surrounds are those in which the ambient illumination is not zero but is also less than 20 % of the scene, print, or display white point, such as home viewing of television with low light levels. Average surround is ambient illumination greater than 20 % of the scene, print, or display white point, such as viewing of surface colors in a light booth. The CIECAM02 model has a set of constants associated with each surround.

Finally the model has an associated white for all calculations. The selection of a white point is a subtle topic and the model suggests two approaches to this issue. First is to use an adopted white point or a computational white point for all calculations. An adopted white point is a fixed value, such as one based on a standard viewing condition or to ensure a specific final mapping for the white point. Second is an adapted white or the white point adapted by the human visual system for a given set of viewing conditions. An adapted white point is one which attempts to as closely as possible match the state of adaptation for a human observer. Note that adapted white points may require experimentation to infer their value, such as for a novel set of viewing conditions or situations with multiple illuminants. In cases where it

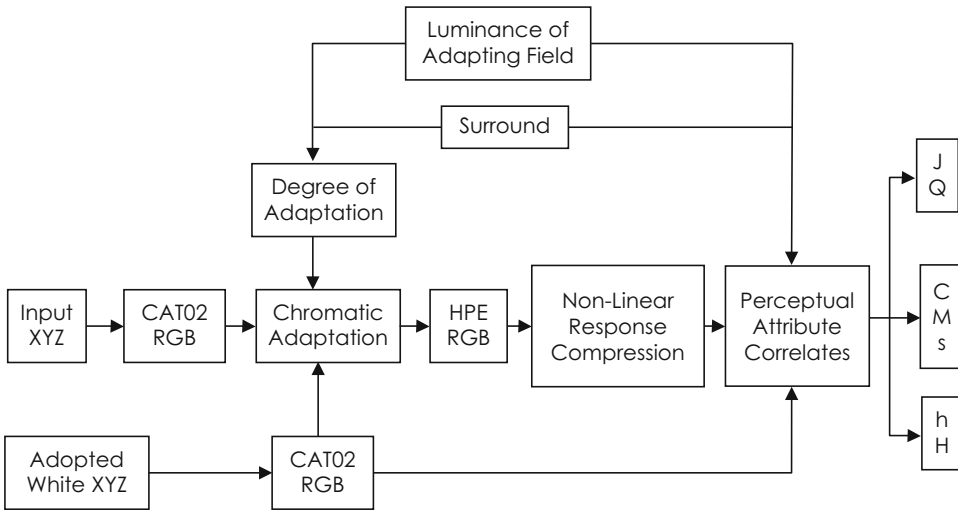
is not possible to determine the adapted white point, use of an adopted or assumed white point can be convenient.

Given a set of viewing conditions and their associated parameters, it is then possible to compute the perceptual attribute correlates. Figure 2 shows an overall flowchart of the CIECAM02 model starting with input tristimulus values and adopted white point tristimulus values on the left.

The first step is to apply a matrix to convert the XYZ values to the CAT02 RGB space. The space was selected as a preferred color space to perform chromatic adaptation. This matrix can be written:

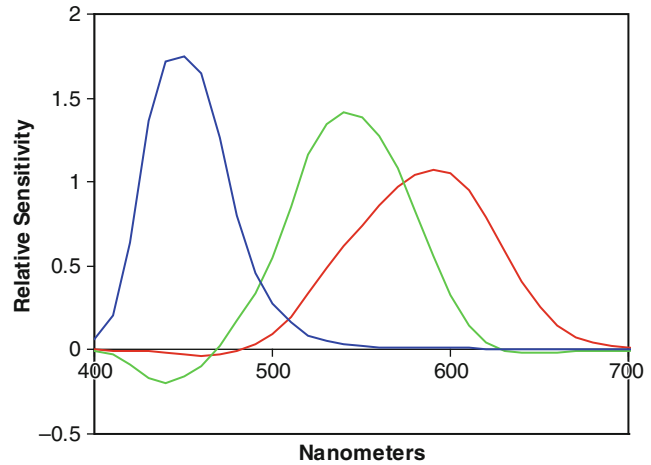
$$M_{CAT02} = \begin{bmatrix} 0.7328 & 0.4296 & -0.1624 \\ -0.7036 & 1.6975 & 0.0061 \\ 0.0030 & 0.0136 & 0.9834 \end{bmatrix} \quad (1)$$

The result of applying this matrix to 1931 color matching functions is shown in Fig. 3. These curves can be compared to the CIE color matching functions, and it can be seen that these curves are qualitatively more narrow or sharpened. This matrix was derived using a set of corresponding color training data but excluding highly chromatic light sources.



CIECAM02, Fig. 2 Overall flowchart of the computational steps for calculating perceptual attribute correlates given input tristimulus values and a specific set of viewing conditions

CIECAM02, Fig. 3 The CAT02 matrix shown as a set of *red*, *green*, and *blue* sensitivity curves computed by applying Eq. 1 to the CIE 1931 color matching functions



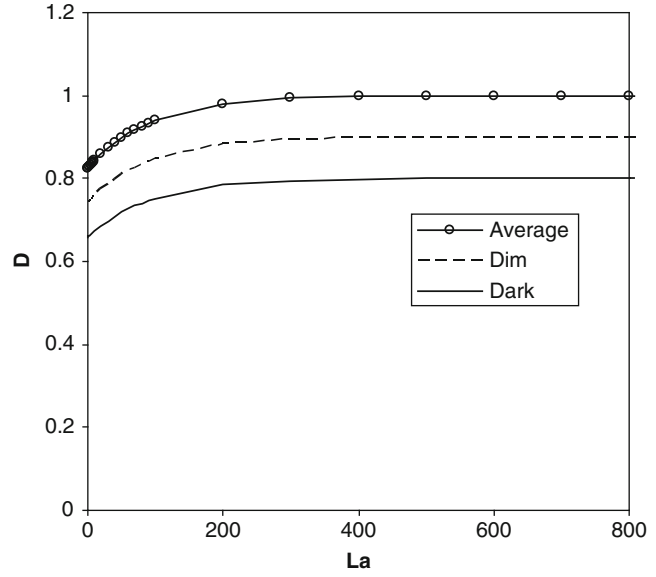
The degree of adaptation is the next step in the model and is calculated given the surround setting and the luminance of the adapting field. The specific calculation for incomplete adaptation or D is shown in Eq. 2. The variable L_A is the luminance of the adapting field and the value of F is a parameter that is computed from the surround setting.

$$D = F \left[1 - \left(\frac{1}{3.6} \right) e^{\left(\frac{-L_A - 42}{92} \right)} \right] \quad (2)$$

The results of using Eq. 2 with three different surround settings over luminance of adapting field values ranging from 0 to 800 cd/m^2 are shown in Fig. 4. Essentially the surround limits the degree of adaptation and increases with larger values of L_A . Complete adaptation is only achieved with the average surround with high L_A values.

Given the input stimulus converted to CAT02 RGB space and a degree of adaptation, it is then possible to compute the chromatic adaptation. This can be done according to Eqs. 3, 4, and 5

CIECAM02, Fig. 4 The degree of adaptation as computed for three surrounds and varying L_A . A D value of 1 is complete adaptation to the white point, while values less than one are for incomplete adaptation



below. The stimulus CAT02 values are shown as R, G, and B and the adopted or adapted white point as R_w , G_w , and B_w .

$$R_c = [D(100/R_w) + 1 - D]R \quad (3)$$

$$G_c = [D(100/G_w) + 1 - D]G \quad (4)$$

$$B_c = [D(100/B_w) + 1 - D]B \quad (5)$$

The next step in the calculation of the forward CIECAM02 model is shown as the box HPE RGB in the center of the flowchart in Fig. 2. This is the conversion of R_c , G_c , and B_c above to the Hunt-Pointer-Estevéz space. This can be done using a 3 by 3 matrix shown in Eq. 6:

$$\mathbf{M}_{HPE} = \begin{bmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0.00000 & 0.00000 & 1.00000 \end{bmatrix} \quad (6)$$

A plot of the corresponding red, green, and blue sensitivity curves for the Hunt-Pointer-Estevéz RGB space is shown in Fig 5. This graph shows the red and green curves with qualitatively broader and more overlapping than the red and green curves for the CAT02 RGB curves shown in Fig. 3. Effort was made during the formulation of

the model to derive a single RGB space for both the chromatic adaptation and for the nonlinear compression, but the results were generally worse and ultimately the final model made use of the two different RGB spaces.

Given the Hunt-Pointer-Estevéz RGB values, the next step is the nonlinear response compression. The specific nonlinearity used [6] in CIECAM02 is shown in Eqs. 7 through 9 and is shown graphically in Fig. 6. The R' , G' , and B' values are the HPE values as computed with Eq. 6, and the F_L value is a model parameter that is dependent on the viewing conditions.

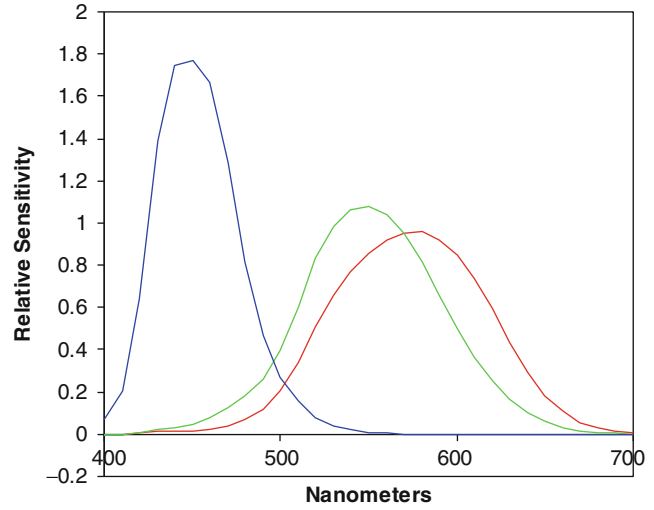
$$R'_a = \frac{400(F_L R'/100)^{0.42}}{[27.13 + (F_L R'/100)^{0.42}]} + 0.1 \quad (7)$$

$$G'_a = \frac{400(F_L G'/100)^{0.42}}{[27.13 + (F_L G'/100)^{0.42}]} + 0.1 \quad (8)$$

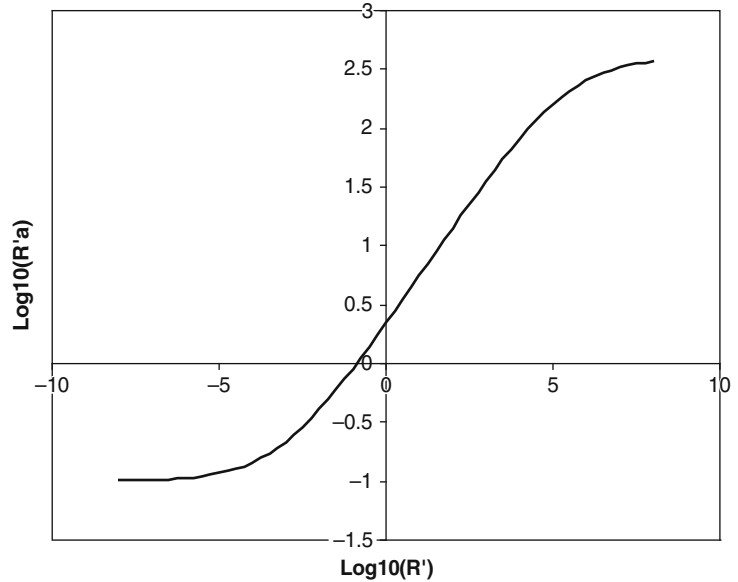
$$B'_a = \frac{400(F_L B'/100)^{0.42}}{[27.13 + (F_L B'/100)^{0.42}]} + 0.1 \quad (9)$$

The final steps in the forward CIECAM02 model are the computation of the perceptual attribute correlates. First an intermediate set of opponent

CIECAM02, Fig. 5 The HPE matrix shown as a set of *red*, *green*, and *blue* sensitivity curves computed by applying Eq. 6 to the CIE 1931 color matching functions



CIECAM02, Fig. 6 The post-adaptation nonlinear response compression function as computed using Eq. 7. Similar curves result for Eqs. 8 and 9 for the calculation of G'_a and B'_a



coordinates, a and b , are computed according to Eqs. 10 and 11. It should be emphasized that these values of a and b are preliminary and should not be used directly. The value of h or hue is computed using the arctangent of b divided by a . A table of constants is used to compute H or hue quadrature. The resulting H values for red, yellow, green, and blue are 100, 200, 300, and 400, respectively.

$$a = R'_a - 12G'_a/11 + B'_a/11 \quad (10)$$

$$b = (1/9)(R'_a + G'_a - 2B'_a) \quad (11)$$

The perceptual attribute correlates for lightness and brightness can then be calculated. First the achromatic signal or A is computed according to Eq. 12. The R'_a , G'_a , and B'_a values are the nonlinearly compressed values from Eqs. 7 through 9. Next the computation of J or lightness is shown in Eq. 13, while the computation of Q or brightness is shown in Eq. 14. The c and z values



CIECAM02, Fig. 7 *Light gray to deep blue constant hue angle gradients for CIELAB, shown at the top, and CIECAM02 shown at the bottom. The left- and right-hand*

side colors are the same XYZ values and have constant respective hue angles, but the top gradient for CIELAB tends to a purplish tone in the center of the gradient

are additional model parameters as computed based on the viewing conditions.

$$A = [2R'_a + G'_a + (1/20)B'_a - 0.305]N_{bb} \quad (12)$$

$$J = 100(A/A_w)^{c_z} \quad (13)$$

$$Q = (4/c)\sqrt{J/100}(A_w + 4)F_L^{0.25} \quad (14)$$

Finally given the correlates for hue, lightness, and brightness, it is possible to calculate the perceptual attribute correlates for chroma, colorfulness, and saturation. First a temporary variable t is computed using Eq. 15. Next C or chroma is calculated using Eq. 16. Colorfulness or M and saturation s can then be computed using Eqs. 17 and 18.

$$t = \frac{(50,000/13)N_c N_{cb} e_t (a^2 + b^2)^{1/2}}{R'_a + G'_a + (21/20)B'_a} \quad (15)$$

$$C = t^{0.9} \sqrt{J/100} (1.64 - 0.29^n)^{0.73} \quad (16)$$

$$M = CF_L^{0.25} \quad (17)$$

$$S = 100\sqrt{M/Q} \quad (18)$$

The result of the preceding calculations is then a set of perceptual attribute correlates for the given input values and viewing conditions. Note though that this does not define a rectangular coordinate system, such as the a^* and b^* values for CIELAB

[7]. Instead a set of correlates such as lightness, chroma, and hue must first be computed and used as polar coordinates. The rectangular coordinates can be computed using Eqs. 19 and 20. Similar coordinates can be calculated for lightness and saturation, with subscript s , and brightness and colorfulness, with subscript M .

$$a_c = C \cdot \cos(h) \quad (19)$$

$$b_c = C \cdot \sin(h) \quad (20)$$

The inverse CIECAM02 equations are beyond the scope of this entry, but reference 1 contains the full steps for inverting the above equations. This allows tristimulus values to be computed given input perceptual attribute correlates and viewing conditions.

The CIECAM02 color appearance model can then be used in situations requiring a viewing condition independent color encoding. It also could be considered in cases where CIELAB or CIELUV [7] lacks perceptual uniformity. For example, the CIELAB blue hue lacks hue constancy and tends toward purple as it approaches the neutral axis. This can be problematic in many circumstances, such as when gamut mapping colors from a display to a printer. The CIECAM02 space is considerably more uniform in this case. Two example gradients are shown in Fig. 7. The CIELAB gradient shown on the top has a clear tendency to purple as it goes to gray while the CIECAM02 gradient on the bottom does not. Both color spaces have constant hue angles for

these colors but CIECAM02 is significantly improved.

It is useful to further compare and contrast CIELAB and CIECAM02. CIELAB has as input the stimulus XYZ values and the white point XYZ values. CIECAM02 has as input the stimulus and white XYZ values and also luminance of the adapting field, the luminance of the background, and the surround setting. CIELAB has a chromatic adaptation transform that consists of a complete von Kries transform in XYZ space, while CIECAM02 has a complete or incomplete von Kries transform in CAT02 RGB space. CIELAB has a cube-root nonlinearity, while CIECAM02 uses a modified hyperbolic function as the nonlinearity. CIELAB uses XYZ data to compute the opponent signal, while CIECAM02 is based on a Hunt-Pointer-Estevez space. Finally CIELAB can be used to compute lightness, chroma, and hue correlates, while CIECAM02 can be used to compute these values as well as brightness, colorfulness, saturation, and hue quadrature values. However, CIELAB has benefited from the additional research in advanced color difference equations and as a result has advanced color difference metrics such as ΔE_{94} and ΔE_{2000} which CIECAM02 does not have. There are encouraging results [8] though for using CIECAM02-based color difference equations.

Future Directions

CIECAM02 has been a useful and valuable addition to color appearance modeling research. It has provided a single reference point for ongoing research in the area of color appearance modeling. However, a number of researchers have pointed to specific aspects of the complexity that are problematic in some cases. For example, for the darkest colors, it may not be possible to invert the calculations for highly saturated inputs. These values may be outside the spectral locus, but for color management applications that use a fixed intermediate grid of coordinates, this is a

shortcoming. Therefore, it seems likely that future work will continue in the area of color appearance modeling, with a future focus on robustness [8] and perhaps simplicity. In spite of these limitations, there is already work integrating CIECAM02 with color management systems, such as the International Color Consortium (ICC) [10, 11]. There has also been work [12] to consider how the model could be further extended to encompass a wider range of viewing conditions, such as mesopic illumination levels. Finally, there is also research [13] in the area of how the model could be used with complex stimuli to create an image appearance model.

Cross-References

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

References

1. CIE Publication 159: A Colour Appearance Model for Colour Management Systems: CIECAM02. (2004)
2. Fairchild, M.D.: A revision of CIECAM97s for practical applications. *Color. Res. App.* **26**, 418–427 (2001)
3. Hunt, R.W.G., Li, C.J., Juan, L.Y., Luo, M.R.: Further improvements to CIECAM97s. *Color. Res. App.* **27**, 164–170 (2002)
4. Fairchild, M.D.: *Color Appearance Models*. Addison Wesley, Reading (1998)
5. Hunt, R.W.G.: Models of color vision. In: *Measuring Color*, pp. 208–247. Fountain Pr. Ltd, Kingston-upon-Thames (2001)
6. Hunt, R.W.G., Li, C.J., Luo, M.R.: Dynamic cone response functions for models of colour appearance. *Color. Res. App.* **28**, 82–88 (2003)
7. CIE Pub. No. 15: *Colorimetry*, Central Bureau of the CIE, Vienna. (2004)
8. Li, C.J., Luo, M.R., Cui, G.: Colour-differences evaluation using colour appearance models. In: *Proceedings of the IS&T/SID Color Imaging Conference*. Scottsdale, Arizona, USA, pp. 127–131. (2003)

9. Li, C.J., Luo, M.R.: Testing the robustness of CIECAM02. *Color. Res. App.* **30**, 99–106 (2005)
10. Kuo, C., Zeise, E., Lai, D.: Robust CIECAM02 implementation and numerical experiment within an international color consortium workflow. *J. Imaging Sci. Technol.* **52**, 020603–020606 (2008)
11. Tastl, I., Bhachech, M., Moroney, N., Holm, J.: ICC color management and CIECAM02. In: *Proceedings of the IS&T/SID Color Imaging Conference*. Scottsdale, Arizona, USA, pp. 217–223. (2005)
12. Kwak, Y., MacDonald, L.W., Luo, M.R.: Mesopic colour appearance. *Proc. SPIE* **5007**, 161–169 (2003)
13. Tulet, O., Larabi, O., Fernandez-Maloigne, M.C.: Image rendering based on a spatial extension of the CIECAM02. In: *IEEE Workshop Application of Computer Vision*. Anchorage, Alaska, USA (2008)

CIECAM02 (Standards: CIE)

► [Gamut Volume](#)

CIEDE2000, History, Use, and Performance

Ming Ronnier Luo
 State Key Laboratory of Modern Optical
 Instrumentation, Zhejiang University, Hangzhou,
 China
 School of Design, University of Leeds,
 Leeds, UK
 Graduate Institute of Colour and Illumination,
 National Taiwan University of Science and
 Technology, Taipei, Taiwan, Republic of China

Synonyms

[CIE 2000 color-difference equation](#); ΔE_{00}

Definition

CIEDE2000 [1, 2] is a color-difference formula recommended by the CIE in year 2001. It has also recently been published as an ISO and CIE Joint Standard [3]. It is a modification of CIELAB [4] and gives an overall best performance in predicting experimental datasets. The typical applications are pass/fail decision, color constancy, metamerism, and color rendering.

Overview

Over the years, color scientists and engineers have been striving to achieve a single number pass/fail color-difference equation, i.e., to apply a single pass/fail color difference to all color regions for industrial quality control. In practice, product batches should be visually acceptable against a standard, when a color difference is less than a predetermined color-difference unit (called color tolerance). Reversely, it will be rejected to be returned for re-shading.

In 1976, CIELAB uniform color space was recommended by the CIE. The decision was made based on limited experimental data. It was realized a shortage of reliable experimental data having similar color-difference magnitude as those used in industry (typically with CIELAB color difference (ΔE^*_{ab}) ≤ 5). Hence, many datasets were produced, in which four datasets were considered to be most comprehensive and robust and were used to derive the new formulae. These datasets are Luo and Rigg [5], RIT-DuPont [6], Witt [7], and Leeds [8]. They have 2776, 156, 418, and 307 pairs of samples and average color differences of 3.0, 1.0, 1.9, and 1.6 ΔE^*_{ab} , respectively. All datasets were based on glossy paint samples except that of Luo and Rigg data, which include many subsets based on different materials, and finally all subsets were combined according to the experimental results based on textile samples [5].

Using these data, a series of equations were developed by modifying CIELAB. They all have a generic form as given in Eq. 1:

$$\Delta E = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H S_H}\right)^2} + R_T \left(\frac{\Delta C_{ab}^*}{k_C S_C} \frac{\Delta H_{ab}^*}{k_H S_H}\right) \quad (1)$$

where

$$\begin{aligned} \Delta L^* &= L_{ab,2}^* - L_{ab,1}^* \\ \Delta C_{ab}^* &= C_{ab,2}^* - C_{ab,1}^* \\ \Delta H_{ab}^* &= 2\sqrt{C_{ab,2}^* C_{ab,1}^*} \sin\left(\frac{\Delta h_{ab}}{2}\right) \\ \text{where } \Delta h_{ab} &= h_{ab,2} - h_{ab,1} \end{aligned}$$

and subscripts 1 and 2 are the two samples in a pair. The R_T is an interactive term between ΔC_{ab}^* and ΔH_{ab}^* . The S_L , S_C , and S_H are weighting factors for the correlates of L^* , C_{ab}^* , and h_{ab} and are dependent on the positions of the samples in a pair. The k_L , k_C , and k_H are parametric factors to take into account the surface characteristics of the materials in question such as textile, paint, and plastic.

Equation 1 is a form of ellipsoid along the directions of CIELAB lightness, chroma, and hue angle. The ellipsoid can also be rotated in C_{ab}^* and h_{ab} plane. Four equations were developed after CIELAB in 1976. These were Leeds [8], BFD [9], CIE94 [10] and CMC [11]. The CMC was adopted by the ISO for textile applications in 1992 [11]. The CIE94 was recommended by the CIE for field trials in 1994. Both equations have the first three terms of Eq. 1, and BFD and Leeds include all four terms. All formulae greatly outperform CIELAB to fit the experimental datasets. Industry was confused which equation should be used. Hence, CIE Technical Committee (TC) 1–47 *Hue and Lightness-Dependent Correction to Industrial Colour Difference*

Evaluation was formed in 1998. It was hoped that a generalized and reliable formula could be achieved.

The TC members worked closely together and a formula named CIEDE2000 was recommended [1, 2]. The computation procedure of this formula is given in Eq. 2.

CIEDE2000 ($K_L:K_C:K_H$) Color-Difference Formula

The input of the equation is two sets of CIELAB values for the samples of the pair in question. The procedure to calculate CIEDE2000 is given below:

Step 1. Prepare data to calculate a' , C' , and h' .

$$L' = L^*$$

$$a' = (1 + G)a^*$$

$$b' = b^*$$

$$C'_{ab} = \sqrt{a'^2 + b'^2}$$

$$h_{ab} = \tan^{-1}\left(\frac{b'}{a'}\right)$$

where

$$G = 0.5 \left(1 - \sqrt{\frac{\overline{C_{ab}^*}^7}{\overline{C_{ab}^*}^7 + 25^7}} \right)$$

where $\overline{C_{ab}^*}$ is the arithmetic mean of the C_{ab}^* values for a pair of samples.

$$\begin{aligned} \Delta L' &= L'_2 - L'_1 \\ \Delta C'_{ab} &= C'_{ab,2} - C'_{ab,1} \\ \Delta H'_{ab} &= 2 \sqrt{C'_{ab,2} C'_{ab,1}} \sin \left(\frac{\Delta h'_{ab}}{2} \right) \end{aligned}$$

where $\Delta h'_{ab} = h'_{ab,2} - h'_{ab,1}$

Step 2. Calculate $\Delta L'$, $\Delta C'$, and $\Delta H'$.

Step 3. Calculate CIEDE2000 ΔE_{00} .

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L} \right)^2 + \left(\frac{\Delta C'_{ab}}{k_C S_C} \right)^2 + \left(\frac{\Delta H'_{ab}}{k_H S_H} \right)^2 + R_T \left(\frac{\Delta C'_{ab}}{k_C S_C} \right) \left(\frac{\Delta H'_{ab}}{k_H S_H} \right)}$$

where

$$S_L = 1 + \frac{0.015 (\overline{L'} - 50)^2}{\sqrt{20 + (\overline{L'} - 50)^2}}$$

and

$$S_C = 1 + 0.045 \overline{C'_{ab}}$$

and

$$S_H = 1 + 0.015 \overline{C'_{ab}} T$$

where

$$T = 1 - 0.17 \cos (\overline{h'_{ab}} - 30^\circ) + 0.24 \cos (2\overline{h'_{ab}}) + 0.32 \cos (3\overline{h'_{ab}} + 6^\circ) - 0.20 \cos (4\overline{h'_{ab}} - 63^\circ)$$

and

$$R_T = -\sin (2\Delta\theta) R_C$$

where

$$\Delta\theta = 30 \exp \left\{ - \left[\left(\overline{h'_{ab}} - 275^\circ \right) / 25 \right]^2 \right\}$$

$$\text{and } R_C = 2 \sqrt{\frac{\overline{C'_{ab}}^7}{\overline{C'_{ab}}^7 + 25^7}}$$

(2)

Note that $\overline{L'}$, $\overline{C'_{ab}}$, and $\overline{h'_{ab}}$ are the arithmetic means of the L' , C'_{ab} , and h'_{ab} values for a pair of samples. For calculating the $\overline{h'_{ab}}$ value, caution needs to be taken for neutral colors having hue angles in different quadrants, e.g., Sample 1 and Sample 2 with hue angles of 90° and 300° would have a mean value of 195° , which differs from the correct answer, 15° . This can be obtained by checking the absolute difference between two hue angles. If the difference is less than 180° , the arithmetic mean should be used. Otherwise, 360° should be subtracted from the larger angle, followed by calculating of the arithmetic mean. This gives

$300 - 360^\circ = -60^\circ$ for the sample and a mean of $(90 - 60^\circ)/2 = 15^\circ$ in this example.

Three-Term CIEDE2000 Color-Difference Formula

The CIEDE2000 color-difference equation described above includes four terms. In many applications, a three-term equation is required such as for shade sorting and color tolerance specification, indicating direction to a specific difference in recipe prediction. Hence, a three-term CIEDE2000 version was also developed [12] and is Eq. 3:

$$\Delta E_{00} = \left[(\Delta L_{00})^2 + (\Delta C_{00})^2 + (\Delta H_{00})^2 \right]^{1/2} \quad (3)$$

where

$$\Delta L_{00} = \frac{\Delta L'}{k_L S_L}; \Delta C_{00} = \frac{\Delta C''}{S_C''}; \Delta H_{00} = \frac{\Delta H''}{S_H''}$$

and

$$\Delta C'' = \Delta C' \cos(\varphi) + \Delta H' \sin(\varphi)$$

$$\Delta H'' = \Delta H' \cos(\varphi) - \Delta C' \sin(\varphi)$$

$$\text{where } \tan(2\varphi) = R_T \frac{(k_C S_C)(k_H S_H)}{(k_H S_H)^2 - (k_C S_C)^2}$$

where φ is taken to be between -90° and 90° . If $k_H S_H = k_C S_C$, then 2φ is equal to 90° and φ is equal to 45° :

$$S_C'' = (k_C S_C) \sqrt{\frac{2(k_H S_H)}{2(k_H S_H) + R_T(k_C S_C) \tan(\varphi)}}$$

$$S_H'' = (k_H S_H) \sqrt{\frac{2(k_C S_C)}{2(k_C S_C) - R_T(k_H S_H) \tan(\varphi)}}$$

where $\Delta L'$, $\Delta C'$, and $\Delta H'$ and S_L , S_C , S_H , k_L , k_C , and k_H are the same as those in Eq. 2.

The ΔE_{00} value calculated from Eq. 3 is the same as calculated from Eq. 2.

Evaluation of the CIEDE2000

The typical way to evaluate each formula's performance is to apply statistical measures. One

measure, standardized residual sum of squares (STRESS) index [13], has been widely used as given in Eq. 4. It measures the discrepancy between the calculated color difference (ΔE) and visual difference (ΔV) from an experimental dataset:

$$\text{STRESS} = 100 \sqrt{\frac{\sum (\Delta E_i - f \Delta V_i)^2}{f^2 \sum \Delta V_i^2}} \quad (4)$$

$$\text{where } f = \frac{\sum \Delta E_i^2}{\sum \Delta E_i \Delta V_i}$$

STRESS values are ranged between 0 and 100. For a perfect agreement, STRESS value should be zero. It can be considered as a percentage error prediction. Table 1 gives the performance of the four equations tested using the previous mentioned four datasets together with COM dataset which was combined with a weight for each of the four datasets.

It can be seen in Table 1 that CIEDE2000 gave an overall best performance. In addition, CIELAB performed the worst. CIEDE2000 performed significantly better than the other formulae except insignificantly better than CIE94 for the Leeds and RIT-DuPont sets.

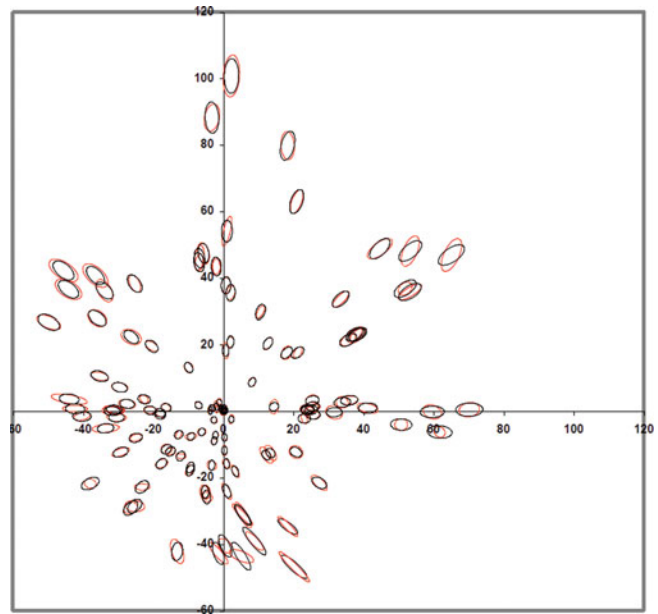
Another widely used method to evaluate color-difference equations is to use color discrimination ellipses. Figure 1 shows that the experimental ellipses (in black color) obtained from the Luo and Rigg dataset plotted against those (in red color) predicted by CIEDE2000. The color discrimination ellipse is an effective way to represent a number of color-difference pairs in a color center. The color differences between points in the ellipse and the color center represent equal visual

CIEDE2000, History, Use, and Performance, Table 1 Color-difference formula performance in STRESS unit (Copyright of the Society of Dyers and Colourists)

	COM	BFD	Leeds	RIT-DuPont	Witt
CIELAB	44	42	40	33	52
CIE94	32	34	21	20	32
CIEDE2000	27	30	19	19	30

CIEDE2000, History, Use, and Performance,

Fig. 1 Experimental color discrimination ellipses plotted in a^*b^* diagram (Copyright of the Society of Dyers and Colourists)



color difference. If CIELAB formula agrees perfectly with the experimental results, all ellipses should be constant radius circles. Hence, the pattern shown in Fig. 1 indicates a poor performance of CIELAB, i.e., very small ellipses close to neutral axis; ellipse sizes increase when chroma increases. Comparing the experimental and CIEDE2000 ellipses, both sets fit well, especially in the blue region. (Note that the experimental ellipses for all other color regions generally point toward the neutral except that the blue ellipses point away from the neutral axis. This effect implies that both CMC and CIE94 formulae do fit well to the experimental results in the region.)

Future Directions

This section showed an outstanding color-difference equation, CIEDE2000, which has been recommended by the CIE. It fits the datasets having magnitude of industrial color differences well. However, it does not have an associated color space. A summary of the future direction on color difference is given below.

- Almost all of the recent efforts have been spent on the modifications of CIELAB. This has resulted in CIEDE2000 including five corrections of CIELAB to fit the available experimental datasets. It is desirable to derive a formula based upon a new perceptually uniform color space from a particular color vision theory such as CIECAM02 [14].
- All color-difference formulae can only be used in a set of reference viewing conditions defined by the CIE [10]. It will be valuable to derive a parametric color-difference formula capable of taking into account different viewing parameters such as illuminant, illuminance level, size of samples, size of color difference, separation, and background. Again, CIECAM02 model and its extension CAM02-UCS [15] are derived to follow this direction.
- Almost all of the color-difference formulae were developed only to predict the color difference between a pair of large single objects/patches. More and more applications require to predict color differences between a pair of pictorial images. The current formula does not include necessary components to consider spatial variations for evaluating images. Johnson and Fairchild developed a spatial model based on CIEDE2000 [16].

Cross-References

- [CIE94, History, Use, and Performance](#)
- [CIECAM02](#)
- [CIELAB](#)

References

1. Luo, M.R., Cui, G.H., Rigg, B.: The development of the CIE 2000 colour difference formula. *Color. Res. Appl.* **26**, 340–350 (2001)
2. CIE Pub. No. 142.: Improvement to Industrial Colour-Difference Evaluation. Central Bureau of the CIE, Vienna (2001)
3. ISO 11664–6:2008(E)/CIE S 014-6/E.: Joint ISO/CIE Standard: Colorimetry-Part 6: CIEDE2000. (2007)
4. CIE Publ. 15.: Colorimetry. Central Bureau of the CIE, Vienna (2004)
5. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colours. *Color. Res. Appl.* **11**, 25–42 (1986)
6. Berns, R.S., Alman, D.H., Reniff, L., Snyder, G.D., Balonon-Rosen, M.R.: Visual determination of suprathreshold color-difference tolerances using probit analysis. *Color. Res. Appl.* **16**, 297–316 (1991)
7. Witt, K.: Geometric relations between scales of small colour differences. *Color. Res. Appl.* **24**, 78–92 (1999)
8. Kim, H., Nobbs J.H.: New weighting functions for the weighted CIELAB colour difference formula, *Proc. Colour 97 Kyoto*. **1**, 446–449 (1997)
9. Luo, M.R., Rigg, B.: *BFD(l:c)* colour difference formula, part I- development of the formula. *J. Soc. Dye. Colour.* **103**, 86–94 (1987)
10. CIE.: Industrial Colour-Difference Evaluation, CIE Publ.116. Central Bureau of the CIE, Vienna (1995)
11. ISO 105-J03.: Textiles: Test for Colour Fastness. Part 3 Calculation of Colour Differences. ISO, Geneva (2009)
12. Melgosa, M., Huertas, R., Berns, R.S.: Performance of recent advanced color-difference formulas using the standardized residual sum of squares index. *J. Opt. Soc. Am. A* **25**, 1828–1834 (2008)
13. Nobbs, J.H.: A lightness, chroma and hue splitting approach to CIEDE2000 colour differences. *Adv. Colour. Sci. Technol.* **5**, 46–53 (2002)
14. CIE Pub. No. 159.: A Colour Appearance Model for Colour Management Systems: CIECAM02. Centre Bureau of the CIE, Vienna (2004)
15. Luo, M.R., Cui, G., Li, C.: Uniform colour spaces based on CIECAM02 colour appearance model. *Color. Res. Appl.* **31**, 320–330 (2006). 425–435
16. Johnson, G.A., Fairchild, M.D.: A top down description of S-CIELAB and CIEDE2000. *Color. Res. Appl.* **28**, 425–435 (2003)

CIELAB

Ming Ronnier Luo

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

School of Design, University of Leeds, Leeds, UK

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

Synonyms

[CIE 1976 L*a*b*](#); [DE*ab](#); [History, use and performance](#)

Definition

CIELAB is a uniform color space (UCS) recommended by CIE in 1976 [1], and it was later published as a Joint ISO/CIE Standard [2]. A UCS is defined by the CIE International Lighting Vocabulary [3] as a color space in which equal distances are intended to represent threshold or suprathreshold perceived color differences of equal size. It is one of the most widely used color spaces. The typical applications include color specification and color difference evaluation. The former is to describe a color in perceptual correlates such as lightness, chroma, and hue and to plot samples to understand their relationships. The latter is mainly used for color quality control such as setting color tolerance, color constancy, metamerism, and color rendering.

For the definition equations of the components of CIELAB, see section [CIE L*a*b* Formula \(CIELAB\)](#).

Overview

Over the years, color scientists and engineers have been striving to achieve a UCS. To apply UCS, a pair of samples will first be measured by a color

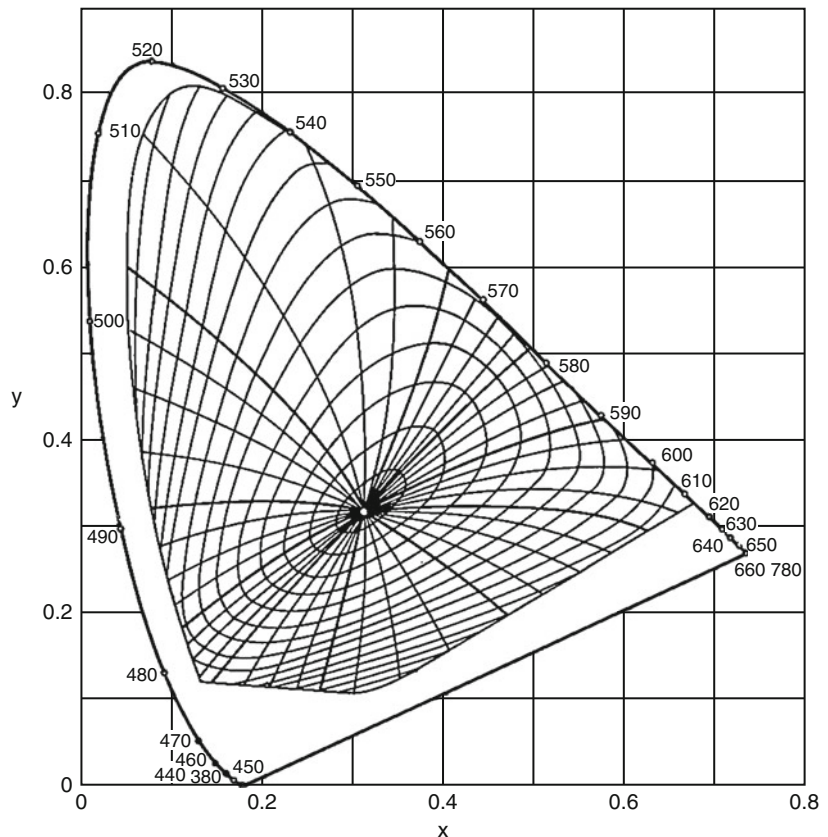
measuring instrument to obtain their CIE tristimulus values (XYZ) which will then be transformed to the perceptual correlates such as CIELAB lightness, chroma, and hue angle. The distance between a pair of colors is calculated and reported as color difference (ΔE). This difference will then be judged against a predetermined color tolerance which could be a particular color region and a product. For a particular product, all pairs should be judged as “acceptable,” when the color difference is less than the color tolerance. Otherwise, it will be rejected. A good color difference formula is also called a “single number pass/fail formula,” i.e., to apply a single color tolerance to all color regions.

Over 20 formulae were derived before the recommendation of CIELAB in 1976 [4]. Some of them were derived to fit the spacing of the Munsell color order system. The concept of the Munsell color order system was invented by A. H. Munsell and was based on steps of equal

visual perception. Any color can be defined as a point in a three-dimensional Munsell color space. Its associated attributes are Munsell hue (H), Munsell chroma (C), and Munsell value (V) which correspond to the perceived hue, saturation, and lightness, respectively. The spacing of the color samples for each attribute was intensively studied by the members in the Colorimetry Committee of the Optical Society of America (OSA), and the CIE tristimulus values of ideally spaced samples were published in 1943 [5].

Figure 1 shows the loci for samples having constant Munsell chroma and curves for samples having constant Munsell hue. Since the Munsell samples are based on equal visual steps, for a perfect agreement between the Munsell data and a color space, all loci should be circles with a constant increment between all neighboring chroma steps. As shown in Fig. 1, this is obviously not the case, i.e., one step of chroma in the blue region is at least five times shorter than one

CIELAB, Fig. 1 Constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in the CIE chromaticity diagram (From Billmeyer and Saltzman [14])



step of chroma in the green region. Additionally, all iso-chroma loci are far from being circles, and no iso-hue contours are straight lines. These indicate that there is a very large difference between the Munsell system and the CIE system represented by x, y chromaticity diagram.

Some earlier Munsell-based formulae are directly calculated using Munsell H, V, and C values with a weighting factor for each component. In 1944, ANLAB was developed by Adams and Nickerson [6] as given in Eq. 1.

$$\Delta E_{\text{ANLAB}} = \sqrt{40 \left\{ (0.23\Delta V_y)^2 + [\Delta(V_x - V_y)]^2 - 0.4[\Delta(V_y - V_z)]^2 \right\}} \quad (1)$$

and

$$I = 1.2219 V_I - 0.23111 V_I^2 + 0.23591 V_I^3 \\ - 0.021009 V_I^4 + 0.00084045 V_I^5$$

where I corresponds to X, Y, or Z tristimulus values.

In Eq. 1, the terms of $(V_x - V_y)$ and $0.4(V_y - V_z)$ correspond to the ANLAB a (redness-greenness) and b (yellowness-blueness) scales, respectively. By adding the third scale $0.23V_y$, ANLAB becomes a three-dimensional UCS. It was recommended by the Colour Measurement Committee (CMC) of the Society of Dyers and Colourists (SDC) and became an ISO standard in 1971 for the application in the textile industry. A series of cube root formulae were also derived to simplify the ANLAB formula which involves a cumbersome fifth-order polynomial function. This resulted in CIELAB color difference formula introduced in 1976 [1]. CIELAB units include L^* , a^* , and b^* ; the asterisk is used to differentiate the CIELAB system from ANLAB.

In 1976, the CIE recommended two uniform color spaces, CIELAB (or CIE $L^*a^*b^*$) and CIELUV (or CIE $L^*u^*v^*$), as it was still not possible to decide which one would correspond better to visual observations.

CIE $L^*a^*b^*$ Formula (CIELAB)

CIELAB equation is given in Eq. 2.

$$L^* = 116f(Y/Y_n) - 16$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)] \quad (2)$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)]$$

where

$$f(I) = I^{1/3}, \text{ for } I > \left(\frac{6}{29}\right)^3$$

Otherwise,

$$F(I) = \frac{841}{108}I + \frac{16}{116}$$

X, Y, Z and X_n, Y_n, Z_n are the tristimulus values of the sample and a specific reference white considered. It is common to use the tristimulus values of a CIE standard illuminant as the X_n, Y_n, Z_n values. Correlates of L^* , a^* , and b^* are lightness, redness-greenness, and yellowness and blueness, respectively.

Correlates of hue and chroma are also defined by converting the rectangular a^*, b^* axes into polar coordinates (see Eq. 3). The lightness (L^*), chroma (C_{ab}^*), and hue (h_{ab}) correlates correspond to perceived color attributes, which are generally much easier to understand for describing colors.

$$h_{ab} = \tan^{-1}(b^*/a^*)$$

$$C_{ab}^* = \sqrt{a^{*2} + b^{*2}} \quad (3)$$

Color difference can be calculated using Eq. 4.

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (4)$$

or

$$\Delta E^*_{ab} = \sqrt{\Delta L^{*2} + \Delta C^{*2}_{ab} + \Delta H^{*2}_{ab}}$$

where

$$\Delta H^*_{ab} = 2 \left(C^*_{ab,1} C^*_{ab,2} \right)^{1/2} \sin \left[\frac{(h_{ab,2} - h_{ab,1})}{2} \right] \quad (5)$$

and subscripts 1 and 2 represent the samples of the pair considered; the ΔL^* , Δa^* , Δb^* , ΔC^*_{ab} , and ΔH^*_{ab} are the difference of L^* , a^* , b^* , C^*_{ab} , and hue in radiant unit (see Eq. 5) between Samples 1 and 2, respectively.

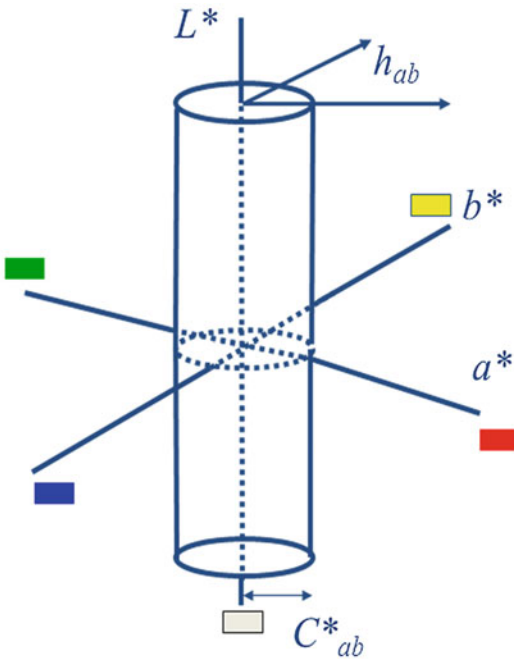
Figure 2 shows the CIELAB color space. It can be seen that the rectangular coordinates consist of L^* , a^* , and b^* . A positive and negative values of a^* represent reddish and greenish colors,

respectively. A positive and negative values of b^* represent yellowish and bluish colors, respectively. For the polar coordinates, hue angle is ranged from 0° to 360° following a rainbow scale from red, yellow, green, blue, and back to red. The 0° , 90° , 180° , and 270° approximate pure red, yellow, green, and blue colors (or unitary hues). Chroma starts from zero origin of neutral axis having chroma of zero and then increases its chromatic content to become more colorful. The colors located in the cylinder of Fig. 4 have a constant chroma, equally perceived chroma content around the hue circle.

Evaluation of CIELAB Color Space

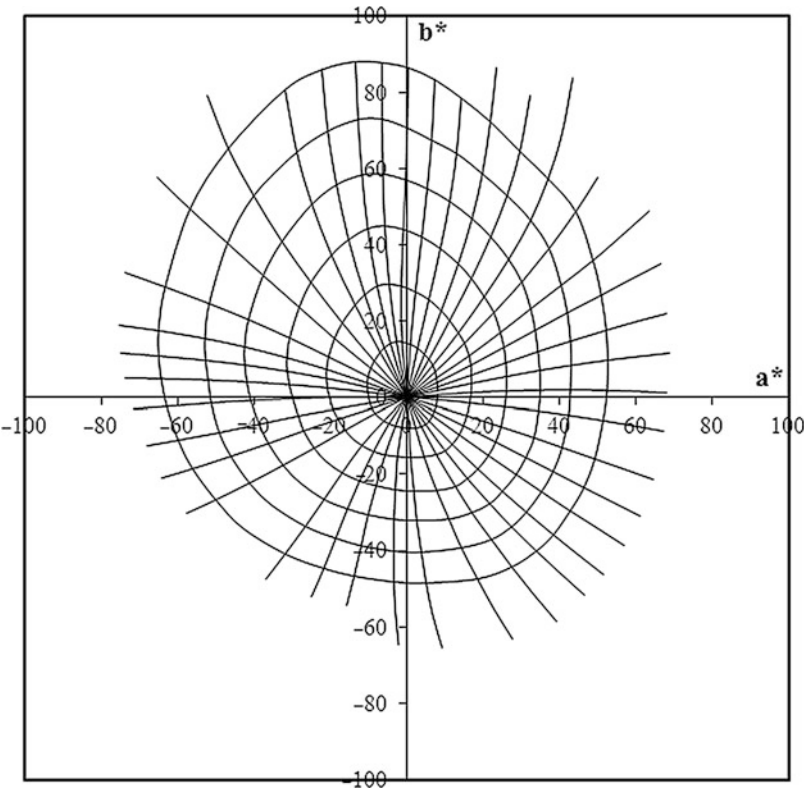
Two sets of data are used here to evaluate the performance of CIELAB space. Figure 3 shows the constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in CIELAB a^*b^* diagram. It can be seen that the pattern is close to the expectation of a good UCS, i.e., the constant chroma loci are close to circle and constant hue radius are close to a straight line (hue constancy). The uniformity of CIELAB is much better than that of CIE chromaticity diagram (see Fig. 1). However, detailed inspection can be found that the same chroma value in yellow and blue regions could still differ by a factor of almost 2. Also, the constant hue line is very much curved in the areas of orange, blue, and green yellow.

After the recommendation of the CIELAB formula, some experimental datasets having $\Delta E^*_{ab} \leq 5$ (representing typical magnitude of industrial color differences) were produced. These sets in general agree with each other. They were later selected to be used to derive different color difference formulae. Figure 4 shows the experimental ellipses obtained from the Luo and Rigg dataset [7]. The dataset includes more color centers and covers a much larger color gamut than the others. A color discrimination ellipse represents the points on the circumference of the ellipse against the center of the ellipse to have the same visual difference. If CIELAB formula agrees perfectly with the experimental results, all ellipses should be constant radius circles. The figure shows the

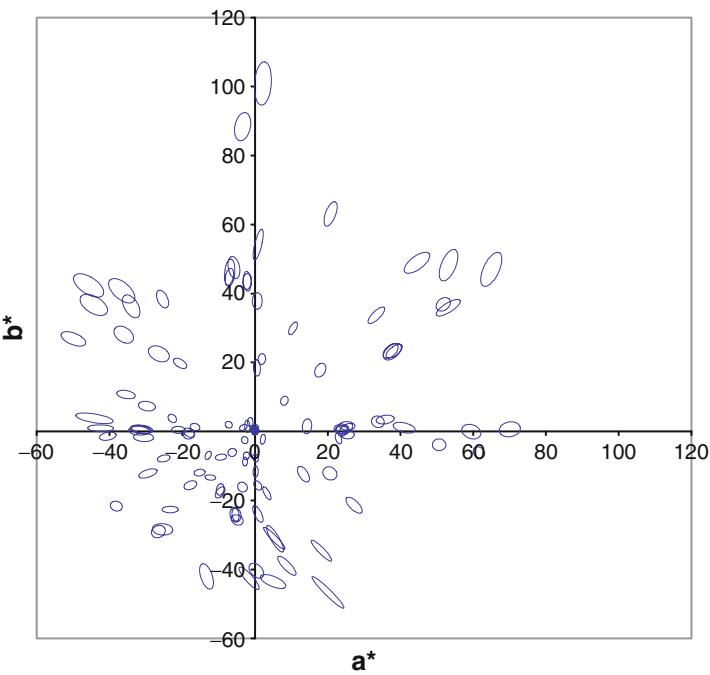


CIELAB, Fig. 2 Illustration of CIELAB color space

CIELAB, Fig. 3 Constant Munsell chroma loci and constant Munsell hue curves at Munsell values of 5 plotted in the CIELAB a^*b^* diagram



CIELAB, Fig. 4 Luo and Rigg experimental color discrimination ellipses plotted in a^*b^* diagram



poor performance of CIELAB. For example, all ellipses close to neutral are much smaller than those in the saturated color regions.

The results given in Figs. 3 and 4 clearly showed the effect of performance that different experimental results could disagree with each other greatly. The discrepancy between the Luo and Rigg and Munsell datasets is mainly due to the color difference magnitude used in the experimental datasets. CIELAB performs not badly for the large color differences with ΔE^*_{ab} about 10 units (see Fig. 3), but predicts very poorly for smaller color differences ($\Delta E^*_{ab} \leq 5$) (see Fig. 4).

Future Directions

Since the recommendation of CIELAB in 1976, many equations were derived by modifying CIELAB such as CMC [8], CIE94 [9], and the more recent CIE recommendation CIEDE2000 [10]. They do fit the datasets ($\Delta E^*_{ab} \leq 5$) well. However, they do not have an associated color space. The future directions on color difference are given below.

- It is desirable to derive a formula based upon a new perceptually uniform color space from a model of color vision theory such as CIECAM02 [11]. A uniform color space is based upon this color appearance model, like CAM02-UCS [12].
- All color difference formulae can only be used in a set of reference viewing conditions defined by the CIE [10]. It will be valuable to derive a parametric color difference formula capable of taking into account different viewing parameters such as illuminant, illuminance level, size of samples, size of color difference, separation, and background. Again, the CIECAM02 model [11] and its extension, CAM02-UCS [12], are equipped with these capabilities.
- Almost all of the color difference formulae were developed only to predict the color difference between a pair of individual patches. More and more applications require evaluating color differences between a pair of pictorial images. Johnson and Fairchild developed a formula for this purpose [13].

Cross-References

- CIE 2000 Color-Difference Equation
- CIE Tristimulus Values
- CIE u' , v' Uniform Chromaticity Scale Diagram and CIELUV Color Space
- CIE94, History, Use, and Performance
- CIECAM02
- CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses)

References

1. CIE Publication No. 015: Colorimetry. Central Bureau of the CIE, Vienna (2004)
2. ISO 11664-4:2008(E)/CIE S 014-4/E: Joint ISO/CIE Standard: Colorimetry-Part4: CIE 1976 $L^*a^*b^*$ Colour Space (2007)
3. CIE: International Lighting vocabulary, Central Bureau of the CIE, Vienna. <http://eilv.cie.co.at/> (2012)
4. Luo, M. R.: The development of colour-difference formulae. Rev. Prog. Color. SDC. 28–39 (2002)
5. Newhall, S.M., Nickerson, D., Judd, D.B.: Final report of the O.S.A. subcommittee on spacing of the Munsell colors. J. Opt. Soc. Am. **33**, 385–418 (1943)
6. Adams, E.Q.: X-Z planes in the 1931 ICI system of colorimetry. J. Opt. Soc. Am. **32**, 168–173 (1942)
7. Luo, M.R., Rigg, B.: Chromaticity-discrimination ellipses for surface colours. Color. Res. Appl. **11**, 25–42 (1986)
8. Clarke, F.J.J., McDonald, R., Rigg, B.: Modification to the JPC79 colour-difference formula. J. Soc. Dye. Colour. **100**, 128–132 (1984). and 281–282
9. CIE Publication No. 116: Industrial colour-difference evaluation. Central Bureau of the CIE, Vienna (1995)
10. Luo, M.R., Cui, G.H., Rigg, B.: The development of the CIE 2000 colour difference formula. Color. Res. Appl. **26**, 340–350 (2001)
11. CIE Publication No. 159: A colour appearance model for colour management systems: CIECAM02. (2004)
12. Luo, M. R., Cui, G., Li, C.: Uniform colour spaces based on CIECAM02 colour appearance model. Color Res. Appl. **31**, 320–330 (2006)
13. Johnson, G.A., Fairchild, M.D.: A top down description of S-CIELAB and CIEDE2000. Color. Res. Appl. **28**, 425–435 (2003)
14. Billmeyer, F.W., Saltzman, M.: Principles of Colour Technology, 2nd edn. Wiley, New York (1981)

CIELAB (Standards: CIE)

- Gamut Volume

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses)

Robert R. Buckley^{1,2} and Edward J. Giorgianni²

¹National Archives of the UAE, Abu Dhabi, UAE

²Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

Synonyms

CIE $L^*a^*b^*$, CIELab, ICC $L^*a^*b^*$, ICCLAB, ITULAB, Lab

Definition

Using the L^* , a^* , and b^* coordinates of the CIELAB uniform color space as the three component values of a 24-bit digital image.

CIELAB

Overview

CIE 1976 (L^* , a^* , b^*) color space (or CIELAB) is widely used for image encoding and processing applications as diverse as color management, gamut mapping, color interchange, and color quality evaluation. What makes CIELAB attractive for these applications is its being an approximately uniform color space. A perceptually uniform scale generally enables the most efficient use of bits and the least visible artifacts when images are digitized for viewing by human observers.

CIELAB Color Space

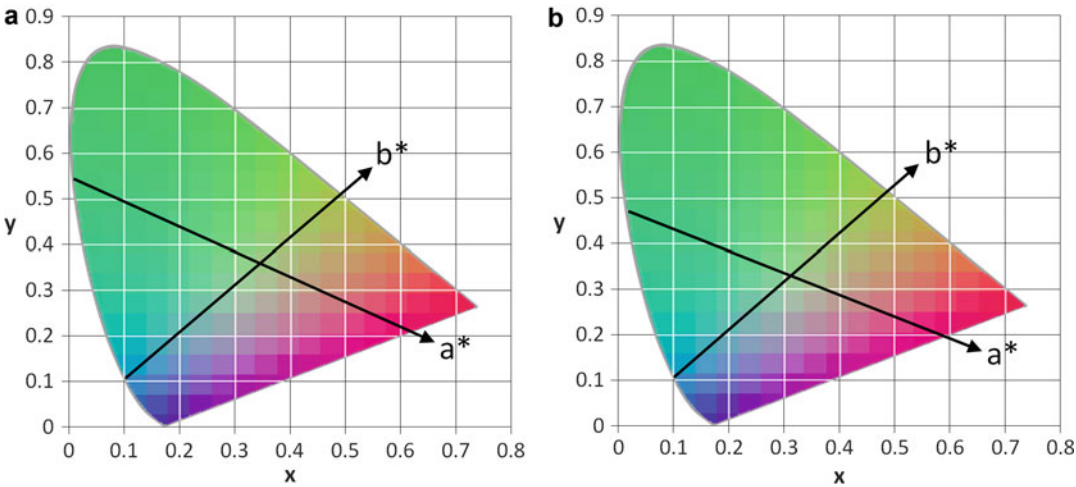
In CIELAB color space a color stimulus is expressed in terms of L^* , a^* , and b^* rectangular coordinates. The entry on CIELAB gives the formulas for calculating L^* , a^* , b^* values from the X , Y , Z tristimulus values of the color stimulus and the X_n , Y_n , Z_n tristimulus values of the reference white object color stimulus. CIELAB color space

can be traced to the Munsell system, which was based on evaluating suprathreshold color differences of isolated reflective patches viewed against a gray background.

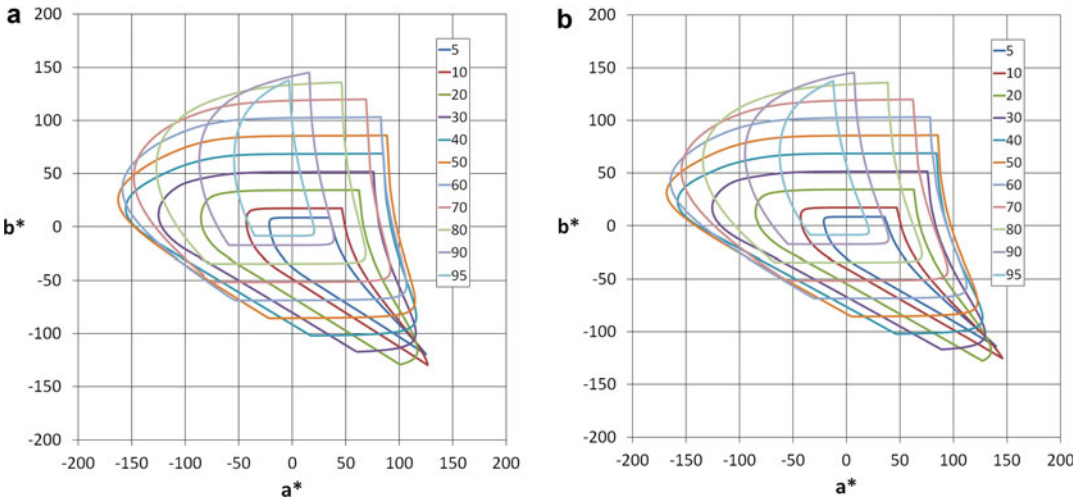
In the CIELAB color space, the L^* value represents the lightness value of the stimulus. A stimulus that reflects no radiation over the visible wavelength band has an L^* value of zero, while the perfectly diffusing reflector under the illuminant with tristimulus values X_n , Y_n , Z_n has an L^* value of 100. Stimuli with tristimulus values that are proportional to those of the reference white color stimulus will have a^* and b^* values of zero. This means that CIELAB is an opponent color space: a gray or achromatic stimulus with the same chromaticities as the normalizing white stimulus will lie on the L^* axis, with black at one end and white at the other. The cube root formula for L^* is a close approximation of the relationship between the Munsell value V and luminance factor Y/Y_n , so L^* is essentially the Munsell value scaled up by a factor of 10. The distance of a stimulus from the L^* axis is given by the chroma value C^*_{ab} .

In an ideal perceptually uniform color space, the Euclidean distance between two color stimuli would correspond to the perceived difference between them. Ideally, then, pairs of color stimuli would look equally different if the three-dimensional distance between them in CIELAB color space, i.e., ΔE^*_{ab} , was the same, independent of the difference being along the L^* , a^* , or b^* dimension or some combination of them. A ΔE^*_{ab} of 1 is generally considered a just perceptible difference, although the threshold depends on how the stimuli are compared and what part of color space they lie in because CIELAB is only approximately uniform. Since 1976, the CIE has published two extensions to the CIE 1976 $L^*a^*b^*$ color-difference formulas to improve perceptual uniformity in the calculation of color differences, i.e., CIE94 and CIEDE2000.

Figure 1 plots the a^* and b^* axes for CIE standard illuminants D_{50} and D_{65} on the x-y chromaticity diagram. In very rough terms, the a^* axis runs from cyan to magenta and the b^* axis, from blue to yellow. The range of a^* and b^* values is determined by the collection of samples at hand or



CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 1 a^* , b^* axes on a CIE x - y chromaticity diagram with (a) CIE standard illuminant D_{50} and (b) CIE standard illuminant D_{65} as normalizing illuminant



CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 2 a^* and b^* values of the MacAdam limits for L^* planes from 5 to 95 for (a) CIE standard illuminant D_{50} and (b) CIE standard illuminant

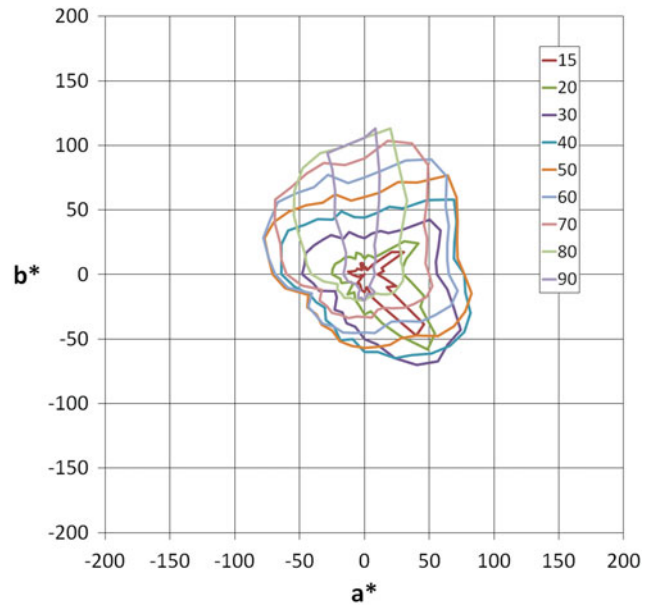
D_{65} (Figure is based on data generously provided by Prof. Francisco Martinez-Verdú of the Colour & Vision Group at the University of Alicante)

that a given device can generate. Figure 2 shows selected L^* planes of the volume generated in CIELAB color space by the MacAdam limits for reflectance materials with respect to CIE standard illuminants D_{50} and D_{65} . For the L^* planes shown in Fig. 2, the a^* values range from -163 to 126.6 for illuminant D_{50} and from -168.2 to 145.6 for illuminant D_{65} ; the b^* values range from -129.6

to 145.1 for illuminant D_{50} and from -127.7 to 145.3 for illuminant D_{65} . These are theoretical limits on the values that can be obtained with reflectance materials.

The ranges obtained with object color stimuli encountered in practice with printed and photographic materials will be much less than the MacAdam limits. For example, Fig. 3 shows the

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 3 a^* and b^* values of the Pointer colors for L^* planes from 15 to 90



Pointer colors measured with illuminant C, which has a correlated color temperature close to that of CIE illuminant D_{65} . An update to the Pointer gamut with a revised set of stimuli and for CIE illuminant D_{50} is under development as this is being written.

Quantization and Dynamic Range

A desirable property of a color space for image encoding is that it covers as large a gamut as needed with the fewest number of levels or digital values. Too many levels are wasteful; too few and an image will show visible contours or bands due to the lower precision in the quantization. Banding is most noticeable in areas with smooth gradients. If the sampling of values or quantization in color space is not fine enough, small differences in the original color values of adjacent pixels in such areas will become larger and potentially visible color differences after quantization. The usual solution is to ensure the quantization is fine enough, especially in the regions of color space where the human visual system is most sensitive to color differences. The addition of noise also can help. If the quantization is mathematically uniform, then quantizing the color space everywhere the same means that the quantization will be finer than it needs to be in regions of color

space less sensitive to color differences, which is inefficient.

The optimum solution for a uniform color quantization of rendered images is to use a perceptually uniform color space. (*Rendered images* are defined here to mean images intended to be directly viewed by human observers.) This solution would use the fewest levels for a given limit on the maximum color error in the quantization or sampling process. This performance is achieved with a three-dimensional quantization based on a body-centered cubic lattice [1, 2], which would also give the minimum mean squared error if the input color stimuli were uniformly distributed [3]. A one-dimensional quantization based on a simple cubic lattice and that samples each CIELAB coordinate L^* , a^* , and b^* independently would be suboptimal but more straightforward to implement. Another reason for less than optimal performance is that CIELAB is only approximately uniform. The divergence of CIELAB from true perceptual uniformity has led to extensions to improve upon various nonuniformities. Nevertheless, the original CIELAB color space is still widely used for many color imaging applications.

For one-dimensional quantization, 8 bits are sufficient to quantize the range of a^* or b^* values

for reflection images typically encountered in practice. While 7 bits would be enough to represent numerical values from 0 to 100, the use of 8 bits generally would be more convenient in a digital system. Moreover, there are perceptual grounds for using a minimum of 8 bits to quantize the L^* coordinate in color images. As mentioned earlier, isolated color patches were used in the development of the Munsell system to which CIELAB color space can be traced. When such patches abut along a sharp edge, smaller color differences are detectable. There are historical data [4] and more recent analysis [5] to justify scaling the L^* value up to 8 bits so that the just noticeable differences along the L^* , a^* , and b^* axes across a sharp edge are more nearly equal. Also, the greater weighting given to L^* differences in color-difference equations in some applications recognizes a generally greater sensitivity to luminance differences compared to chromatic differences.

While there may be room for discussion and debate about the exact value of the scaling factor to use for the lightness coordinate, from the beginning standard engineering practice has been to use 8 bits when encoding L^* in color images [2, 6]. Some CIELAB applications, especially those in which CIELAB-encoded images may later be subjected to significant amounts of additional image processing, may instead specify 12 or 16 bits per coordinate.

CIELAB image data is generated by converting linear tristimulus or image values. Since CIELAB is a nonlinear transform of (linear) XYZ tristimulus values, 12 bits of tristimulus values are needed to address all 256 levels in an 8-bit CIELAB system in order to minimize quantization errors and banding. Nonlinear 8-bit-per-coordinate RGB encodings such as sRGB can be converted to CIELAB using a combination of matrices and one-dimensional lookup tables as long as sufficiently greater bit precision is used in the intermediate conversion stages. Three-dimensional lookup tables are often used for such conversions.

Because CIELAB was developed to be a mathematical representation of the Munsell system – an approximately uniform space based

on reflective color patches – it follows that CIELAB should be well suited for encoding images on reflection media having luminance dynamic ranges and chromaticity boundaries similar to those of the Munsell system color set. In practice, this often is the case. Color images on most reflection media can be encoded satisfactorily in terms of 24-bit (8 bits per channel) CIELAB values if those values are used judiciously. In particular, the luminance dynamic range of the image medium generally should not exceed a visual density range of about 2.50 (a luminance ratio of just over 300:1). The exact luminance dynamic range that can be encoded in terms of 8-bit CIELAB L^* values will depend on a number of factors, most notably the amount of noise in the images. The greater the noise, the less visible quantization artifacts such as contouring will be. In many imaging systems, the noise inherent in the original images and the additional noise contributed by scanners, display devices, and media are sufficient to mask quantization artifacts. However, in other cases, such as virtually noise-free digitally created images, noise in the form of dithering may need to be computationally added in order to avoid unacceptable quantization when 8-bit-per-channel CIELAB encoding is used.

Encoding in terms of 8-bit-per-channel CIELAB values sometimes can be problematic even when applied to images on reflection media or other imaging media or devices having comparable luminance dynamic ranges. In particular, undesirable banding and other artifacts may result if the encoded images are subjected to significant additional image processing. For example, if an image encoded at 8-bit-per-channel CIELAB is appreciably lightened or darkened by subsequent signal processing, quantization effects that originally may have been subthreshold may become visible in image shadow or highlight areas of the processed image. In an extreme case, image processing might be used to reverse the sign of an image, e.g., to make a positive original image into a negative image. In that case, the L^* quantization of the original image would no longer be appropriate for the transformed image, and undesirable quantization effects likely would result.

Another consideration for 8-bit-per-channel CIELAB encoding, as well as for virtually any other 8-bit-per-channel encoding, is its applicability for rendered images having luminance dynamic ranges or chromaticity gamuts significantly greater than those of the Munsell system. For example, imaging devices such as digital projectors and media such as photographic slide films and motion picture films have luminance dynamic ranges considerably greater than those of typical reflection media. When the perceived lightness of such images is first appropriately adjusted for observer general brightness adaptation and then expressed in terms of CIELAB values, the resulting L^* values can range from well below 1.0 to well over 100 [7]. The luminance dynamic ranges of high dynamic range (HDR) digital projectors and other HDR display devices are even greater, and the primaries of such devices can be capable of producing chromaticity color gamuts much larger than that of the Munsell system. In addition, images to be encoded may be colorimetric representations of original scenes rather than of rendered reproductions. Unrendered scene colorimetric values may be in the form of camera RAW values from a digital camera, computationally derived values from scans of color negatives or other photographic media, or values produced by computer-generated imaging (CGI). Such values can greatly exceed those that can be adequately represented by virtually any 24-bit encoding metric, including 8-bit-per-channel color encodings based on CIELAB and other color spaces derived from the Munsell system.

CIELAB Encodings

Different image formats and the applications that use them have made different encoding choices for CIELAB. One of the first standards to define a CIELAB encoding for image and document processing applications was the Xerox Color Encoding Standard [8]. Since then the TIFF, PDF, and JPX formats also have defined CIELAB encodings.

An obvious way to digitally encode L^* , a^* , and b^* values is to convert them to integer values. The TIFF CIELAB encoding [9] does this, using an unsigned integer value for L^* and signed integer

values for a^* and b^* . Other encodings have been designed that use unsigned integers and map the L^* , a^* , and b^* values encountered in practice to match the 8-bit range so as to reduce quantization errors by making best use of the 8 bits available for encoding the values.

When CIELAB is used for image encoding, then the two main choices are:

- What does the L^* range represent?
- What mapping should be used to encode the a^* and b^* values?

The following equations map L^* , a^* , and b^* values to 8-bit encoded L , A , and B values:

$$\begin{aligned} L &= \text{scale}_L \times L^* + \text{offset}_L \\ &= \frac{255}{(L_{\max}^* - L_{\min}^*)} (L^* - L_{\min}^*) \end{aligned}$$

$$\begin{aligned} A &= \text{scale}_A \times a^* + \text{offset}_A \\ &= \frac{255}{(a_{\max}^* - a_{\min}^*)} (a^* - a_{\min}^*) \end{aligned}$$

$$\begin{aligned} B &= \text{scale}_B \times b^* + \text{offset}_B \\ &= \frac{255}{(b_{\max}^* - b_{\min}^*)} (b^* - b_{\min}^*) \end{aligned}$$

The encodings can be described using either scale and offset parameters or the minimum and maximum values that are mapped to the 8-bit values 0 and 255. In some formats, the encoding parameters are defined by the format; in others, the user can explicitly set them with default values usually available.

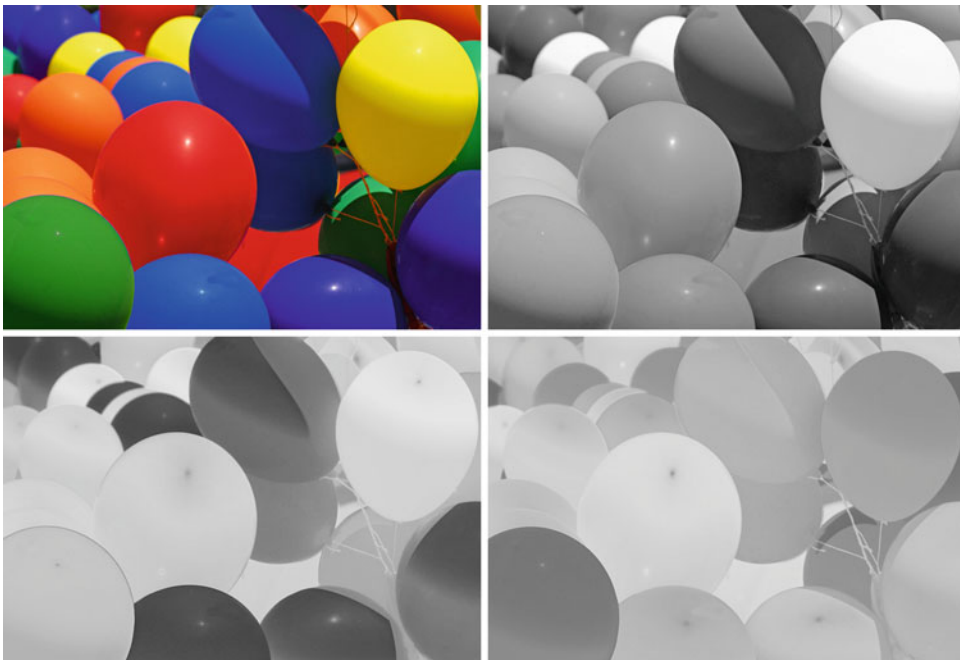
In most formats the L scale factor scale_L is 255/100 and the L offset parameter offset_L is 0. The equivalent minimum and maximum L^* values are 0 and 100. This means that L^* values of 0 and 100 are encoded as 0 and 255. This is the default L^* encoding in TIFF and PDF. In these formats, an L^* value of 100 corresponds to the perfect diffusing reflector under the specified illuminant. In the case of the TIFF CIELAB encoding, the illuminant is not part of the encoding specification. PDF requires that the

diffuse white point be specified. Specifying the illuminant is optional in JPX, which uses CIE illuminant D_{50} as the default. A question sometimes asked is what illuminant should be used with CIELAB. The answer is the reference illuminant used to measure or calibrate the XYZ or RGB values from which the CIELAB values were derived.

The a^* and b^* values typically encountered in practice lie within the 8-bit range from -128 to 127 so they can be encoded using a scale factor of 1 and an offset of 0, which is what the TIFF CIELAB encoding does [9]. The encoded A and B values that result are signed 8-bit integers. The TIFF ICCLAB encoding [10] uses an offset value of 128 instead so that the encoded A and B values are unsigned 8-bit integers, as they are in the 8-bit encoding that the International Color Consortium (ICC) profile format standard uses when CIELAB is the profile connection space [11]. Figure 4 shows the encoded L , A , and B components for the TIFF ICCLAB encoding of a sample image, where a^* and b^* values of 0 are encoded as 128.

Since the ranges of typical a^* and b^* values are usually much less than 255, a scale factor greater than 1 (or equivalently a maximum-minimum difference of less than 255) would expand the range encountered in practice to fill the full range available with 8 bits. The TIFF ITULAB encoding [12] allows an application to specify the minimum and maximum values of a^* and b^* that map to the encoded values 0 and 255. This encoding matches the CIELAB encoding used by color facsimile [13]. PDF [14] provides the same CIELAB encoding for images. JPX [15], the extended version of the JP2 base file format for JPEG 2000, provides an equivalent capability by allowing an application to specify the ranges of a^* and b^* values that map onto the range of 8-bit values and the offsets to go with them.

Besides using the full range available with an 8-bit encoding, scaling the a^* and b^* values also reduces quantization errors, especially when image compression is used. This was important for color facsimile, where reasonable transmission times require the use of moderate to high levels of



CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 4 Clockwise from *upper left*, original color image and its L^* , a^* , and b^* components encoded using the TIFF ICCLAB encoding

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Table 1 Default component ranges for 8-bit CIELAB encodings

Format – encoding	L^*		a^*		b^*		Range settable
	Min	Max	Min	Max	Min	Max	
TIFF – CIELAB	0	100	–128	127	–128	127	No
TIFF – ICCLAB	0	100	–128	127	–128	127	No
TIFF – ITULAB	0	100	–85	85	–75	125	L^*, a^*, b^*
PDF – Lab	0	100	–100	100	–100	100	a^*, b^*
JPX – CIELab	0	100	–85	85	–75	125	L^*, a^*, b^*

JPEG compression, which can lead to visible “blockiness” in the decompressed image. Scaling the a^* and b^* values reduces the effects of the quantization error caused by aggressive JPEG compression. The default scaling defined in the ITU color facsimile standards was based on the range obtained by measuring a wide range of hardcopy materials; the range was –85 to +85 for a^* and –75 to +125 for b^* . The default illuminant for the CIELAB encoding is CIE illuminant D_{50} . The color facsimile standards allow color fax machines to negotiate the use of custom a^* and b^* ranges and other illuminants.

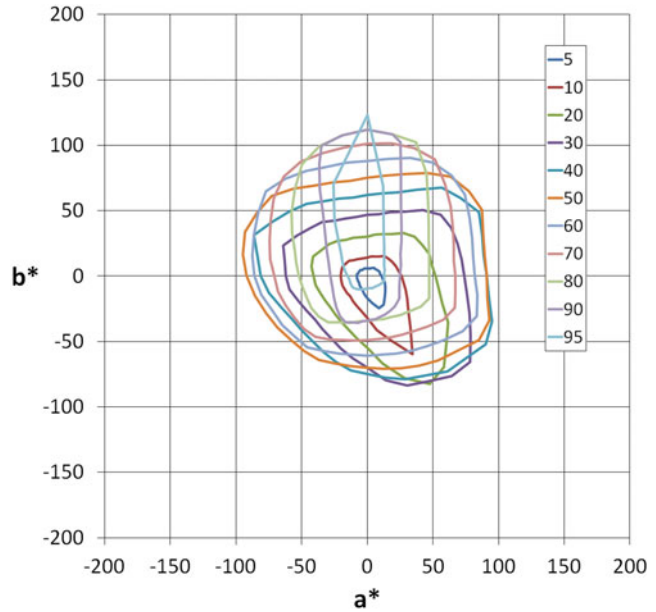
The case for scaling up the L^* values encountered in practice to fill the full 8-bit range is less compelling since the values encountered in practice already cover most of the available range. For example, the minimum and maximum reflectance values for the reference medium defined by the ICC are 0.30911 % and 89 %, corresponding to a visual density range of about 2.46 (a luminance ratio of about 288:1) and minimum and maximum L^* values of 2.79 and 95.58, using the perfectly diffusing reflector as the white reference. While higher dynamic ranges with L^* values greater than 100 originating from scenes and photographic films can be modified to fit within an 8-bit range, the results can be loss of detail due to the tone compression and limited precision provided by 8 bits in those cases.

Table 1 compares 8-bit CIELAB encodings in terms of the defaults for minimum and maximum L^* , a^* , and b^* values and whether the format allows an application to set the ranges for L^* , a^* , or b^* . While the focus here is on 8-bit encodings, these formats also support 16-bit CIELAB encodings.

In the CIELAB encodings described thus far, an L^* value of 100 corresponds to the perfectly diffusing reflector under the specified illuminant. This is not always the case with the CIELAB encoding that the ICC defines for the profile connection space (PCS). The PCS is the color space to or from which input and output device coordinates can be transformed. (Besides CIELAB, the PCS may also use an XYZ encoding.) This makes it possible to connect an arbitrary input device to an arbitrary output device via the PCS in an ICC-based color management system, since a device only needs to know how to transform its coordinates to or from PCS coordinates. Such transforms are defined in device profiles. Transforms can have different intents such as perceptual, where the objective is a pleasing reproduction, or colorimetric, where the objective is a colorimetrically accurate reproduction of the tristimulus values.

Lookup table-based profiles for the *perceptual* rendering intent on input, output, and display devices all use CIELAB as the PCS; both 8- and 16-bit versions are defined. The CIELAB values of the PCS are calculated with XYZ tristimulus values that are normalized with respect to the media white point under CIE illuminant D_{50} rather than the perfectly diffusing reflector. (The PCS CIELAB values for the *colorimetric* intent are calculated with reference to the perfectly diffusing reflector.) In keeping with the orientation to graphic arts and desktop publishing applications, V4 of the ICC specification [11] defined the reference medium for calculating the PCS values of the perceptual rendering as a reflection print on a substrate with a neutral reflectance of 89 % and where the darkest

CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses), Fig. 5 a^* and b^* values of the ISO 12640-3 reference color gamut for L^* planes from 5 to 95



printable color has a neutral reflectance of 0.30911 %. Therefore, the L^* value of the unprinted substrate is 100 and of the darkest printable color is 3.1373. With the 8-bit CIELAB encoding used for the PCS, these values are scaled and encoded as 255 and 8, respectively. A scale factor of 1 and an offset of 128 are used for the a^* and b^* values, as in the TIFF ICCLAB encoding.

The adoption of a reference medium, and subsequently of a reference medium gamut, was designed to improve interoperability within an ICC-based workflow using the perceptual intent. Because a perceptual transform from source to destination media usually required gamut mapping, the gamut of the reference medium used for the PCS was explicitly defined. Besides specifying the endpoints of the L^* scale of the perceptual reference medium, this meant specifying the a^* and b^* limits as well. For this the ICC used the reference color gamut defined in ISO 12640-3 [16] and shown in Fig. 5. As with the PCS, the L^* , a^* , and b^* values in Fig. 5 were calculated using the reference medium (instead of the perfectly diffusing reflector) with CIE illuminant D_{50} as the reference white object color in the L^* , a^* , and b^* equations. The ISO 12640-3 standard also specifies a set of TIFF (specifically TIFF/IT) images with 16-bit CIELAB values that use the

ICC PCS encoding and are within the limits of the reference gamut in Fig. 5.

Conclusions

Based on this discussion, it can be concluded that 8-bit-per-channel CIELAB is appropriate and well suited for digitally encoding color images when all of the conditions below apply:

- The images to be encoded are rendered reproductions.
- The luminance dynamic range and overall color gamut of the images are consistent with those of the Munsell system.
- Appropriate mathematical transformations, such as scale factors and offsets, are used to maximize the efficiency of the CIELAB encoding.
- Relatively limited subsequent image processing is applied to CIELAB-encoded images.

Cross-References

- [CIE Tristimulus Values](#)
- [CIE94, History, Use, and Performance](#)
- [CIEDE2000, History, Use, and Performance](#)

References

1. Buckley, R.: The quantization of CIE uniform color spaces using cubic lattices. In: *Colour 93; 7th Congress of the AIC*. pp. 246–247. International Colour Association (AIC), Budapest (1993)
2. Buckley, R.: Color Image coding and the geometry of color space. Ph.D. thesis. Massachusetts Institute of Technology, Cambridge (1981)
3. Barnes, E.S., Sloane, N.J.A.: The optimal lattice quantizer in three dimensions. *SIAM J. Algebraic Discrete Methods* **4**, 30–41 (1983)
4. Nickerson, D., Newhall, S.M.: A psychological color solid. *J. Opt. Soc. Am.* **33**, 419–422 (1943)
5. Melgosa, M.: Testing CIELAB-based color-difference formulas. *Color. Res. Appl.* **25**, 49–55 (2000)
6. Schreiber, W.F., Buckley, R.R.: A two-channel picture coding system: II – adaptive companding and color coding. *IEEE Trans. Commun.* **COM-29**, 1849–1858 (1981)
7. Giorgianni, E.J., Madden, T.E.: *Digital Color Management: Encoding Solutions*, 2nd edn, pp. 235–237. Wiley, Chichester (1998)
8. Xerox Systems Institute: *Xerox Color Encoding Standard XN55 288811*, Sunnyvale (1989)
9. Adobe Systems Inc: *TIFF, Revision 6.0*, Mountain View (1992)
10. Adobe Systems Inc: *Adobe Pagemaker® 6.0 TIFF Technical Notes*, Mountain View (1995)
11. International Color Consortium: *Specification ICC.1:2010 (Profile Version 4.3.0.0) Image technology colour management – Architecture profile format and data structure* http://color.org/specification/ICC1v43_2010-12.pdf (2010). Accessed 2 Jan 2013.
12. Internet Engineering Task Force (IETF): *RFC 3949, File Format for Internet Fax* (2005)
13. International Telecommunication Union (ITU): *ITU-T Recommendation T.42 Continuous tone colour representation method for facsimile* (2003)
14. International Organization for Standardization (ISO): *ISO 32000-1:2008, Document management – Portable document format – Part 1: PDF 1.7* (2008)
15. International Organization for Standardization (ISO): *ISO 15444-2: 2004, Information technology – JPEG 2000 image coding system: Extensions* (2004)
16. International Organization for Standardization (ISO): *ISO 12640-3:2007, Graphic technology – Prepress digital data exchange – Part 3: CIELAB standard colour image data (CIELAB/SCID)*

Circadian Rhythms

► [Non-Visual Lighting Effects and Their Impact on Health and Well-Being](#)

Circadian System

► [Non-Visual Lighting Effects and Their Impact on Health and Well-Being](#)

Clapper-Yule Model

Mathieu Hébert

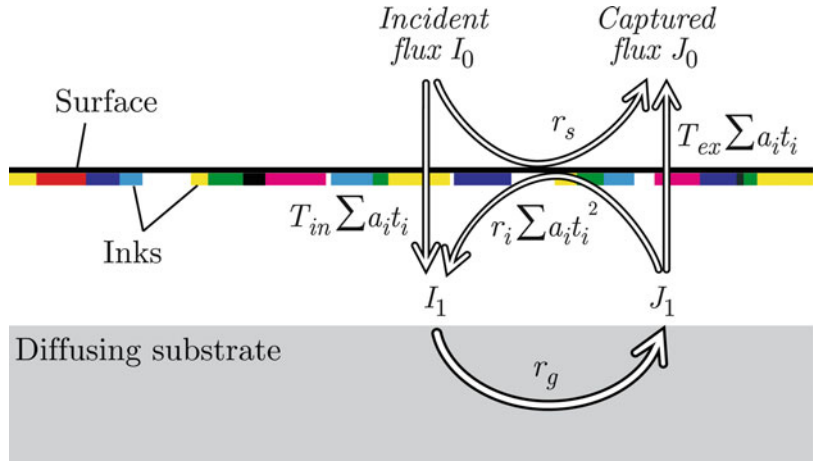
ERIS Group, CNRS, UMR 5516, Laboratoire Hubert Curien, Université de Lyon, Université Jean Monnet de Saint-Etienne, Saint-Etienne, France

Definition

The Clapper-Yule model is a physically based model describing the reflection of spectral light fluxes by a printed surface and enabling the prediction of half-tone prints on diffusing substrates [1]. The model relies on a closed-form equation obtained by describing the multiple transfers of light between the substrate and the print-air interface through the inks. Physical parameters are attached to the inks, the diffusing support, and the surface. The model assumes that the lateral propagation distance of light within the substrate, due to scattering, is much larger than the halftone screen period. Most photons therefore cross different ink dots while traveling in the print. The reflections and transmissions of light at the surface are explicitly taken into account depending on the print's refractive index, as well as the considered illumination and measuring geometries.

The Clapper-Yule Equation

The Clapper-Yule equation derives from the description of multiple reflections of light between the substrate and the print-air interface, represented in Fig. 1. The substrate, strongly diffusing, has a spectral reflectance $r_g(\lambda)$. The print-air interface reflects on its two sides: On the top side, a fraction r_s of the incident light is reflected toward the detector (this fraction may be zero

Clapper-Yule Model,**Fig. 1** Transfers of light flux between the print-air interface and the diffusing substrate

according to the geometry of illumination and detection). A fraction T_{in} enters the print. On the back side, a fraction r_i is reflected and a fraction T_{ex} exits the print toward the detector. Regarding the inks, the model assumes that they are located between the surface and the substrate. If N different inks are used, the halftone is a mosaic of 2^N colors, also called *Neugebauer primaries*, resulting from the partial overlap of the ink dots. In the case of typical cyan, magenta, and yellow inks, there are eight Neugebauer primaries: white (no ink), cyan, magenta, yellow, red (magenta + yellow), green (cyan + yellow), blue (cyan + magenta), and black (cyan + magenta + yellow). Each primary occupies a fractional area a_i on the surface and transmits a fraction $t_i(\lambda)$ of the light. The global attenuation for light crossing the halftone ink layer is therefore:

$$\sum a_i t_i(\lambda).$$

By denoting, $I_0(\lambda)$ and $I_1(\lambda)$ the downward fluxes illuminating the surface, respectively, the substrate, and as $J_0(\lambda)$ and $J_1(\lambda)$ the upward fluxes exiting the surface, respectively, the substrate, the following relations are obtained:

$$\begin{aligned} J_0(\lambda) &= r_s I_0(\lambda) + \left[T_{ex} \sum a_i t_i(\lambda) \right] J_1(\lambda), I_1(\lambda) \\ &= \left[T_{in} \sum a_i t_i(\lambda) \right] I_0(\lambda) \\ &\quad + \left[r_i \sum a_i t_i^2(\lambda) \right] J_1(\lambda), J_1(\lambda) \\ &= r_g(\lambda) I_1(\lambda), \end{aligned}$$

from which is deduced the spectral reflectance $R(\lambda)$ of the halftone print:

$$R(\lambda) = \frac{J_0(\lambda)}{I_0(\lambda)} = r_s + \frac{T_{in} T_{ex} r_g(\lambda) \left[\sum a_i t_i(\lambda) \right]^2}{1 - r_i r_g(\lambda) \sum a_i t_i^2(\lambda)}.$$

This equation is the general Clapper-Yule equation. In their original paper, Clapper and Yule considered halftones of one ink, with surface coverage a and transmittance $t(\lambda)$:

$$R(\lambda) = r_s + \frac{T_{in} T_{ex} r_g(\lambda) [1 - a + at(\lambda)]^2}{1 - r_i r_g(\lambda) [1 - a + at^2(\lambda)]}.$$

If one Neugebauer primary covers the whole surface ("solid primary layer"), the equation becomes:

$$R(\lambda) = r_s + \frac{T_{in} T_{ex} r_g(\lambda) t^2(\lambda)}{1 - r_i r_g(\lambda) t^2(\lambda)},$$

and in the special case where no ink is printed (white colorant), it expresses the reflectance of the printing support:

$$R(\lambda) = r_s + \frac{T_{in} T_{ex} r_g(\lambda)}{1 - r_i r_g(\lambda)}.$$

Fresnel Terms and Measuring Geometries

The terms T_{in} , T_{ex} , r_s , and r_i are derived from the Fresnel formulae for unpolarized light. The

Fresnel reflectivity of an interface between media 1 and 2, with respective indices n_1 and n_2 , is denoted as $R_{12}(\theta)$ when the light comes from medium 1 at the angle θ . The term T_{in} depends on the illumination geometry. It is typically:

$$T_{in} = 1 - R_{12}(\theta)$$

when directional light incomes at angle θ , or

$$T_{in} = 1 - \int_{\theta=0}^{\pi/2} R_{12}(\theta) \sin 2\theta d\theta$$

when the incident light is perfectly diffuse. T_{ex} depends on the measuring geometry. When the print is observed in one direction θ' , it is:

$$T_{ex} = (n_1/n_2)^2 T_{12}(\theta')$$

where the factor comes from the change of solid angle containing the exiting radiance due to the refraction. When an integrating sphere captures all light reflected by the print, it is:

$$T_{ex} = 1 - \int_{\theta=0}^{\pi/2} R_{21}(\theta) \sin 2\theta d\theta.$$

The term r_s corresponds to the surface reflection, i.e., the gloss. It depends on the illumination and observation geometries, by it is typically 0.05 when it is captured, or 0 when it is discarded from measurement.

The reflectance r_i is the fraction of diffuse light emerging from the substrate which is reflected by the surface. It corresponds to the Lambertian reflectance of flat interfaces:

$$r_i = \int_{\theta=0}^{\pi/2} R_{21}(\theta) \sin 2\theta d\theta.$$

For example, for a print made of material of refractive index 1.5 illuminated by directional light at 45° and observed at 0° (so-called $45:0^\circ$ measuring geometry), one has [2]:

$$\begin{aligned} T_{in} &= 1 - R_{12}(45^\circ) = 0.95, \\ T_{ex} &= T_{01}(0^\circ)/n^2 = \frac{1}{n^2} \left[1 - \frac{(1-n)^2}{(1+n)^2} \right] = 0.42, \\ r_s &= 0, \\ r_i &= 0.6. \end{aligned}$$

The Clapper-Yule equation becomes:

$$R(\lambda) = \frac{0.4r_g(\lambda) \left[\sum a_i t_i(\lambda) \right]^2}{1 - 0.6r_g(\lambda) \sum a_i t_i^2(\lambda)}.$$

Calibration of the Model and Prediction

The calibration of the model requires measuring the spectral reflectance of a few halftones, as the one represented in Fig. 2 in the case of CMY halftones.

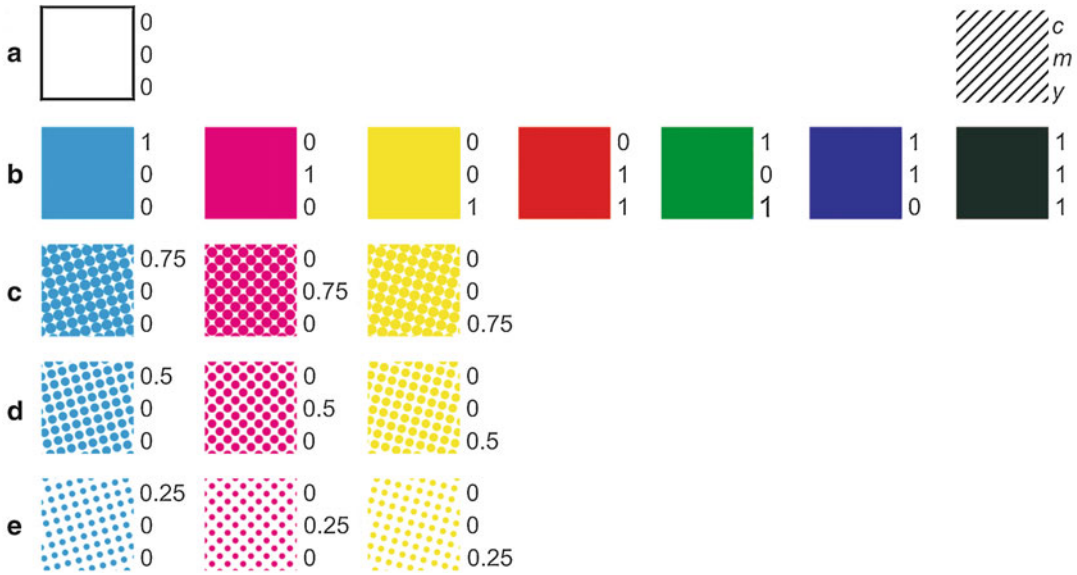
By measuring the spectral reflectance R_w (left (λ) of the unprinted support (patch in row A of Fig. 2), whose expression is given above, one deduces the intrinsic spectral reflectance of the substrate thanks to the following formula:

$$r_g(\lambda) = \frac{R_w(\lambda) - r_s}{T_{in}T_{ex} + r_i(R_w(\lambda) - r_s)}.$$

Then, by measuring the spectral reflectances $R_i(\lambda)$ of the solid primary patches (patches in row B of Fig. 2), whose expression is given above, one deduces the intrinsic spectral reflectance of the substrate thanks to the following formula:

$$t_i(\lambda) = \sqrt{\frac{R_i(\lambda) - r_s}{r_g(\lambda)[T_{in}T_{ex} + r_i(R_i(\lambda) - r_s)]}}.$$

All spectral parameters are now known. The spectral reflectance of a given halftone can be predicted by the Clapper-Yule general equation as soon as the surface coverages of different primaries are known. In classical clustered-dot or error diffusion prints, the primary surface coverages can be deduced from the surface coverages of the inks according to the Demichel equations. In the case of CMY halftones, the Demichel equations relating the surface coverages of the eight



Clapper-Yule Model, Fig. 2 Colors to be printed and measured using a spectrophotometer to calibrate the Clapper-Yule model in the case of CMY halftones

primaries to the surface coverages c , m , and y of the cyan, magenta, and yellow inks are:

$$\begin{aligned}
 a_w &= (1 - c) (1 - m) (1 - y), \\
 a_c &= c(1 - m) (1 - y), \\
 a_m &= (1 - c)m(1 - y), \\
 a_y &= (1 - c) (1 - m)y, \\
 a_{m+y} &= (1 - c)my, \\
 a_{c+y} &= c(1 - m)y, \\
 a_{c+m} &= cm(1 - y), \\
 a_{c+m+y} &= cmy.
 \end{aligned}$$

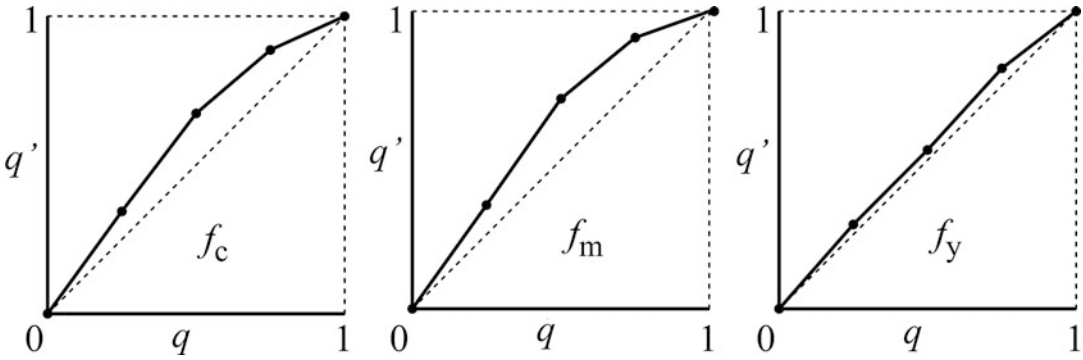
Note that the prediction accuracy of the model sensibly depends on the exactitude of the primary surface coverage values. When the inks spread on the surface, which is almost always the case, the values for c , m , and y are larger than the ones defined in prepress. This phenomenon is generally called *dot gain*. The amount of ink spreading cannot be estimated by advance as it depends on several parameters, such as the chemical and mechanical properties of the inks and of printing support as well as the halftone pattern. They therefore need to be estimated from measurement.

Each ink i is printed alone on paper at the nominal surface coverages $q_i = 0.25, 0.5$, and

0.75 , which correspond to the nine color patches represented in rows C, D, and E of Fig. 2. Their respective spectral reflectances are denoted as $R_{q_i}^{(m)}(\lambda)$. These halftones contain two primaries: the ink which should occupy a fractional area q_i and the paper white which should occupy the fractional area $1 - q_i$. Applying the Clapper-Yule equation with these two primaries and these surface coverages should predict a spectral reflectance $R_{q_i}^{(p)}(\lambda)$ equal to the measured one, but due to the fact that the effective ink surface coverage is different from the nominal one, these two reflectances are not the same. The effective surface coverage q'_i as the q_i value minimizing the deviation between predicted and measured spectra, by quantifying the deviation either by the sum of square differences of the components of the two spectra, i.e.,

$$q'_i = \arg \min_{0 \leq q_i \leq 1} \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \left[R_{q_i}^{(p)}(\lambda) - R_{q_i}^{(m)}(\lambda) \right]^2$$

or by the sum of square difference of the components of their logarithm, i.e.,



Clapper-Yule Model, Fig. 3 Example of ink spreading curves f obtained by linear interpolation of the effective surface coverages q'_{ij} deduced from the measured spectral

reflectances of patches with single-ink halftones (ink i) printed at nominal surface coverages 0.25, 0.5, and 0.75

$$q'_i = \arg \min_{0 \leq q_i \leq 1} \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} \left[\log R_{q_i}^{(p)}(\lambda) - \log R_{q_i}^{(m)}(\lambda) \right]^2$$

or by the corresponding color difference given, e.g., by the CIELAB metric

$$q'_i = \arg \min_{0 \leq q_i \leq 1} \Delta E_{94} \left(R_{q_i}^{(p)}(\lambda), R_{q_i}^{(m)}(\lambda) \right).$$

The first method is the most classical way of determining the effective surface coverage. Taking the log of the spectra as in the second method has the advantage of providing a higher weight to lower reflectance values where the visual system is more sensitive to small spectral differences. Fitting q'_i from the color difference metric sometimes improves the prediction accuracy of the model but complicates the optimization. Even at the optimal surface coverage q'_i , the difference between the two spectra is rarely zero and provides a first indication of the prediction accuracy achievable by the model for the corresponding print setup.

Once the nine effective surface coverages are computed, assuming that the effective surface coverage is 0, respectively 1, when the nominal surface coverage is 0 (no ink), respectively 1 (full coverage), three sets of q'_i values are obtained which, by linear interpolation, yield the continuous ink spreading functions f_i (Fig. 3). As an alternative, one can print halftones at nominal surface coverage 0.5 only and perform parabolic

interpolation [3]. The number of patches needed for establishing the ink spreading curves is thus reduced to three (row C in Fig. 2).

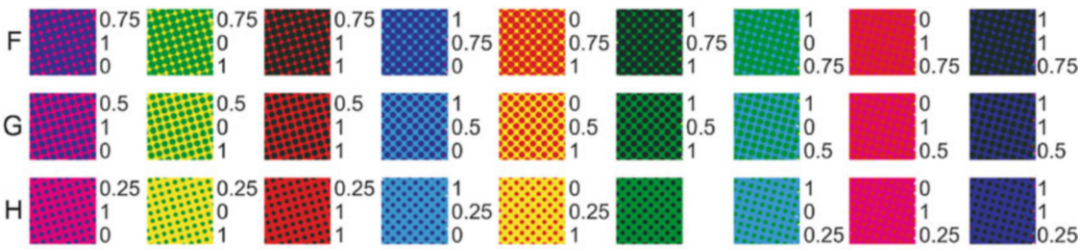
Once the spectral parameters and the ink spreading functions are computed, the model is calibrated. The spectral reflectance of prints can be predicted for any nominal ink surface coverages c , m , and y . The ink spreading functions f_i directly provide the effective surface coverages c' , m' , and y' of the three inks:

$$\begin{aligned} c' &= f_c(c), \\ m' &= f_m(m), \\ y' &= f_y(y). \end{aligned}$$

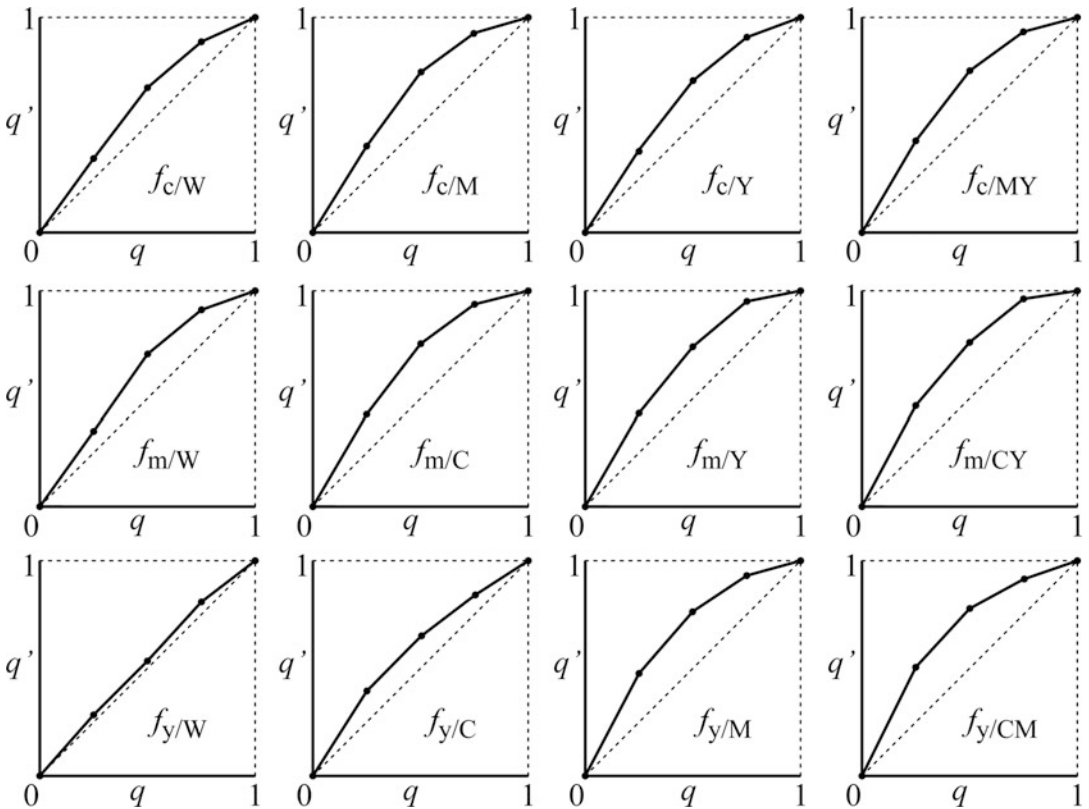
Plugging these effective ink surface coverages into the Demichel equations, one obtains the eight effective primary surface coverages, and the general equation of the model finally predicts of the reflectance spectrum of the considered halftone.

Improved Ink Spreading Assessment Method

Hersch et al. observed that a given ink spreads differently according to whether it is printed alone on paper or superposed with another ink. They proposed an ink spreading assessment method taking into account the superposition conditions of the inks in the halftone [4]. This method increases noticeably the model's prediction accuracy. It relies on the halftones represented in Fig. 2 as well as the ones represented in Fig. 4.



Clapper-Yule Model, Fig. 4 Additional colors to be printed and measured using a spectrophotometer to assess the superposition-dependent ink spreading in the case of CMY halftones



Clapper-Yule Model, Fig. 5 Example of ink spreading curves $f_{i/j}$ for each ink i superposed with solid layers of the other inks in the case of CMY halftones

The nominal ink surface coverages c , m , and y are converted into effective ink surface coverages c' , m' , and y' by accounting for the superposition-dependent ink spreading. In the case of CMY halftones, 12 in. spreading curves are obtained, similar to the ones represented in Fig. 5.

The effective surface coverage of each ink is obtained by a weighted average of the ink spreading curves, the weights being the surface coverages of the respective primaries on which the ink halftone is superposed. For example, the weight of the ink spreading curve f_c (cyan halftone over the white primary) is proportional to the surface of the

underlying white primary, i.e., $(1 - m)(1 - y)$. First, $c' = c$, $m' = m$, and $y' = y$ are taken as initial values on the right side of the following equations:

$$\begin{aligned} c &= (1 - m)(1 - y)f_c(c_0) + m(1 - y)f_{c/m}(c_0) \\ &\quad + (1 - m)yf_{c/y}(c_0) + myf_{c/m+y}(c_0), \\ m &= (1 - c)(1 - y)f_m(m_0) + c(1 - y)f_{c/m}(m_0) \\ &\quad + (1 - c)yf_{m/y}(m_0) + cyf_{m/c+y}(m_0), \\ y &= (1 - c)(1 - m)f_y(y_0) + c(1 - m)f_{y/c}(y_0) \\ &\quad + (1 - c)m f_{y/m}(y_0) + cmf_{y/c+m}(y_0). \end{aligned}$$

The obtained values of c' , m' , and y' are then inserted again into the right side of the equations, which yield new values of c' , m' , y' , and so on, until they stabilize. The final values of c' , m' and y' are plugged into the Demichel equations in order to obtain the effective surface coverages of the eight primaries. The spectral reflectance of the considered halftone is finally provided by the Clapper-Yule equation.

Experimental Testing

In order to assess the prediction accuracy of the model, predicted and measured spectra may be compared on sets of printed colors. As comparison metric, one generally uses the CIELAB ΔE_{94} , obtained by converting the predicted and measured spectra first into CIE-XYZ tristimulus values, calculated with a D65 illuminant and in respect to a 2° standard observer, and then into CIELAB color coordinates using as white reference the spectral reflectance of the unprinted paper illuminated with the D65 illuminant.

Because it assumes that the lateral propagation of light is large compared to the halftone screen period, the Clapper-Yule model is theoretically restricted to halftones with high screen frequency. For example, the model tested on two sets of 729 CMY colors printed with the same offset press on the same paper but at different frequencies, respectively, 76 and 152 lines per inch (lpi), provides better predictions for the highest frequency (average ΔE_{94} of 0.98 unit) than for the

lowest one (average ΔE_{94} of 1.26 unit). Nevertheless, the experience shows that the model may also perform well for middle and low frequencies: For a set of 40 CMY colors printed in ink-jet at 90 lpi on supercalendered paper, the model achieves a fairly good prediction accuracy, denoted by the average ΔE_{94} of 0.47 unit. Note that the average ΔE_{94} is 0.70 unit when the ink superposition conditions are not taken into account in the ink spreading assessment.

Conclusion

Despite the simplicity of its base equation, the Clapper-Yule model is one of the most accurate prediction models for halftone prints. Its main advantage compared to other models such as the Neugebauer model or the Yule-Nielsen-corrected Neugebauer model is the fact that physical parameters are attached to the different elements composing the print (inks, paper, and surface). The Fresnel terms can be adapted to the considered measuring geometry, which is particularly interesting when predictions are made for a geometry different from the one used for calibration. The model also enables controlling ink thickness at printing time by comparing the colorant transmittances in various halftones, whose log is proportional to the ink thickness [5]. Recent improvements and extensions have been proposed which enable predicting both reflectance and transmittance of halftone prints thanks to extended flux transfer model relying on similar physical concepts as the Clapper-Yule model [6, 7].

Cross-References

- [Printing](#)
- [Reflectance Standards](#)

References

1. Clapper, F.R., Yule, J.A.C.: The effect of multiple internal reflections on the densities of halftone prints on paper. *J. Opt. Soc. Am.* **43**, 600–603 (1953)

2. Judd, D.B.: Fresnel reflection of diffusely incident light. *J. Res. Natl. Bur. Stand.* **29**, 329–332 (1942)
3. Rossier, R., Bugnon, T., Hersch, R.D.: Introducing ink spreading within the cellular Yule-Nielsen modified Neugebauer model. In: *IS&T 18th Color Imaging Conference*, San Antonio, TX, USA, pp. 295–300 (2010)
4. Hersch, R.D., Emmel, P., Cr  t  , F., Collaud, F.: Spectral reflection and dot surface prediction models for color halftone prints. *J. Electron. Imaging* **14**, 33001–33012 (2005)
5. Hersch, R.D., Brichon, M., Bugnon, T., Amrhyn, P., Cr  t  , F., Mourad, S., Janser, H., Jiang, Y., Riepenhoff, M.: Deducing ink thickness variations by a spectral prediction model. *Color. Res. Appl.* **34**, 432–442 (2009)
6. H  bert, M., Hersch, R.D.: Reflectance and transmittance model for recto-verso halftone prints: spectral predictions with multi-ink halftones. *J. Opt. Soc. Am. A* **26**, 356–364 (2009)
7. Mazaure, S., H  bert, M., Simonot, L., Fournel, T.: Two-flux transfer matrix model for predicting the reflectance and transmittance of duplex halftone prints. *J. Opt. Soc. Am. A* **31**, 2775–2788 (2014)

Clinical Color Vision Tests

► Color Vision Testing

Coarseness

► Texture Measurement, Modeling, and Computer Graphics

Coatings

► Pigment, Inorganic

Coefficient of Utilization, Lumen Method

Peter Thorns
Strategic Lighting Applications, Thorn Lighting
Ltd, Spennymoor, Durham, UK

Definition

The lumen method is an indoor lighting calculation methodology that allows a quick assessment

of the number of luminaires necessary to achieve a given average illuminance level or alternatively the average illuminance level that will be achieved for a given number of luminaires. It is valid for empty rectangular rooms with simple three-surface diffuse reflectances for ceiling, wall, and floor.

Introduction

When an electric lamp is turned on, it emits light and it is possible to quantify the amount of light by measuring it, the result being given in the units of lumens. In itself this is a useful piece of information, but what would be more useful would be to have a method of converting this into a measure of the amount of light that would be received onto a desk from one or more luminaires or alternatively the number of luminaires necessary to achieve a given quantity of light on the desk. This calculation is known as the lumen method.

Light falling onto a surface is called illuminance and has units of lux (lx) or lumens per square meter (lm/m²). It is not possible to see illuminance as light is actually invisible, which is fortunate, else any view would be degraded by looking through a fog of light. What is actually seen is the effect light has on surfaces, the reflected light, and this is called luminance with units of candelas per square meter (cd/m²). To test this, place a sheet of white paper onto a dark surface (see Fig. 1). At the point where the edge of the paper meets the dark surface, the amount of light falling onto the two materials will be approximately the same. However, they appear completely different as the eye detects the light reflected back towards it, not the light falling onto the surfaces, and the white paper reflects more light than the dark desktop.

However, luminance is a difficult quantity to measure and changes with viewing position. Imagine a day-lit room with resultant shadows and patches of light reflected from polished surfaces. As the observer position moves within the space, the shadows and patches of light change with viewing position. As vision is essentially viewing luminance, this means that the luminance is view



Coefficient of Utilization, Lumen Method,
Fig. 1 White paper on a black desk

dependent. This makes it a difficult design quantity, so for ease, common practice is to design using illuminance which is view independent as the amount of light falling onto a surface does not change with viewing position. (There are a few exceptions, an example being traffic routes, where design may use luminance, in this case the light that is reflected from the road surface).

Calculating the Maximum Achievable Illuminance

To be able to calculate the amount of light falling onto a surface from a given number of luminaires, it is necessary to be able to convert between the measures of lamp lumens (lm) and illuminance (lx). Remember lx is also lm/m^2 , so a given total quantity of lumens can be converted into illuminance by dividing it by the size of the area to be lit in m^2 . But it is important to consider where the area to be lit really is. Ideally luminaires should provide task lighting, that is, light the task being undertaken at any given point to the correct level of illuminance. This means that if there is a desk in a room, it would generally be lit to a level of 500 lx, the recommended level of illuminance for reading and writing tasks [4]. However, the circulation space within the room does not need

this quantity of light, 200 lx being a perfectly adequate illuminance level to safely move around the space. This has two problems:

- Frequently in large spaces, it is not known where tasks such as desks will be positioned and may not be even clearly known what tasks will be performed within the space. Even if the task types and locations are known, most spaces and how they are used change through time. This means that, within reason, lighting has to be flexible enough to preserve the correct lighting conditions for changing requirements.
- Calculating the illuminance level for a particular area within a larger space is generally beyond the ability of a quick and easy calculation method. For lighting the calculation of an average illuminance across a space is relatively easy, while the calculation of the illuminance across a desk located in a room would normally require the use of a computer calculation program.

Therefore, to keep the calculation simple, it will calculate the average illuminance across a horizontal plane, called the task plane. This could either be the floor or a virtual plane at desk height depending upon the expected height of the task. So remembering the units of lm/m^2 , it is necessary to know the total quantity of lumens within the space. So

- For a given number of luminaires, N_{lum}
- Where each luminaire contains a given number of lamps, N_{lamp}
- And each lamp produces a given quantity of lumens, lm_{lamp}

the total amount of lumens available, lm_{total} , is equal to

$$\text{lm}_{\text{total}} = N_{\text{lum}} * N_{\text{lamp}} * \text{lm}_{\text{lamp}} \quad (1)$$

So for a given area of task plane, A_{tp} , and from Eq. 1, the known total amount of lumens available, lm_{total} , the absolute maximum illuminance possible, lx_{max} , will be

$$lx_{\max} = lm_{\text{total}} / A_{\text{tp}} \quad (2)$$

There are two points to consider regarding the value produced by Eq. 2.

- This value assumes *all* the light produced by the lamps will be received onto the task, with no losses. This is generally not possible or even desirable.
- As this is the maximum illuminance theoretically (but not practically) achievable, if the illuminance is too low, at this point more luminaires and, hence more lumens, will be required.

So how is this result converted into a more practical value? The losses in light come from three main effects:

- Losses within the light fitting
- Losses through aging and dirt
- Losses through light not going directly to the task plane but via reflection from room surfaces

Accounting for Losses Within the Light Fitting

A luminaire is a housing containing a light source with associated control gear, optics, wiring, and electrical connections, both to the light source and to the external power supply. This creates a number of problems:

- The luminaire has its own microclimate, the temperature inside the fitting generally being different to that of the air outside. Designing a fitting so that the internal temperature is the optimum for the light source it contains requires skilled design, and the larger the difference between the optimum light source conditions and the actual luminaire conditions results in an increasing reduction in lumen output from the light source and/or reduced operating life expectancy for the source.
- An optic, no matter how well designed, has an element of inefficiency. No surface will reflect

100 % of any light incident upon it and losses are cumulative. So, for example, if a metal reflector has a reflectance of 92 %, then light that bounces once off the reflector before exiting the luminaire loses 8 % of its initial lumens. If it requires two bounces to exit the fitting, only $0.92 * 0.92 = 0.85$ or 85 % of the initial lumens exit the fitting, 15 % being lost. Similarly optics that rely on transmission of light, such as diffusers or prismatic controllers, absorb some of the light, no material being 100 % transmissive.

- Components within the luminaire will obstruct light, and this light may become trapped and never exit the fixture. For example, light emitted from the light source upwards into a ceiling recessed luminaire will need to be reflected around the light source, and some light will be lost in this process, trapped behind the source. (This can cause extra problems if the light is absorbed by the source, causing it to heat up further). Many reflectors trap light in their back which may be open and shaped as the inside of a V with low reflectance.

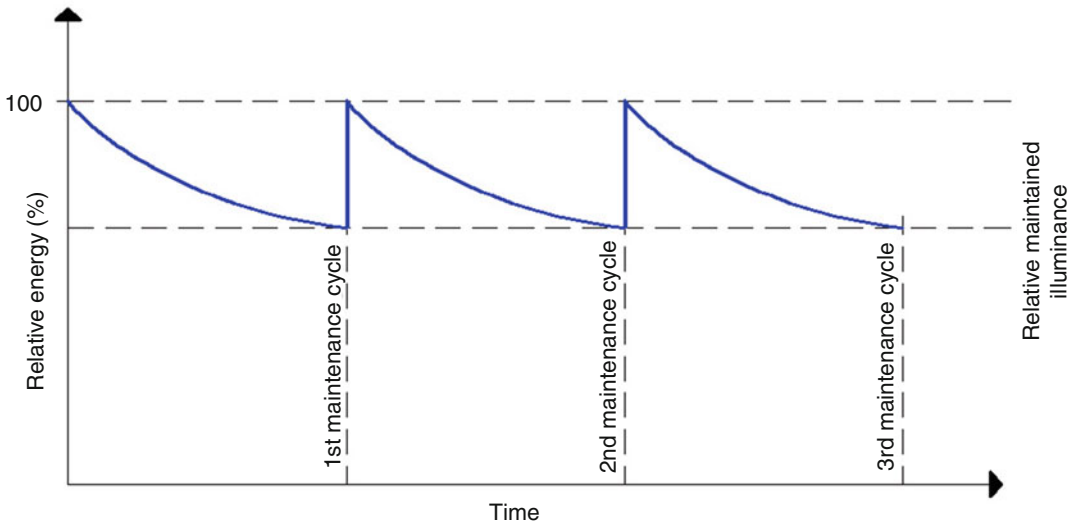
Therefore, a measure is needed to quantify how all of these situations affect the lumen output of the luminaire. This is the light output ratio (LOR). Essentially the LOR is the ratio of light emitted from the luminaire to the light emitted by the light source, and the value of LOR is luminaire specific.

So to account for these effects, Eq. 2 is modified as shown in Eq. 3 below:

$$lx_{\text{luminaires}} = lm_{\text{total}} * \text{LOR} / A_{\text{tp}} \quad (3)$$

Accounting for Losses Through Aging and Dirt

When a lighting system is first installed, the lamps are new and all functioning, the luminaires are clean, and generally the room surfaces (floor, walls, and ceiling) are clean. However, through time the condition of the installation will deteriorate. As light sources age, their lumen output reduces (lumen depreciation) and some lamps will fail completely. Dust and dirt will gather on the reflecting surfaces of the luminaires, reducing



Coefficient of Utilization, Lumen Method, Fig. 2 A typical scheme maintenance cycle

how efficiently light is directed from the light source out of the fitting, and room surfaces will become dirty and marked through everyday wear and tear. All of these will reduce the amount of light reaching the task plane.

A suitable maintenance routine, such as renewing aging lamps and cleaning the luminaires, will help minimize these impacts, but the amount of light received onto the task plane will still vary through life, reducing through time between maintenance cycles, increasing back to close to the original lighting levels immediately after a maintenance cycle (see Fig. 2).

However, a lighting design should be designed to produce a level of maintained illuminance, so that even at the point of maximum reduction in light just before maintenance is performed, the required light level is still achieved. To account for this, a maintenance factor (mf) is used, which is the amount of light lost when the light sources are at their oldest and the luminaires and room surfaces are at their dirtiest. So modifying Eq. 3 produces

$$I_{xmf} = I_{m_{total}} * LOR * mf / A_{tp} \quad (4)$$

Further advice on the determination of maintenance factors is available from the Commission Internationale de L'Eclairage [3].

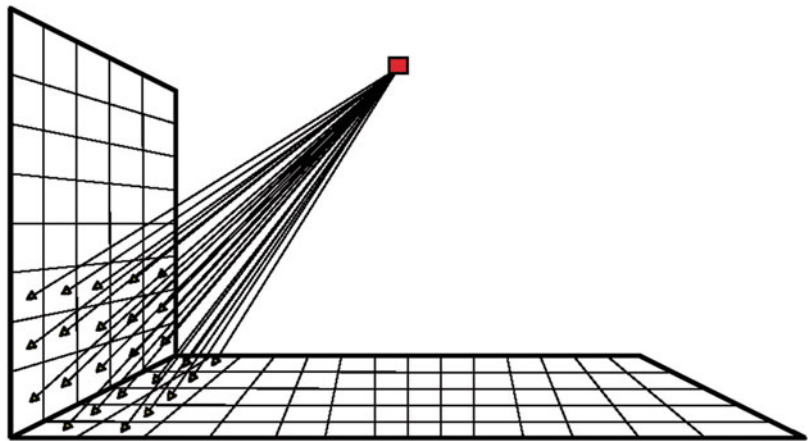
Accounting for Reflection Losses from Room Surfaces

Equation 4 still makes one major assumption that all of the light from the luminaires goes directly onto the working plane.

However, this is rarely the case, light being directed onto the room surfaces (and it is rarely desirable for all light to go directly to the task as this would result in a pool of light within a dark room which would be an uncomfortable work environment). So some of the light hits the room surfaces (see Fig. 3), and some of this light will be reflected back onto the task (see Fig. 4). However, it should be remembered that no surface will have 100 % reflectance. The quantity of light reflected will depend upon the material properties of a surface, a light-colored wall typically having a reflectance in the region of 60 %, so a quantity of light will be lost with each reflection from a surface. To account for this, a measure called the *utilization factor* (UF) is used. This is a measure of the total amount of light from a luminaire that reaches the task, both directly from the luminaire and indirectly through reflection from room surfaces. Manufacturers publish tables of utilization factors, and these vary as values are dependent upon the luminaire distribution, the room reflectance, and also upon the ratio of wall surface area

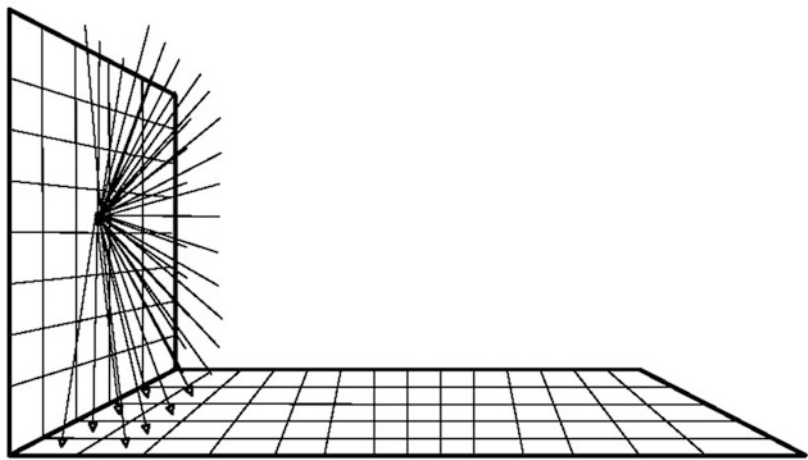
**Coefficient of Utilization,
Lumen Method,**

Fig. 3 Light distribution
onto the wall and floor



**Coefficient of Utilization,
Lumen Method,**

Fig. 4 Light distribution
from the wall to other room
surfaces



to ceiling/floor surface area. (Given a ceiling with a high reflectance within a room that is tall and thin proportionally, there is a large amount of wall surface area to ceiling surface area, so the high reflectance ceiling will have less effect than for a large open plan office with a large ceiling surface area compared with the wall surface area).

To calculate the ratio of wall to ceiling/floor surface area, the room index (K) is used. This is defined as

$$K = (\text{area of ceiling} + \text{area of floor}) / \text{total area of wall} \quad (5)$$

For a room of length L, width W, and height H (where the height is the distance between the

task plane and the luminaire plane, see Fig. 5), Eq. 5 becomes

$$K = ((L * W) + (L * W)) / (2(L * H) + 2(W * H))$$

Therefore,

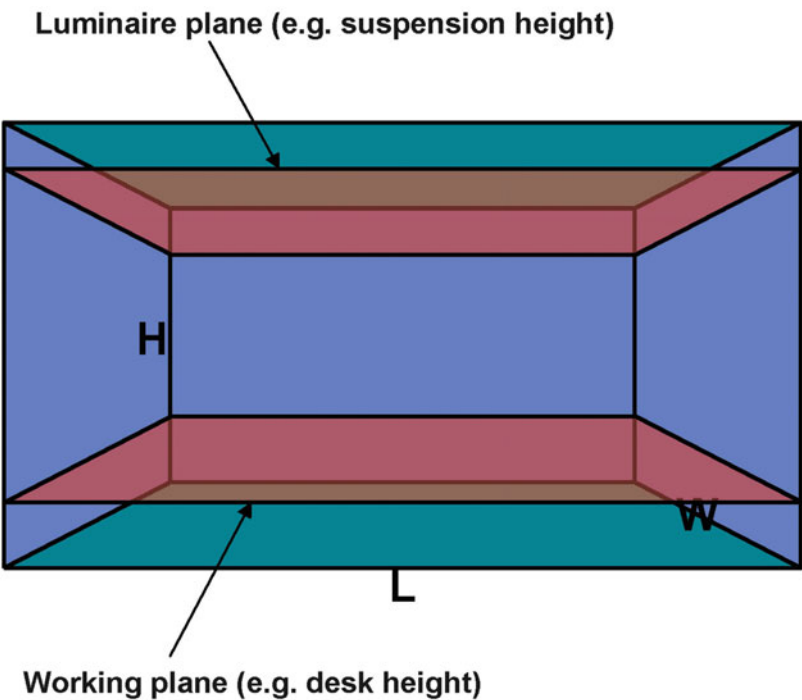
$$K = 2(L * W) / 2H(L + W)$$

Therefore,

$$K = L * W / (L + W)H \quad (6)$$

So given the example table of utilization factors shown in Table 1, it can be seen that values of

Coefficient of Utilization, Lumen Method,
Fig. 5 Room dimensions in terms of L, W, and H



Coefficient of Utilization, Lumen Method, Table 1 An example utilization factor table

Utilization Factors									
Room Reflectance	Room Index								
Ceiling/Walls/Floor	0.75	1.00	1.25	1.50	2.00	2.50	3.00	4.00	5.00
70/50/20	50	56	61	64	69	72	74	77	78
70/30/20	44	51	56	60	65	68	71	74	76
70/10/20	41	47	52	56	62	65	68	72	74
50/50/20	48	54	59	62	66	69	71	74	75
50/30/20	44	50	55	58	63	66	69	72	74
50/10/20	41	47	52	55	61	64	66	70	72
30/50/20	47	53	57	60	64	67	69	71	73
30/30/20	43	49	54	57	62	65	67	69	71
30/10/20	40	46	51	55	59	63	65	68	70
0/0/0	39	44	49	53	57	60	62	65	66

utilization factor are supplied for a variety of room indices and surface reflectance’s. From the table, it can be seen that for a room with reflectance’s of

Length 12 m
Width 12 m
Height 3 m

Ceiling 70 %
Walls 50 %
Floor 20 %

which give a room index of

$$K = (12 * 12)/((12 + 12) * 3) = 2$$

and with dimensions

the utilization factor would be 0.69.

(Values used in the calculation are always percentages, so in this case, 69 % = 0.69. Some tables may already show the values as fractions, in this case 0.69).

Therefore, using Eq. 4 adjusted by the utilization factor gives

$$I_{x_{\text{final}}} = I_{m_{\text{total}}} * LOR * mf * UF / A_{tp} \quad (7)$$

So with knowledge of the type and number of luminaires in a space, it is possible to calculate approximately how much illumination the space will have.

Information on the calculation of utilization factor tables is available from the Commission Internationale de L'Eclairage [1, 2].

Calculating the Number of Luminaires Required in a Room

Equation 7 may also be rearranged to allow the calculation of the required number of luminaires to achieve a given illumination level.

$$I_{m_{\text{total}}} = I_{x_{\text{final}}} * A_{tp} / LOR * mf * UF \quad (8)$$

And using Eq. 1 gives

$$N_{lum} * N_{lamp} * I_{m_{lamp}} = I_{x_{\text{final}}} * A_{tp} / LOR * mf * UF$$

So the number of luminaires required to achieve a given level of illumination is

$$N_{lum} = \frac{I_{x_{\text{final}}} * A_{tp} / LOR * mf * UF * N_{lamp}}{I_{m_{lamp}}} \quad (9)$$

Cross-References

- [Interior Lighting](#)
- [Light Distribution](#)
- [Lumen Depreciation](#)
- [Luminaires](#)

References

1. CIE Publication 40: Calculations for interior lighting – basic method (1978)

2. CIE Publication 52: Calculations for interior lighting – applied method (1982)
3. CIE Publication 97: Guide on the maintenance of indoor electric lighting systems (2005)
4. ISO 8995-1:2002(E)/CIE S 008/E: 2001: lighting of work places part 1: indoor

Coherence

- [Color Harmony](#)

Color Accord

- [Color Harmony](#)

Color Adaptation

- [Environmental Influences on Color Vision](#)

Color Aesthetics

- [Color Preference](#)

Color and Culture

- [Environmental Influences on Color Vision](#)

Color and Light Effects

- [Interactions of Color and Light](#)

Color and Light Interdependence

- [Interactions of Color and Light](#)

Color and Lightness Constancy

► Retinex Theory

Color and Visual Search, Color Singletons

Jasna Martinovic and Amanda Hardman
School of Psychology, University of Aberdeen,
Aberdeen, UK

Synonyms

Color pop-out

Definition

Visual search is a task involving the detection of a unique item within a multi-item display.

Characteristics of the Visual Search Paradigm

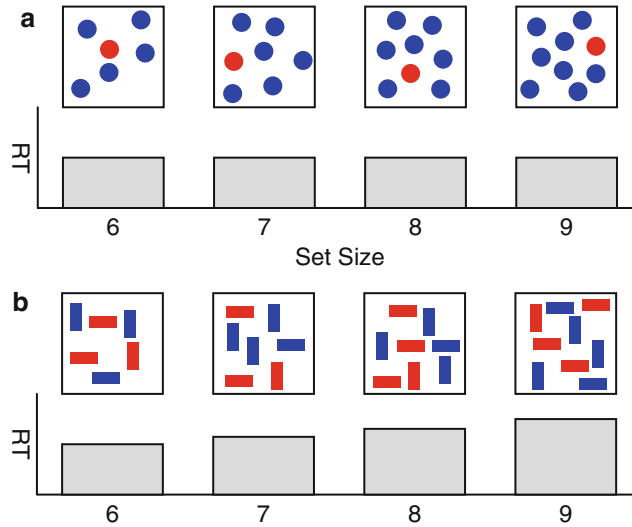
In a visual search task, the item that is being searched for is known as the *target*; other items are known as *distractors*. Figure 1 presents an example of visual search displays, containing a target item and varying numbers of distractor items. The total number of items in the display is known as *set size*. Items in a search display can differ along various *feature dimensions*, for example, color, orientation, or shape. If items in a visual search task vary along a single dimension such as color, the observer may be looking for a target of a specific hue (e.g., red) in a set of different hue distractors. This is known as *feature search*. In the case of feature search, distractors can be heterogeneous – varying in hue – or homogeneous – all of them the same hue (e.g., blue, as in the example in Fig. 1a). If all the distractors share the same color, the uniquely colored target is defined as a *target singleton*. But

irrelevant singletons are sometimes also used in visual search tasks, e.g., a single orange distractor can be present in a display with a red target and a number of blue distractors. Detection of color singletons is typically very efficient, with reaction times for singleton color targets that are independent of set size (see Fig. 1a). Such efficient visual search is also known as a *pop-out effect* or *parallel visual search*. If items vary along multiple dimensions, the participant may be looking for a red bar-oriented upright among a set of distractors that differ in both hue and orientation (see Fig. 1b). This is known as *conjunction search*. Conjunction search is typically inefficient, producing reaction time costs as set size increases, which is a characteristic of *serial visual search*.

Historical Background on Visual Search Tasks in Attention Research

The visual search task is one of the most commonly used paradigms in vision research. There are over 5,000 articles in the Institute for Scientific Information's (ISI) database that refer to visual search in their title. The popularity of the visual search paradigm stems from the fact that it operationalizes a vital task performed by both humans and nonhuman animals. Eckstein's review [1] summarizes many examples of everyday search situations. In natural environments, foraging for food involves searching for edible fruit, whilst in man-made environments, operators monitor complex images in order to detect security-relevant or medically relevant information. In many real world search situations, color is an important determinant of performance due to its ability to make certain features of the scene more or less conspicuous. For example, in order to avoid detection by predators, prey often adopts coloration that acts as camouflage, precluding it to “pop-out” when seen in its natural environment.

Visual search task came to prominence in the 1980s, providing the initial evidence base for Treisman's highly influential Feature Integration Theory (FIT). FIT posited that attentional deployment is guided by multiple, distinct feature maps that are activated in parallel [2]. Visual search



Color and Visual Search, Color Singletons, Fig. 1 Examples of visual search displays. Examples of (a) feature singleton and (b) conjunction search for a range of different set sizes, going from 6 to 9 items. The targets are (a) a red circle and (b) a red vertical bar. The relative

reaction time for each set size is shown underneath each search set. Reaction times for a feature singleton are most often independent of set size, while the reaction times for conjunction search most often increase linearly with the addition of each extra item

provided an excellent paradigm to test this theory, with the potential to reveal the underlying neural representations of feature maps using relatively simple, behavioral methods (for a review, see [3]). The slope of the reaction–time function (milliseconds per item) was considered to be a particularly important variable, providing insight into the amount of time needed for attention to process one item before moving on to the next item. In parallel search, the slope was shown to be constant across different set sizes, which according to FIT was due to the target’s unique representation in a retinotopic feature map, activated in parallel to other such basic level maps. Not all the tenets of FIT held up in the face of stringent experiments that followed, so FIT was supplanted by other models of search. Out of these, the most notable is Wolfe’s Guided Search model which was initially published in [4] and revised in [5].

Initially, visual search studies relied on purely behavioral methods, but they were soon joined with neuroscientific methods, which had the potential to confirm and extend the discoveries made about feature maps underlying attentional

deployment. With its millisecond resolution, EEG was a perfect complement to the traditional reaction time approach of visual search paradigms, allowing a more in-depth look at the timing of processes occurring during search. EEG methods thus extended the scope for testing the diverse competing theories of visual attention such as FIT and its many successors. While EEG was used to establish the timing of various attentional processes, functional magnetic resonance imaging (fMRI) studies were used to determine the extent of the neural networks activated during visual search (for an overview, see [6]).

Color Search and Its Underlying Representations: Cone-Opponent or Hue-Based?

Color was considered one of the basic visual features by FIT due to the fact that it could support parallel visual search. In fact, a long line of studies demonstrated that color was one of the most potent feature dimensions for causing a stimulus to pop-out from its surroundings (for a review, see

[7]). As long as the difference in chromaticity between the target and the distractors is sufficiently large, search for color is efficient [8]. However, in spite of decades of visual search research using color targets and distractors, it is still not fully understood which chromatic representations guide the attentional selection of color. In a seminal early study, D’Zmura [9] showed that search for equally saturated colors is parallel if target and distractor chromaticities can be linearly separated within a hue-based color space. However, while D’Zmura [9] led the way in providing support for selection based on relative distances in a hue-based color space, Lindsey et al.’s [10] more recent findings were strongly in favor of cone-opponent influences on attentional selection. In particular, Lindsey et al.’s study demonstrated that cone-opponent chromatic representations determine the efficiency of attentional selection. These cone-opponent representations originate in the two separate retinogeniculate pathways: the first distinguishes between reddish and greenish hues through opposing the signal from the L and M cones (L-M) and the second distinguishes between bluish and yellowish hues through opposing the S-cone signal with a combination of L and M cone signals (S-(L + M)). Lindsey et al. found that search was particularly ineffective for desaturated S-(L + M) increments (bluish), whilst being particularly effective for pinkish colors that combine an L-M increment with some S-(L + M) information. Recent visual search experiments demonstrate that absolute featural tuning to color gets overruled by relational tuning to color when targets and distractors can be distinguished on the basis of a relative search criterion, e.g., “redder than” or “yellower than”. For example, in a study by Harris, Remington, and Becker [11] observers searched for orange among yellow distractors by selecting items that were “more reddish” when the trials were blocked together, and only tuned into orange as a particular feature when the search displays of orange singletons among red distractors were randomly mixed with search displays of orange singletons among yellow distractors, rendering such relational search templates ineffective.

As visual search is thought to be driven by feature maps situated in the earliest areas of the cortex [12], findings of subcortical representations influencing color search over and above hue-based cortical representations will need to be addressed in future research. One particular problem with the use of visual search to study representations that underlie attention to color is that the visual search paradigm combines bottom-up, salience-driven, and top-down goal-driven influences on attention. The only way to disentangle bottom-up influences from top-down influences in visual search is to use task-irrelevant color singletons (for a review, see [13]). A recent study by Ansorge and Becker [14] used a spatial cueing paradigm instead of classical visual search in order to circumvent the bottom-up/top-down confound inherent in the search task, but the results again failed to support a single representational space being used for color selection. Finally, conflicting experimental findings are likely to be at least in part due to the many methodological differences between studies investigating visual search for color. The studies rely on both different stimulus and task set-ups (search for single or dual targets; differences between stimuli in terms of saturation and lightness) and on different dependent variables that are meant to reflect performance (manual reaction times, reaction time slopes, eye movements, event-related potentials). For example, while the study by Lindsey and colleagues strongly suggests that cone-opponent signals are important in driving attentional effects, the relation of these effects to the level of luminance contrast in the stimulus remains unclear. Li, Sampson, and Vidyasagar [15] demonstrated that while search times for targets defined by L-M contrast are able to benefit from the added luminance signals, this is not the case for targets defined by S-cone contrast. Asymmetrical interactions between luminance and chromatic signals in determining salience would provide a mechanism through which cone-opponent signals are able to influence visual search performance, without discounting any further potential influences from hue-based representations.

Concluding Comments

Visual search experiments led the way in researching the deployment of attention to color. While visual search remains a highly useful paradigm for studying attention to color, it may be advantageous to consider the knowledge on color representations gained from visual search tasks in a more broad context. This is due to its peculiar susceptibility to bottom-up/top-down confounds generated by the search context, e.g., the choice of target/distractor chromoluminance levels.

Cross-References

- [Color Vision, Opponent Theory](#)
- [Effect of Color Terms on Color Perception](#)
- [Magno-, parvo-, koniocellular pathways](#)
- [Unique hues](#)

References

1. Eckstein, M.P.: Visual search: a retrospective. *J. Vis.* **11**(5), 14 (2011)
2. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980)
3. Nakayama, K., Martini, P.: Situating visual search. *Vision Res.* **51**(13), 1526–1537 (2011)
4. Wolfe, J.M., Cave, K.R., Franzel, S.L.: Guided search – an alternative to the feature integration model for visual-search. *J. Exp. Psychol. Hum. Percept. Perform.* **15**(3), 419–433 (1989)
5. Wolfe, J.M.: Guided search 2.0 – a revised model of visual-search. *Psychon. Bull. Rev.* **1**(2), 202–238 (1994)
6. Muller, H.J., Krummenacher, J.: Visual search and selective attention. *Vis. Cogn.* **14**(4–8), 389–410 (2006)
7. Wolfe, J.M.: Visual search. In: Pashler, H. (ed.) *Attention*. University College London Press, London (1998)
8. Nagy, A.L., Sanchez, R.R.: Critical color differences determined with a visual-search task. *J. Opt. Soc. Am. A.* **7**(7), 1209–1217 (1990)
9. D’Zmura, M.: Color in visual search. *Vision Res.* **31**, 951–966 (1991)
10. Lindsey, D.T., et al.: Color channels, not color appearance or color categories, guide visual search for desaturated color targets. *Psychol. Sci.* **21**(9), 1208–1214 (2010)
11. Harris, A.M., Remington, R.W., Becker, S.I.: Feature specificity in attentional capture by size and color. *J. Vis.* **13**(3), 12 (2013)
12. Zhaoping, L.: *Understanding vision: theory, models, and data*. Oxford University Press, Oxford (2014)
13. Theeuwes, J., Olivers, C.N.L., Belopolsky, A.: Stimulus-driven capture and contingent capture. *Wiley Interdiscip. Rev. Cogn. Sci.* **1**(6), 872–881 (2010)
14. Ansorge, U., Becker, S.I.: Contingent capture in cueing: the role of color search templates and cue-target color relations. *Psychol. Res.* **78**(2), 209–221 (2014)
15. Li, J.C., Sampson, G.P., Vidyasagar, T.R.: Interactions between luminance and colour channels in visual search and their relationship to parallel neural channels in vision. *Exp. Brain Res.* **176**(3), 510–518 (2007)

Color Appearance

- [Appearance](#)

Color Appearance Solid

- [Psychological Color Space and Color Terms](#)

Color Atlases

- [Color Order Systems](#)

Color Bleeding

- [Global Illumination](#)

Color Blindness

- [Deuteranopia](#)
- [Protanopia](#)
- [Tritanopia](#)

Color Board

► Palette

Color Categorical Perception

J. Richard Hanley

Department of Psychology, University of Essex,
Colchester, UK

Definition

Categorical perception (CP) occurs when discrimination of items that cross category boundaries is faster or more accurate than discrimination of exemplars from the same category. Categorical perception of color is observed when, for example, a green stimulus and a blue stimulus are more easily distinguished than two stimuli from the same color category (e.g., two different shades of green). Color differences in terms of discriminability can be equated across between-category and within-category comparisons by using the *Commision Internationale de L'Eclairage* (CIE) values. It is therefore important to note that superior cross-category relative to within-category discrimination is observed when the physical distance between cross-category items and the physical distance between within-category items are equivalent.

Categorical Perception Using a Two Alternative Forced-Choice Procedure

Categorical perception of color has been demonstrated many times using a two alternative forced-choice (2-AFC) procedure. In an experiment of this kind, participants view a colored patch for a short duration. Shortly afterward, the original item is displayed next to a distractor, and the participant has to indicate which of the two colored patches was presented a few moments earlier. Discrimination between the target and distractor

is significantly more accurate when they belong to different color categories (e.g., green target and blue distractor) than when they come from within the same color category (e.g., target and distractor are different shades of green).

Categorical perception on the 2-AFC task has provided evidence for the existence of different color categories in different cultures [1]. For example, the Berinmo are a traditional hunter-gatherer culture in the upper Sepik region of Papua New Guinea who have a different set of basic color terms from speakers of English. Berinmo speakers showed significantly better discrimination of 32 cross-category items than 32 within-category items at the boundary between two Berinmo color categories (*nol* and *wor*) that do not exist for English speakers [1]. Conversely, English speakers did not show CP at this boundary. Most important, there was no evidence of CP at the boundary between green and blue for speakers of Berinmo whose language does not make this distinction. Such findings provide evidence against the idea that color categories are universal and exist irrespective of the color vocabulary that speakers have acquired during language development.

Hanley and Roberson [2] argued that CP in the 2-AFC task reflects the role of categorical codes in distinguishing targets from distractors. They proposed that, in principle, participants can perform the 2-AFC task using either categorical or perceptual information about the target item. The categorical code contains less information but may generally be easier to retain than the perceptual code. Targets and distractors can be more accurately distinguished when they cross a category boundary because they differ at both the categorical and the perceptual level. The categorical code cannot be used to distinguish targets and distractors from the same color category, however, so participants must rely entirely on the perceptual code. This account can readily explain the finding that CP for color was abolished when a verbal interference task was interposed between presentation of the target and the test pair [3]. Presumably, verbal interference disrupts participants' ability to generate or retain the category code, and so performance for both within- and cross-

category targets relied on retention of a perceptual representation of the color of the target.

Hanley and Roberson [2] also reanalyzed the results of a series of studies that had demonstrated CP effects for color using the 2-AFC task. This reanalysis revealed that within-category performance was particularly weak when the target was a poor example of a color category and the distractor was a better exemplar. If, however, the target was a good exemplar and the distractor a poor exemplar, performance on within-category trials and cross-category trials was equivalent. Such an outcome is of considerable theoretical importance because it challenges the commonly held view that CP has a perceptual basis (e.g., [4]). According to this account, CP occurs because perceptual systems are fine-tuned by experience such that sensitivity to change is greater around category boundaries than it is to changes that occur within a category. Exactly the same perceptual discriminations are required for within-category trials in which performance is good (central target, peripheral distractor) as for those in which performance is poor (peripheral target, central distractor). Consequently, the advantage for central targets over peripheral targets, which forms the basis for CP in the 2-AFC task, cannot reflect greater perceptual sensitivity for central targets.

Why is performance so poor on the 2-AFC task when targets are peripheral examples of a color category? Hanley and Roberson [2] argued that the poor exemplars in these studies were less likely to be classified as category members when they appeared at test alongside a good exemplar than when they appeared by themselves at presentation. The likely outcome is a mismatch between the way in which a peripheral target is categorized when it is originally presented and the way it is classified at test. A mismatch of this kind will lead to poor performance on the 2-AFC task because the categorical code now provides misleading information about the identity of the target color. This explanation is consistent with the findings that classification of an exemplar at the center of a color category was unaffected by the presence of a peripheral category member [5]. But, crucially, when a poor exemplar appeared alongside a

central exemplar, the poor exemplar was frequently assigned a different category label from the one it was given when it appeared on its own. Similar effects have been observed in other domains and have been termed “category contrast effects” [6].

Consistency of categorization is crucial to the account of CP for color in the 2-AFC task suggested by Hanley and Roberson [2]. For within-category pairs, it might appear that a categorical code could not possibly distinguish target from distractor. However, it follows from the category contrast effect that when the target is a central exemplar and the distractor is a poor exemplar, the distractor is likely to be given a different categorical label from the target. Hence, Hanley and Roberson argued that good performance in the 2-AFC task when the target is a central exemplar occurs because target and distractor can often be accurately distinguished on the basis of the categorical code. It also follows from the category contrast effect that performance will be poor on within-category trials when the target is a peripheral exemplar. For example, a greenish-blue boundary target may be categorized as “blue” when it appears on its own. However, the presence at test of a distractor that is a good example of “blue” means that the peripheral “blue” target will sometimes elicit a different category label (“green”) at test. The outcome will be that the category label given to the target at encoding is elicited at test only by the distractor, which is likely to be selected in preference to the target as a consequence.

Categorical Perception in Visual Search Tasks

Even though test stimuli are presented relatively soon after the target has been removed, performance on the 2-AFC task necessarily depends on a memory code for the target color. Recent studies have also found CP for color in visual processing tasks where the memory component is minimized. For example, Gilbert et al. [7] presented an odd-ball stimulus in the presence of a series of identically colored background items. Even though the

size of the physical difference between the oddball and the background items was held constant, participants were quicker to identify the oddball when the background items belonged to a different color category from the oddball than when oddball and targets were different exemplars of the same color category. Brown et al. [8] have failed to replicate these findings and claim that color discrimination on this task is determined by perceptual rather than categorical differences. Nevertheless, similar results to those reported by Gilbert et al. [7] have been reported in several studies [9–11]. For example, Franklin et al. [10] adapted the task by presenting participants with a colored background and asking them to indicate as quickly as possible which side of fixation a differently colored patch suddenly appeared against this background. Response times were significantly faster when background color and patch came from different categories than when the patch and background were different exemplars of the same category.

In some languages such as Russian, Greek, and Korean, additional color categories exist that are not present in English. For example, *siniy* (dark blue) and *goluboy* (light blue) are distinct basic color terms for speakers of Russian, and *yeondu* (yellow-green) and *chorok* (green) are distinct basic color terms for speakers of Korean. An important question is whether CP can be observed in visual search tasks at boundaries between color categories of this kind. This issue has recently been investigated with speakers of Russian [11] and with speakers of Korean [9]. Russian participants showed CP at the boundary between *siniy* and *goluboy* in the visual search task [11]. Conversely, English speakers, who would call all of these stimuli “blue,” did not show the same cross-category advantage at the *siniy*-*goluboy* boundary. Korean participants showed CP at the boundary between *yeondu* and *chorok*, but no evidence of CP was shown by native English speakers at this boundary [9]. These findings reinforce the claim that color categories are determined by the color vocabulary that speakers acquire during language development.

Two sets of additional findings have provided information about the precise point at which

categorical codes influence color categorization in this experimental paradigm. First, CP was not observed in perceptual tasks when participants carried out a concurrent verbal interference task [7, 11]; under verbal suppression, all equally spaced separations of color were equally easy to discriminate regardless of the presence of a color boundary. Second, Gilbert et al. [7] and Roberson et al. [9] reported that the CP effect was only found for colors that were presented in the right visual field, which is presumed to preferentially access language-processing areas in the left hemisphere. No difference between within-category and between-category pairings of targets and distractors was observed for colors presented in the left visual field, which gains preferential access to the nonverbal right hemisphere. Gilbert et al. [7] also showed that CP was found only when stimuli were presented to the left hemisphere of a patient in whom the corpus callosum (the structure that connects the two hemispheres of the brain) had been surgically severed.

The speed with which categorical information is accessed in these visual processing tasks suggests that there is rapid automatic retrieval of a categorical code when a colored stimulus is presented. Nevertheless, the results from these tasks can be readily explained in terms of Hanley and Roberson’s [3] dual code account of categorical perception. Assume that decisions about whether a target and a background item are the same color are taken on the basis of either a right hemisphere perceptual code or a left-hemisphere categorical code and that when the two codes conflict, accuracy and speed will be reduced. Automatic activation of color category names should therefore slow judgments about whether, for example, two different shades of blue are different. This is because the categorical information that they are the same is in conflict with the perceptual information that they are different. Decisions for items from different categories (e.g., blue and green) will be quicker and more accurate because both the categorical and perceptual codes provide evidence that is consistent. When the left-hemisphere language system is suppressed by verbal interference, or is not accessed because information is presented

directly to the right hemisphere, the categorical code is not generated, and there is never any source of conflict with the perceptual code. Hence, without the language system, there is no advantage for comparisons that fall across category boundaries. Such an account can also explain why an unpredicted improvement in within-category performance has been observed when verbal interference abolished CP for color [7, 11]. Within-category performance cannot be delayed by a mismatch between the categorical and perceptual codes if a categorical code is not generated.

Categorical Perception and Event Related Potentials (ERPs)

Categorical effects for color have also been demonstrated in a paradigm that measures event-related potentials [12]. Participants were asked to look at a display containing a series of colored patches and respond as soon as they saw a cartoon character that appeared infrequently among the patches. The key manipulation was that the patches changed color; sometimes the color category changed, and sometimes the shade changed but the color category remained the same. These color changes took place at a different time from the appearance of the cartoon figures. Color processing was incidental because participants were never asked to pay any attention to color when making their responses to the appearance of the cartoon figures. The latency of the response in the ERP signal to cross- versus within-category color changes was investigated. Crucially, significantly earlier ERP peak latencies were observed when the color change involved a different color category (195 ms) than when the change involved a different exemplar from the same color category (214 ms). Such findings make it clear that information about color categories is available within the brain at a very early stage of visual processing.

Thierry et al. [13] employed a similar task to native Greek speakers. The Greek language makes a categorical distinction between light blue (*ghalazio*) and dark blue (*ble*). Thierry et al. investigated the strength of the ERP

response to unexpected changes from *ble* to *ghalazio* and reported a difference in the amplitude of the ERP. They referred to the response to such changes as a visual mismatch negativity effect (VMMN). The Greek speakers' ERP responses were significantly weaker when there was an unexpected change from light to dark green that did not cross a category boundary for either Greek or English speakers. Conversely, speakers of English did not show a stronger VMMN response to changes from *ble* to *ghalazio* than to changes from light to dark green. Athanasopoulos et al. [14] subsequently found that the signal strength of the VMMN response shown by their Greek participants when *ble* changed to *ghalazio* was negatively correlated with their length of stay in the UK. Once again, these findings emphasize the close link between visual processing of color categories and participants' linguistic experience.

Mo, Xu, Kay, and Tan [15] investigated the strength of the VMMN effect when the color change occurred in only the right (RVF) or in only the left visual field (LVF). They reported significant VMMN effects in both hemifields. Like Thierry et al. [13], Mo et al. found a stronger VMMN when a color change crossed a category boundary (in this case the boundary between blue and green) than when there was a change between two equally distinct exemplars of the same color category (e.g., different shades of green). Crucially, however, this categorical effect was observed only when the color change occurred in RVF; in LVF the strength of the VMMN was equivalent for within- and between-category changes.

In the ERP data, differential responses to cross-versus within-category pairs of stimuli appear between 100 and 300 ms after stimulus presentation. Such rapid processing of categorical information has led to suggestions that CP in perceptual tasks of this kind may reflect activity at a site in visual cortex rather than in language areas [12]. These findings are clearly consistent with the claim that CP effects for color can sometimes be genuinely perceptual and may arise from increased perceptual sensitivity at color category boundaries. It is, however, difficult to reconcile

this claim with the finding that even the early and automatic categorical color effects that are detected by ERPs are lateralized to the left cerebral hemisphere [15]. Furthermore, the published studies that employ these ERP tasks have, to date, only reported cross-category versus within-category contrasts. Exactly the same perceptual discriminations are required regardless of whether the color change on within-category trials is from a good category member to a peripheral category member or vice versa. It remains to be seen whether the ERP signal is equally weak in both these situations relative to the ERP signal generated by a color change that crosses a category boundary. It is therefore unclear whether the asymmetry of within-category performance found consistently to cause CP in 2-AFC tasks [2] generalizes to visual mismatch negativity in ERP signals. If color CP effects in these visual processing tasks are also the result of a category contrast effect, it would be difficult to see how an explanation of the ERP data that is based on perceptual sensitivity could reasonably be maintained.

Cross-References

- [Berlin and Kay Theory](#)
- [Infant Color Categories](#)

References

1. Roberson, D., Davies, I., Davidoff, J.: Color categories are not universal: replications and new evidence from a stone-age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
2. Hanley, J.R., Roberson, D.: Categorical perception effects reflect differences in typicality on within-category trials. *Psychon. Bull. Rev.* **18**, 355–363 (2011)
3. Roberson, D., Davidoff, J.: The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cogn.* **28**, 977–986 (2000)
4. Hamad, S.: Psychophysical and cognitive aspects of categorical perception: a critical overview. In: Hamad, S. (ed.) *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge (1987)
5. Hampton, J.A., Estes, Z., Simmons, C.L.: Comparison and contrast in perceptual categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1459–1476 (2005)
6. Stewart, N., Brown, G.D.A., Chater, N.: Sequence effects in categorization of simple perceptual stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 3–11 (2002)
7. Gilbert, A.L., Regier, T., Kay, P., Ivry, R.B.: Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 489–494 (2006)
8. Brown, A.M., Lindsey, D.T., Guckes, K.M.: Color names, color categories, and color-cued visual search: sometimes color perception is not categorical. *J. Vis.* **11**, 1–21 (2011)
9. Roberson, D., Pak, H., Hanley, J.R.: Categorical perception of color in the left and right hemisphere is verbally mediated: evidence from Korean. *Cognition* **107**, 752–762 (2008)
10. Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., Regier, T.: Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3221–3225 (2008)
11. Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Boroditsky, L.: Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7780–7785 (2007)
12. Fonteneau, E., Davidoff, J.: Neural correlates of color category perception. *Neuroreport* **18**(13), 1323–1327 (2007)
13. Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., Kuipers, J.R.: Unconscious effects of language-specific terminology on preattentive color perception. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4567–4570 (2009)
14. Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.-R., Thierry, G.: Perceptual shift in bilingualism: brain potentials reveal plasticity in pre-attentive color perception. *Cognition* **116**, 437–443 (2010)
15. Mo, L., Xu, G., Kay, P., Tan, L.-H.: Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14026–14030 (2011)

Color Categorization and Naming in Inherited Color Vision Deficiencies

Valerie Bonnardel
Psychology, University of Winchester,
Winchester, Hampshire, UK

Synonyms

[Dichromat color naming](#)

Definition

Color categorization and naming behaviors of human observers that experience forms of color vision deficiency called “dichromacy.” Such deficiencies are sex linked and predominantly affect males and are due to errors in photopigment expression or functioning, or to the failure to inherit the genetic precursors for the expression of a normal set of retinal photopigments.

Color Naming and Categorization in Color Deficients

Human color perception is categorical. Light diffracted through a prism provides a visible spectrum spanning continuously from 400 to 700 nm in the wavelength interval that readily appears to the human eye as a smooth juxtaposition of colored bands of different width. By adding to the visible spectrum a mixture of spectral lights, Isaac Newton described a color circle with its different sectors as *violet*, *indigo*, *blue*, *green*, *yellow*, *orange*, and *red* (Fig. 1) [1]. For the great majority of the Caucasian population (96 %), Newton’s

seven-color categories (although not unanimously agreed upon) are a fair account of human color experience. However, for a substantial percentage of men who are color deficient (8 %), this description is inadequate. Compared to individuals with normal color vision, color deficient have reduced color discrimination abilities and live in a linguistic environment with a larger color vocabulary than the gamut of color percepts they probably experience. Yet, in everyday life, color deficient’s use of color names permits intelligible use of color in conversation, and their deficiency is largely unnoticeable by color normal interlocutors. As such, color deficiencies offer a useful paradigm for the study of the complex relationships between idiosyncratic perceptual experience and its linguistic expression.

Color Vision Deficiencies

Only inherited forms of color vision deficiency associated with changes in the genes on the X-chromosomes will be considered here. Normal color vision is trichromatic and results from light absorption by three different types of photopigments located in the retinal cell receptors called “cones.” Short-wavelength (SW), medium-wavelength (MW), and long-wavelength (LW) cone photopigments absorb light maximally in short (440 nm), medium (540 nm), and long (560 nm) wavelengths. A genetic polymorphism affecting X-linked genes encoding for MW and LW cone photopigments induces small variations in their absorption peak, leading to subtle variations among normal trichromats; larger variations produce deficiencies. In the extreme case, a gene deletion occurs, and as a result, color vision will lose one dimension and will be dichromatic. A more moderate condition arises from gene alterations that generate hybrid photopigments. In this case, color vision is referred as anomalous trichromatic. When LW photopigment is affected, deficiencies are of the protan type, with about 1 % of dichromats (of “protanope” type) and 1 % of anomalous trichromats or “protanomalous.” When the MW photopigment is concerned, deficiencies are of



Color Categorization and Naming in Inherited Color Vision Deficiencies, Fig. 1 Color circle from Newton’s Opticks [1] (Source:<http://posner.library.cmu.edu/Posner/books/>)



Color Categorization and Naming in Inherited Color Vision Deficiencies, Fig. 2 Color circle obtained from sRGB color display (center) and its two simulations as

supposedly seen by a deuteranope (left) and protanope (right). Simulations were obtained with Vischeck (www.vischeck.com)

deutan type and account for frequencies of 1.4 % “deuteranope” and 4.6 % of “deuteranomaly.”

In practice, color deficient people confuse colors that are easily discriminable to normal trichromats. To gain intuition on color deficient’s perceptual experience, two clarifications are useful to bear in mind. The first consists in defining the term “color” which can be polysemous. Color can be defined as corresponding to the appearance of objects and lights resulting from the combination of three perceptual attributes associated to three physical variables, that is, “hue” (associated to dominant or complementary wavelength), saturation (associated to purity), and brightness (associated to luminance). Color deficient people primarily confuse hues and retain the ability to discriminate between different levels of saturation and brightness. Secondly, photoreceptor absorption process is not useful when considering the perceptual implications of their dysfunction. This is better accounted for by considering further stages of color processing which involve synergic and antagonist neural linkages between photoreceptors to form an achromatic (white vs. black) [MW+LW] and two-chromatic channels: “red” vs. “green” [LW-MW] and “yellow” vs. “blue” [(LW+MW) – SW]. Dichromats who are missing either LW or MW photoreceptors are assumed to retain a single color-opponent channel: [MW-SW] or [LW-SW] (protanope or deuteranope, respectively). Anomalous trichromats possess a residual yet functional “red-green” [LW-MW] channel.

Considering the diversity of anomalous trichromat types, illustrations of color deficiencies will be limited to dichromatic vision.

Algorithms based on the reduction assumption permit dichromatic color gamut simulations as seen by individuals with normal color vision [2]. The simulated color circles for deuteranope and protanope are filled with blue, yellow, and shades of those colors and a neutral or achromatic zone (Fig. 2). Although the qualities of another person’s sensations can never be fully known, such simulations provide insights on magnitudes of the difference that might exist between dichromatic and trichromatic color experience.

Despite substantial differences in color experience, the condition of color deficiency had not been reported until the end of the eighteenth century. John Dalton (posthumously diagnosed in 1995 as a deuteranope by means of molecular genetic techniques [3]) provided the most comprehensive description of color deficiency based on his own experience. It is at age of 28, when noticing that a geranium changed its color from sky blue in daylight to yellowish in candlelight, that Dalton suspected his color vision to be different from that of others. From systematic observations of the solar spectrum, Dalton established that if people distinguished six color categories, namely, *red*, *orange*, *yellow*, *green*, *blue*, and *purple* (further divided in *blue* and *indigo* to fit with Newton’s nomenclature), he was able to see only two or at most three distinctions: *yellow* and *blue*

or *yellow*, *blue*, and *purple*. This description is in fair agreement with today simulations presented in Fig. 2.

Dichromats Color Naming and Categorization Abilities

In everyday life, colors are usually object colors, and as young children, dichromats can learn to associate color names with objects of predetermined color such as green grass or red cherries, and it is not surprising that dichromats use “red” and “green” to refer to colors that seem so distinctive to others. However, “red” and “green” are also used accurately most of the time for non-predetermined object colors such as clothes or furnishings, and asking color names to color deficient gives little indication of a color perception deficiency. It is in laboratory situation that one can measure the discrepancy between perception and naming in dichromats. Jameson and Hurvich [4] in an article entitled “Dichromatic color language: “reds” and “greens” don’t look alike but their colors do” [4] reported an experiment using arrangement color tests (i.e., Farnsworth D-15, with 15 colors), where colors printed on small caps should be ordered by perceived similarity. In this type of tests, dichromats make systematic and identifiable color alternations from which color diagnosis is based. For instance, dichromats put cap #2, considered by normal trichromats as “blue-green,” next to cap #13, a “violet-blue” cap. When subsequently, dichromatic observers were asked to name the same stimuli, the authors noted several interesting observations. Firstly, and despite confusing their colors, all dichromats used “red” and “green” words. Secondly, a dichromat could describe a color cap as “red-green.” This designation never occurs in normal trichromat observer as, due to their antagonist nature, these percepts are mutually exclusive. Thirdly, a large interindividual variation was reported among dichromats. For the less keen observers, “red”-“green” were used indifferently to refer to colors seen as red or green by normal trichromats. Yet, some dichromats were able to produce a naming pattern

indistinguishable from that of a normal trichromat, thus revealing a gap between perceived similarity and naming. For instance, if cap #2 was placed next to cap #13 in the arrangement test, in agreement with normal trichromat, cap #2 was named “blue-light” and cap #13 “reddish blue.” In this experiment, correct naming was explained by subtle luminance differences between red and green caps that could have been used by dichromats. To the careful dichromatic observer the simple rule “darker, then red” provides a better-than-chance correct naming.

This rule would be more difficult to apply when colors are shown in isolation. This setup was used in an experiment with 140 Munsell color samples varying in hue, saturation, and brightness intended to probe naming and categorization. In the naming task, participants had to choose one of the eight chromatic basic color terms (BCT, namely, *red*, *green*, *blue*, *yellow*, *orange*, *pink*, *purple*, and *brown*) to describe each sample presented in isolation. Categories thus obtained for the two dichromats tested were very similar to normal trichromatic prototypical naming categories, with 66 % and 72 % of the samples described with the same BCT as used by normal trichromats. In the categorization task, color samples were sorted in eight categories based on their perceptual similarity with, this time, the overall sample collection displayed on a large table, inciting comparison strategy. Despite the possibility of samples’ visual comparison, local category inversions corresponding to cases where stimuli were assigned to groups in a way that is contrary to the structure of the nearest neighbor were observed in the categorization task, while absent in the naming task. Moreover, for the same number of categories, color naming produced a more consensual categorization pattern compared to perceptual similarity. Indeed, color names that define category boundaries in terms of hue (i.e., red, green, blue, yellow, orange, and purple), saturation (i.e., pink), and brightness (i.e., brown) further constrain the elicited categories by, for instance, excluding categories such as turquoise or pastel that would be legitimate based on similarity criteria. This added constraint in normal trichromats increased the consensus

between participants from 72 % (categorization task) to 82 % (naming task) and similarly from 66 % to 76 % in color deficient [5]. Language appears to operate normalization effect on categories obtained by naming in color deficient, yet it does not seem to be readily available to fine-tune color categories when these are based on perceptual similarities.

Explanations for Dichromat Color Naming and Categorization Abilities

Aside from subtle luminance differences or lightness cues, dichromat color naming and categorization performances have been explained by the contribution of other visual signals. In particular, rods distributed in peripheral retina and absent from the fovea are mediating low-light intensity vision and can act as a third photoreceptor. Under moderate light conditions, for surfaces larger than 4° of visual angle or presented in periphery, dichromats reveal trichromatic color matches comparable to normal trichromats' small-field (2°) color match, that is, for these larger field sizes, they need two primary lights (red and green) to match a spectral yellow light. Under these conditions, dichromat large-field trichromacy is explained by rod-cone interaction providing a residual red-green sensitivity [6]. In the naming of 424 OSA Uniform Scales samples presented in isolation, dichromats produced trichromat-like responses in conditions where rod-cone interactions were possible. When rod contribution was excluded by bleaching the photopigment by light preexposure, trichromat-like naming was no longer possible in the red-green dimension [7].

Dichromats still perform better than expected on perceptual tasks in conditions where luminance and rod-cone interaction are ruled out. For instance, in hue-scaling experiment, stimuli were monochromatic lights presented in a 2° field. Observers' task consisted in describing each stimulus by a given proportion of "yellow," "blue," "red," and "green." In this experiment, a molecular genetic analysis had confirmed that dichromat participants had only one functional X-chromosome-linked photopigment opsin gene

(either L or M), excluding the possible contribution of a residual red-green color vision. Dichromats reported the presence of a red component in the 420–450 nm interval and explicitly referred to blue and red mixture. These data suggest a dichromatic color space structure is richer than that illustrated by simulations based on linear model and reduction assumption as presented in Fig. 2. A model including a nonlinear transfer function of signals from S and L or M cones can account for a richer dichromatic perceptual space with enhanced naming and categorization abilities [8].

Non-chromatic visual cues, rod-cone interaction, or nonlinear transfer function could all provide supplementary visual information to dichromats to resolve ambiguities in an otherwise impoverished chromatic environment. However, color naming shows a more trichromat-like pattern than similarity or perceptual categorization tasks can elicit. The comparison between colors and color-naming structures has been addressed in similarity tasks using color cards and their corresponding color names printed on cards. Similarity judgments were requested for 36 pairs of stimuli. Apart from a violet-blue inversion, the multidimensional scaling analysis of dichromat color-name data provided a Newton color circle in a two-dimensional model with yellow-blue and red-green axis. For color similarities, the color circle was distorted along the red-green axis, bringing these two colors next to each other [9].

Dichromats have correctly learned the relational structure between color names as established by normal trichromats. Yet, as noted by Jameson and Hurvich [4], a dichromat's correct conceptual representation of color is not used to optimize performance on perceptual tasks. This finding indicates that the existence of an isomorphism between percept and concept structures is not compulsory; each type of representation coexists with no apparent conflict and can be used independently to fulfill specific task demands.

Cross-References

- [Color Categorical Perception](#)
- [Cone Fundamentals](#)

- [Environmental Influences on Color Vision](#)
- [World Color Survey](#)

References

1. Newton, I.: *Opticks: or a Treatise of the Reflections, Refractions, Inflections and Colours of Light*. Smith & Walford, London (1704)
2. Viénot, F., Brettel, H., Ott, L., Ben M'Barek, A., Mollon, J.D.: What do color-blind people see? *Nature* **376**, 127–128 (1995)
3. Hunt, D.M., Dulai, K.S., Bowmaker, J.K., Mollon, J.D.: The chemistry of John Dalton's color blindness. *Science* **267**, 933–1064 (1995)
4. Jameson, D., Hurvich, L.M.: Dichromatic color language: "reds" and "greens" don't look alike but their colors do. *Sens. Processes* **2**, 146–155 (1978)
5. Bonnardel, V.: Naming and categorization in inherited color vision deficiencies. *Vis. Neurosci.* **23**, 637–643 (2006)
6. Smith, V.C., Pokorny, J.: Large-field trichromacy in protanopes and deuteranopes. *J. Opt. Soc. Am.* **67**, 213–220 (1977)
7. Montag, E.D., Boynton, R.M.: Rod influence dichromatic surface color perception. *Vis. Res.* **27**, 2153–2152 (1987)
8. Wachtler, T., Dohrmann, U., Hertel, R.: Modeling color percepts of dichromats. *Vis. Res.* **44**, 2843–2855 (2004)
9. Shepard, R.N., Cooper, L.A.: Representation of colors in the blind, color-blind, and normally sighted. *Psy. Sci.* **3**, 97–110 (1992)

Color Category Learning in Naming-Game Simulations

Tony Belpaeme

School of Computing and Mathematics,
University of Plymouth, Plymouth, UK

Synonyms

[Computer modeling](#); [Linguistic relativism](#);
[Simulation](#)

Definition

Color categories are inextricably linked to language as color categories are typically, though

not necessarily, associated with color terms. It is believed that the acquisition of categories, including color categories, is influenced by language [1]. Prelinguistic infants do seem to have a set of color categories, which are then either consolidated or modified through observing or engaging in linguistic interactions about color.

Language has been shown to have an influence on a range of modalities, such as time, space, and color. This phenomenon, known as linguistic relativism, shows how the language one uses, and by extension the culture one lives in, has an impact on perception and cognition. It has also been shown that language has an influence on color perception: having a particular color word speeds up spotting a chip of that color among distracting color chips [2, 3]. What is not entirely clear is how language influences the acquisition of color categories. As data on color category acquisition in infants is hard to come by, we can resort to computer simulations to learn more about how language impacts the acquisition of color categories.

A language is a communication system that is shared within a group of language users. As such, a language can be seen as an agreement between all language users on the words and rules of a language, and their meaning. Color words are also subject to this agreement: in the English language speakers agree to use "red" for, among others, the chromatic perception of a ripe tomato, blood, and a light with dominant spectral wavelength of 780 nm. There is no central authority insisting on this: language users themselves agree on this convention. When a new language user, such as a newborn child, enters a linguistic community, it will to varying degrees adopt this convention.

If language is a convention that is agreed upon by a linguistic community, and if language impacts category acquisition, then it follows that categories are to a certain extent also agreed upon by the community. Computer simulations can help us understand how a community can arrive at an agreement on linguistic conventions and how language shapes concepts and categories.

There are a number of simulation models that can elucidate the process of language acquisition. The Language Game model [4, 5] has proven to

be both effective and popular and can be used to study how a group of individuals reach a consensus on linguistic forms and associated categories. If, for example, a new word is introduced in a language, the Language Game model can simulate how that word spreads through the population. The model as such serves to study the dynamics of language change, and by setting parameters of the model, one can study what conditions make a language change. Iterated Learning Models, an alternative model to Language Games, study the sequential transfer of language [6]. Individuals are placed in a chain, and each individual's output is used as input for the next individual. Language Games study horizontal transmission, while Iterated Learning Models study vertical transmission of language. Both models demonstrate how small biases and communication bottlenecks can have a large effect on the language and conceptual structures that arise. A third class of simulated Language Game models are based on Evolutionary Game Theory (e.g., [7, 8]) or Statistical Mechanics (e.g., [9]). These are in essence mathematical models which start from a minimal set of parameters and study the influence of different settings of these parameters.

Language Game Models

In Language Game models, a community of language users is modeled as a population of N software agents. Each agent can store and recall words (or other linguistic information, such as rules) and categories. In the domain of color, the agents store color terms and color categories. In addition, each agent stores associations between color terms and color categories. An association typically is a value showing how strong the association between a term and category is. Agents start with empty inventories and gradually fill these with words and categories.

Various Language Game models represent color in different ways. Color can be modeled as a point on a single circular dimension [7, 8, 10, 11]; a color stimulus is then a real number in the interval $[0, 1]$. For a more realistic model of color,

one can endow the artificial agents with a color appearance model, such as the CIE $L^*a^*b^*$ color space [12]. In this space, each color is represented by three real numbers L^* , a^* , and b^* , with L^* representing lightness, a^* the amount of green or magenta, and b^* the amount of yellow and blue. The CIE $L^*a^*b^*$ color appearance model aims to provide an accurate representation of color perception differences and allows for a similarity measure to be calculated between two colors, which is done by taking the Euclidean distance between two color values, permitting a good first approximation to categorical color perception (see [13] for an experimental appraisal and extension of CIE $L^*a^*b^*$).

In addition to the color categories and color terms used by an agent, simulations also need to prescribe what agents do when interacting with each other. In one form of a Language Game model interaction two agents are selected at random from the population; one acts as a *speaker*, the other as a *hearer*. Both agents are presented with a *context*; this is a set of M random color stimuli, each at a distance d from each other. The distance d guarantees that colors are not too similar or identical. From the context one color stimulus is selected, this will be the topic, and the speaker will attempt to communicate what the topic is to the hearer.

The speaker first finds a category that best matches the topic (often, but necessarily, this category is a unique match, meaning that it matches no other stimuli in the context). If no category can be found, the speaker will adapt its category set by adding a new category. Next it finds a word associated with the category and communicates it to the hearer. The hearer will attempt to guess the topic by looking up the word and the associated category in its inventory. It will check which stimulus matches the category best and will "point out" the stimulus. The hearer will then signal if this stimulus is indeed the intended topic. If it is, the game is successful. If the hearer points out any other stimulus, the game fails. When successful, categories and word-category associations in both agents are reinforced, with the categories used in the interaction adapting such that they match the topic

closer. When the game fails, the associations are weakened [4, 5].

During the iterative playing of Language Games [6] the agents create novel categories and words and change existing categories and category-word associations to optimize the communication. Only when communication is good enough (as determined by a preset threshold, e.g., $\tau = 90\%$ of all games end in success) will the agents' internal representations stop changing. It is important to note that the internal representation of all agents at this moment will not be identical and that the agents do not necessarily have the same number of words and categories. Their internal representations are merely sufficiently coordinated for communication to succeed with a success rate of τ . A population typically has a size N between two and several thousand agents and will play tens of thousands or perhaps millions of Language Games before stabilizing. An interesting observation is that the lexical and category systems of the agents adapt until they are "good enough" to successfully communicate; the agents do not need identical words and identical categories; they only need to sufficiently overlap to allow successful communication in most interactions. As such the semantics of words differ between agents: indeed your red is different from my red. As such, word and category are not true descriptions of the world; they are merely useful [14].

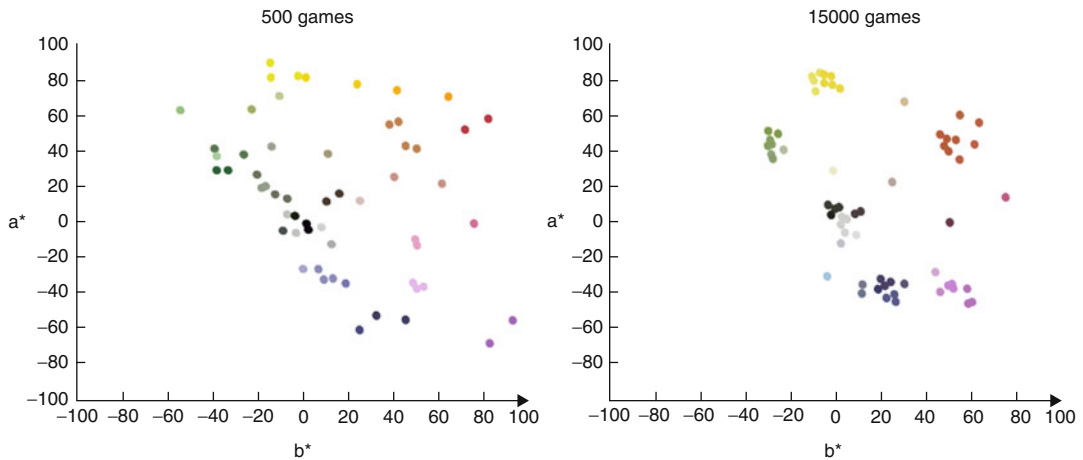
The dynamics of Language Game models have been extensively studied, as they inform research into the diachronic evolution of language. For example, the conditions under which a new word or a new linguistic construction is taken up by a language community can be modeled using Language Games [15, 16], and model predictions have been confirmed in studies with human participants [17]. Language Games have been used to clarify the minimum constraints needed to evolve a shared color category system by populations of agents [7], how varying agent perceptual properties impact color category system evolution [18–20], and how varying color salience affects color category evolution [20]. In the case of color, however, the Language Game model serves a different purpose. It helps us to understand how

relatively small biases present in color communications can have large-scale effects on the evolution of color category systems. Small biases are amplified through repeated interactions between language users. Specifically, simulated Language Game models help us formally investigate factors likely to influence color communications and help us understand why color categories appear to be universal and the degree to which pragmatics of communication or culture may contribute to color category evolution.

Explaining the Universal Character of Color Categories

It has been suggested that human color categories exhibit a universal pattern: many cultures have color categories that are seemingly similar. This was first suggested based on tenuous evidence in 1969 by Berlin and Kay [21] and later refined in the World Color Survey [22–24]. As such, color categories are not arbitrary, and this infused a principled search for the basis of their universal character. There are cultures which deviate sufficiently from the universal pattern, virtually ruling out the possibility that color categories are genetically determined. Other processes must be at work, and computer simulations can help us elucidate these.

When agents in a Language Game are communicating about color stimuli, their categories soon converge on a limited set of categories (Fig. 1). The locations in which the categories converge are not predetermined; instead they result from the slight biases introduced by the perceptual system of the agents [12, 18–20, 25]. Two different runs of a simulation will result in two different end results. However, repeated simulations do exhibit a pattern in which some color categories consistently emerge from the interaction between the agents. Categories such as red, green, yellow, blue, white, black, and so forth tend to almost always emerge. This matches the observations in the World Color Survey: these basic color terms and categories are found in the majority of the world's languages.



Color Category Learning in Naming-Game Simulations, Fig. 1 Both plots show all color categories of ten agents plotted in the a^*b^* plane of the CIE $L^*a^*b^*$ color space (the L^* lightness dimension is projected onto the a^*b^* plane; as such *white* and *black* categories are plotted on *top* of each other). The *left* figure shows the state after

500 Naming Games; the categories are still scattered in the color space. The *right* figure shows the state after 15,000 games when the communicate success of the agents is over 85 %; the categories have gravitated to a limited number of clusters

As the repeated playing of Language Games forces agents to coordinate their color categories and color terms, small biases in the agents' color perception will have a large influence. Belpaeme, Bleys and Steels [12, 23] showed how the bias of the CIE $L^*a^*b^*$ color appearance model together with a repeated negotiation of linguistic conventions results in the emergence of universal patterns of color categories. Baronchelli et al. [11] refined this; again using a Language Game model they showed how the human Just Noticeable Difference (JND) function, a function which shows the wavelength differences that are just about distinguishable to the human eye for each wavelength in the visual spectrum, also provides a small but important bias that can explain the universal character of color categories.

Language Games show how a variety of factors may contribute to the universal character of color categories without the need for color categories to be explicitly genetically determined. They permit the evaluation of, for example, neurophysical properties of human color perception as well as other small biases which, through repeated linguistic negotiations, amplify and can

contribute to the similarities seen across groups of languages that have roughly similar color categories.

References

1. Bowerman, M., Levinson, S. (eds.): Language Acquisition and Conceptual Development. Cambridge University Press, Cambridge, UK (2001)
2. Gilbert, A., Regier, T., Kay, P., Ivry, R.: Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci.* **103**(2), 489–494 (2006)
3. Roberson, D., Pak, H., Hanley, J.R.: Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition* **107**(2), 752–762 (2008)
4. Steels, L.: Evolving grounded communication for robots. *Trends Cogn. Sci.* **7**(7), 308–312 (2003)
5. Steels, L.: Modeling the cultural evolution of language. *Phys. Life Rev.* **8**(4), 339–356 (2011)
6. Smith, K., Kirby, S., Brighton, H.: Iterated learning: a framework for the emergence of language. *Artif. Life* **9**(4), 371–386 (2003)
7. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. *J. Math. Psychol.* **51**, 359–382 (2007)

8. Steingrímsson, R.: Evolutionary game theoretical model of the evolution of the concept of hue, a hue structure, and color categorization in novice and stable learners. *Adv. Complex Syst.* **15**(03n04), 1150018 (2012)
9. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**(2), 591 (2009)
10. Puglisi, A., Baronchelli, A., Loreto, V.: Cultural route to the emergence of linguistic categories. *Proc. Natl. Acad. Sci.* **105**(23), 7936–7940 (2008)
11. Baronchelli, A., Gong, T., Puglisi, A., Loreto, V.: Modeling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci.* **107**, 2403 (2010)
12. Belpaeme, T., Bleys, J.: Explaining universal colour categories through a constrained acquisition process. *Adapt. Behav.* **13**(4), 293–310 (2005)
13. Komarova, N.L., Jameson, K.A.: A quantitative theory of human color choices. *PLoS One* **8**, e55986 (2013)
14. Hoffman, D., Singh, M.: Computational evolutionary perception. *Perception* **41**, 1073–1091 (2012) (Special issue in honor of David Marr)
15. de Vylder, B., Tuyls, K.: How to reach linguistic consensus: a proof of convergence for the naming game. *J. Theor. Biol.* **242**(4), 818–831 (2006)
16. Loreto, V., Steels, L.: Social dynamics: emergence of language. *Nat. Phys.* **3**(11), 758–760 (2007)
17. Centola, D., Baronchelli, A.: The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proc. Natl. Acad. Sci.* **112**(7), 1989–1994 (2015)
18. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. I. Population categorization systems based on normal and dichromat observers. *J. Opt. Soc. Am. A* **26**(6), 1414–1423 (2009)
19. Jameson, K.A., Komarova, N.L.: Evolutionary models of categorization. II. Investigations based on realistic observer models and population heterogeneity. *J. Opt. Soc. Am. A* **26**(6), 1424–1436 (2009)
20. Komarova, N.L., Jameson, K.A.: Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *J. Theor. Biol.* **253**(4), 680–700 (2008)
21. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
22. Kay, P., Berlin, B., Maffi, L., Merrifield, W.: *The World Color Survey*. Center for the Study of Language and Information, Stanford (2003)
23. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci.* **100**(15), 9085–9089 (2003)
24. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within – language diversity. *Proc. Natl. Acad. Sci.* **106**(47), 19785–19790 (2009)
25. Steels, L., Belpaeme, T.: Coordinating perceptually grounded categories through language. A case study for colour. *Behav. Brain Sci.* **24**(8), 469–529 (2005)

Color Centers

Richard J. D. Tilley

Queen's Buildings, Cardiff University, Cardiff, UK

Definition

Color centers are point defects or point defect clusters associated with trapped electrons or holes in normally transparent materials. These centers cause the solid to become colored when the electronic ground state of the defect is excited to higher energy states by the absorption of visible light [1–5]. [Note that transition metal and lanthanoid ion dopants that engender color in an otherwise colorless matrix are frequently called color centers. These are dealt with elsewhere (see “[Cross-References](#)”).]

Color Centers

The concept of color arising from point defects was initially developed in the first half of the twentieth century, principally by Pohl, in Germany. It was discovered that clear alkali halide crystals could be made intensely colored by diverse methods, including irradiation by X-rays, heating crystals in the vapor of any alkali metal, and electrolysis. Crystals with induced color were found to have a lower density than the crystals before treatment and appeared to contain a population of anion vacancies. The absorption spectrum was always a simple bell shape. It was notable that the color engendered in the crystal was always the same and was not dependent upon the method of color production. That is, if a crystal of KCl was heated in an atmosphere of any alkali metal, or irradiated by X-rays, or electrolyzed using any cathode material, the crystal took on a violet color. Similarly, crystals of NaCl always took on an orange-brown hue under all preparation methods. The fact that the color was a unique property of the host structure implied that it was a

Color Centers, Table 1 Some color centers

Host crystal	Symbol	Description
Alkali halide: MX	F	Electron trapped an anion vacancy
	M, F ₂	A pair of adjacent interacting F centers
	F _A	F center next to an alkaline metal substitutional impurity
	F', F ⁻	F center with 2 trapped electrons
	R, M ⁺ , F ₂ ⁺	Three adjacent interacting F centers on (111)
	V _K	Two adjacent anion vacancies with 1 trapped electron
Alkaline earth halide: MX ₂	F	Electron trapped an anion vacancy
	M	A pair of adjacent F centers aligned along [100]
	F ₃	Three adjacent F centers aligned along [100]
Alkaline earth oxide: MO	F	Oxygen vacancy with two trapped electrons
	F', F ⁻	Oxygen vacancy with three trapped electrons
	F ⁺	Oxygen vacancy with one trapped electron
Quartz: SiO ₂	E', E ⁻	Oxygen vacancy with one trapped electron
Diamond: C	C, P1	Isolated N atom substituted for a C atom
	A	Two C atoms substituted by N, forming an N–N pair
	N3	Three N atoms on C sites surrounding a C vacancy
	N2	Two N atoms on C sites adjacent to a C vacancy
	NV	One N atom on a carbon site adjacent to a C vacancy
	NV ⁻	Negatively charged NV center

property of the crystal itself. The color was ultimately attributed to the formation of defects called *Farbzentren* (color centers), which were equated with mistakes in the crystal structure.

Since these early studies, many different color center types have been characterized (Table 1). All are electronic defects that possess similar characteristics in that they form in colorless, insulating, often relatively ionic, solids and incorporate either trapped electrons, to produce *electron excess centers*, or trapped holes, to produce *hole excess centers*. Solids may contain several different types of color center, including populations of both electron and hole excess centers in the same host matrix. Color centers are usually labeled with a letter symbol.

As the alkali halide studies demonstrated, color centers can be created in a host matrix in a number of ways. For example, strong ultraviolet light can transform clear glass into purple-colored “desert glass” and intense radiation from nuclear weapons or accidents may color ceramics such as porcelain a blue color, both changes being due to the formation of color centers. Controlled irradiation in nuclear reactors or similar is used to produce artificial gemstones from colorless and less



Color Centers, Fig. 1 Blue topaz, produced by irradiation of clear crystals; the color arises from a population of color centers (Photograph R J D Tilley)

valuable starting materials. For example, irradiation of colorless topaz, Al₂SiO₄(F, OH)₂, induces a beautiful blue color due to color center formation (Fig. 1).

From an optical viewpoint, color centers behave something like isolated atoms dispersed

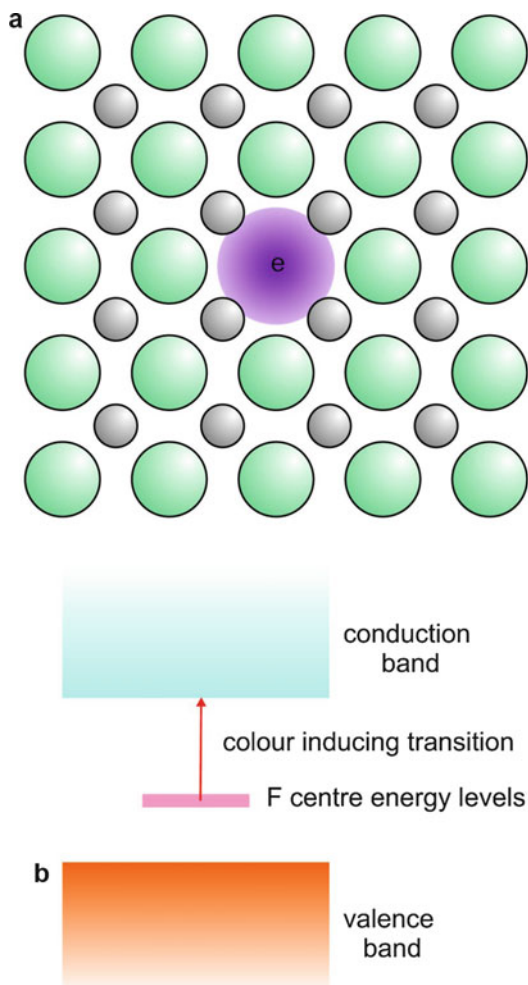
throughout the host matrix. As such, they make ideal probes of the interaction of light with the matrix and are currently being explored for wide-ranging electro-optic applications.

F Centers: Electron Excess Centers

The first color center to be characterized was the F center found in alkali halides, and these remain the best-known electron excess centers. Anion vacancies, which occur in low concentrations in alkali halides, MX, have an effective positive charge, and an F center is an anion vacancy (V_X^-) plus a trapped electron to form an effectively neutral defect ($V_X^- e$) (Fig. 2a). The F center behaves rather like an isolated hydrogen atom in the structure, and the electron is able to absorb electromagnetic radiation, jumping from one energy level to another, just as an electron absorbs radiation in the Bohr model of the H atom. The peak of the absorption curve, I_{\max} , corresponds to a total removal of the electron from the F center, and usually falls in the visible, so coloring the originally transparent crystals (Table 2). In terms of band theory, alkali halides are insulators with a considerable energy gap between a filled valence band and an empty conduction band. The F center in its ground state creates a new energy level or narrow energy band in the band gap (Fig. 2b). The color-producing optical absorption peak corresponds to electron promotion into the conduction band. The variation in the color observed depends upon the host crystal band gap and the energy level of the F center, both of which are linked to the lattice parameter of the host matrix (Table 2).

The detailed mechanism for the formation of F centers depends to some extent upon the manner in which they are generated. In the case of irradiation by X-rays, for example, the energetic radiation is able to displace an electron from a normal anion, and some of these become trapped at existing anion vacancies. The corresponding anion that has lost an electron creates a hole energy level in the valence band.

F centers occur in many alkaline earth halides and oxides (Table 1). For example, the mineral *Blue John* is a rare naturally occurring purple-blue form of fluorite, CaF_2 . The coloration is caused by F centers believed to have formed when the



Color Centers, Fig. 2 (a) An F center in an alkali halide MX crystal (schematic); *large circles*, anions X^- ; *small circles*, cations, M^+ ; (b) schematic energy level diagram for an F center in an alkali halide crystal

fluorite crystals were fortuitously located near to uranium compounds in the rock strata. Radioactive decay of the uranium produced the energetic radiation necessary to form the color centers.

Hole Excess Centers

One of the best understood hole excess centers gives rise to the color in smoky quartz, a naturally occurring form of silica, (SiO_2). This material contains small amounts of Al^{3+} substitutional impurities. These replace Si^{4+} ions in $[\text{SiO}_4]$ tetrahedral units which form the building units of the crystal. Overall charge neutrality is preserved by

Color Centers, Table 2 Alkali metal halide F centers

Compound	Absorption peak λ_{\max}/nm	Color ^a	Lattice parameter/nm
LiF	235, ultraviolet	Colorless	0.4073
NaF	345, ultraviolet	Colorless	0.4620
KF	460, blue	Yellow brown	0.5347
RbF	510, green	Magenta	0.5640
LiCl	390, ultraviolet (just)	Yellow green	0.5130
NaCl	460, blue	Yellow brown	0.5641
KCl	565, green	Violet	0.6293
RbCl	620, orange	Blue green	0.6581
LiBr	460, blue	Yellow brown	0.5501
NaBr	540, green	Purple	0.5973
KBr	620, orange	Blue green	0.6600
RbBr	690, red	Blue green	0.6854

^aThe appearance of the crystal is the complementary color to that removed by the absorption band.

the incorporation of one H^+ for each Al^{3+} . These H^+ ions sit in interstitial positions in the rather open SiO_2 structure. The color center giving rise to the smoky color in quartz is formed when an electron is liberated from an $[\text{AlO}_4]$ unit by ionizing radiation and is trapped on one of the H^+ ions present, leaving a hole (h) behind. The color center is a $(\text{AlO}_4 \text{ h})$ group. The color arises when the trapped hole absorbs radiation exactly as the electron in an F center.

The situation in amethyst, which is a form of silica containing Fe^{3+} and H^+ impurities, is similar. The impurity Fe^{3+} forms $[\text{FeO}_4]$ groups. These crystals are known as the pale yellow semiprecious gemstone citrine and also in a pale green form, the color arising from the impurity Fe^{3+} ions (see “[Cross-References](#)”). On irradiation, $(\text{FeO}_4 \text{ h})$ color centers form by interaction with H^+ ions. The color centers impart the purple amethyst coloration to the crystals (Fig. 3).



Color Centers, Fig. 3 Amethyst crystals; the purple color arises from a population of color centers. The pale green stone is unirradiated (Photograph R J D Tilley)

Some Applications of Color Centers

F Center Lasers

F centers do not exhibit laser action but F centers that have a dopant cation next to the anion vacancy are used in this way. These are typified by F_{Li} centers, which consist of an F center with a lithium ion neighbor. Crystals of KCl or RbCl doped with LiCl, containing F_{Li} centers, have been found to be good laser materials, yielding

emission lines with wavelengths between 2.45 and 3.45 μm . A unique property of these crystals is that in the excited state an anion adjacent to the F_{Li} center moves into an interstitial position. This is *type II* laser behavior, and the active centers are called F_{Li} (II) centers. These centers are stable if the crystal is kept at -10°C .

Persistent Luminescence

Color centers are active in materials that show persistent luminescence. In these compounds, irradiation by ultraviolet light present in normal daylight gives rise to luminescence for many

hours after dark. Such materials are making their way into applications as diverse as road signs that do not need a power supply at night and bicycle frames that glow in the dark.

Although the color centers responsible for persistent luminescence vary from material to material, the principle can be illustrated with the oxide SrAl_2O_4 doped with B^{3+} , Eu^{2+} , and Dy^{3+} , which gives a green luminescence. The SrAl_2O_4 structure consists of a framework of corner-linked $[\text{AlO}_4]$ tetrahedra enclosing Sr^{2+} ions in the cavities so formed. The B^{3+} substitutes for Al^{3+} to create $[\text{BO}_4]$ tetrahedra and $[\text{BO}_3]$ triangular groups, which can be thought of as $[\text{BO}_4]$ tetrahedra in which one of the oxygen ions is absent to form an oxygen vacancy (V_O), creating a $(\text{BO}_3 \text{V}_\text{O})$ center. The substitution of Dy^{3+} for Sr^{2+} results in a charge imbalance that is compensated for by the creation of an equal number of vacancies on Sr^{2+} positions (V_Sr) to form defect clusters $(\text{Dy BO}_4 \text{V}_\text{Sr})$. Following irradiation with ultraviolet light, an electron is transferred from a $(\text{BO}_3 \text{V}_\text{O})$ cluster to a $(\text{Dy BO}_4 \text{V}_\text{Sr})$ unit, resulting in the formation of two complex color centers: $(\text{Dy BO}_4 \text{V}_\text{Sr} \text{h})$, which are hole excess centers, and $(\text{BO}_3 \text{V}_\text{O} \text{e})$, which are electron excess centers. The origin of the luminescence lies in this pair of color centers. Under normal conditions, the electron and hole centers are metastable, and over the course of several hours the holes and electrons gradually recombine. The energy liberated is transferred to the Eu^{2+} ions, which lose energy by the emission of photons, thus producing a long-lasting green fluorescence.

Photostimulable Phosphors

Photostimulable phosphors that make use of color centers are widely used in X-ray imaging, particularly by dentists, where they have largely replaced X-ray film recording. The first commercial material to fulfill these requirements, introduced in 1983, was BaFBr doped with Eu^{2+} . Although the detailed mechanism by which these phosphors work is still not entirely clear, it is established that an important component of the process is the formation of F centers, produced as a result of the X-ray irradiation. In dental X-ray imaging, a plate covered with a thin layer of

phosphor is placed into the mouth and exposed to X-rays. The X-rays initially displace an electron from an anion to form an electron-hole pair. The electron is subsequently trapped at an anion vacancy to form an F center. This fairly stable pair of electronic defects constitutes a *latent image* in the phosphor. Subsequently the exposed plate is irradiated with 633 nm light from a helium-neon laser. The electrons trapped in the F centers are initially promoted to the conduction band, after which they are free to recombine with the holes. The energy liberated is transferred to Eu^{2+} ions which decay from the subsequent excited state by the emission of visible light at 420 nm, subsequently recorded as a digital image. The number of F centers and holes, and therefore the amount of light emission, is proportional to the X-ray intensity in the phosphor. The optical image thus records accurately the degree to which the X-rays have penetrated the subject.

Color Centers in Diamond

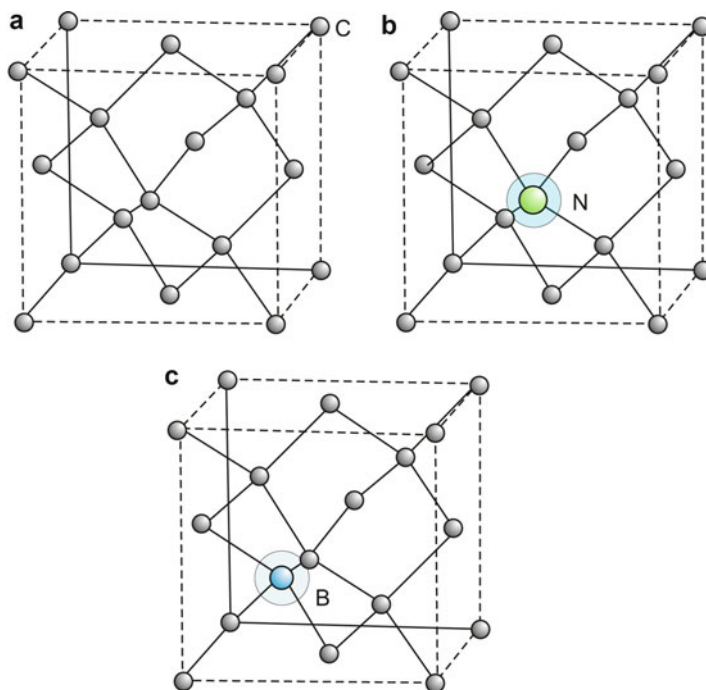
Colored Diamonds

The diamond structure is built up of carbon atoms each coordinated to four carbon atom neighbors, the linking being via tetrahedral sp^3 hybrid bonds (Fig. 4a). Diamond has a band gap of about 5.5 eV which is too large to absorb visible light, and perfect diamonds are clear. The commonest impurity in natural diamonds is nitrogen, which mostly substitutes for carbon on normal tetrahedral sites in the crystal. Natural diamonds are often subjected to temperatures of 1,000–1,200 °C, over geological timescales, allowing these nitrogen atoms diffuse through the structure, leading to populations of defect clusters as well as isolated point defects.

The color of the highly prized natural yellow diamonds called *Canaries* is due to isolated nitrogen atoms located on carbon sites, which form color centers called C centers. The color arises in the following way. Nitrogen, with an electron configuration $1\text{s}^2 2\text{s}^2 2\text{p}^3$, has five bonding electrons, one more than carbon, with a configuration $1\text{s}^2 2\text{s}^2 2\text{p}^2$. Four of the electrons around each impurity nitrogen atom are used to fulfill the

Color Centers, Fig. 4

(a) The diamond structure as a linkage of tetrahedrally coordinated carbon atoms (*small circles*); (b) structure of a C center in diamond, consisting of a nitrogen atom (*large circle*) occupying a carbon position, schematic; (c) structure of a boron impurity center (*large circle*) in diamond (schematic)



local sp^3 bonding requirements of the crystal structure, and one electron remains unused. Substitution of nitrogen for carbon on a normal carbon atom site in the crystal thus creates an electron excess color center (Fig. 4b). On an energy level diagram, this gives rise to a donor level in the band gap, which, because of lattice vibrations and other interactions, consists of a narrow band of energies centered at 2.2 eV and extending to 1.7 eV, the ionization energy of the N atom in diamond. The electron can be excited into the conduction band by absorption of incident visible light of wavelengths longer than about 564 nm, giving the stones a faint yellow aspect. As the number of C centers increases, the color intensifies.

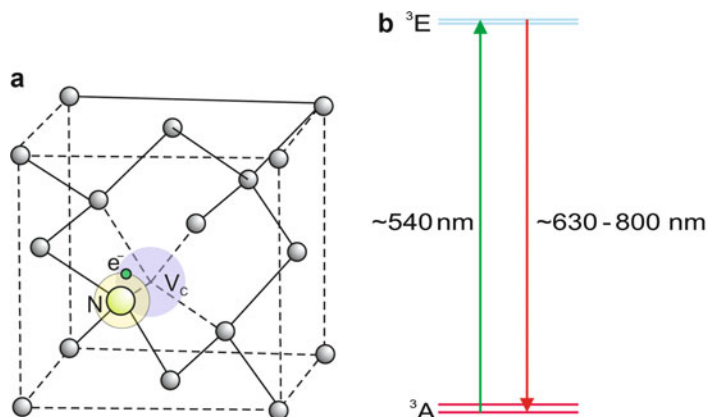
The N3 center, which consists of three nitrogen atoms on neighboring carbon sites adjacent to a carbon vacancy, seems to be responsible (at least in part) for the pale straw color of expensive *Cape Yellow* diamonds. The N3 center has a complex electronic structure which absorbs just in the blue end of the visible, at 415 nm, resulting in yellowish stones. The N3 centers are often accompanied by neutral N2 centers consisting of two nitrogen atoms on neighboring carbon sites adjacent to a carbon vacancy. These absorb at approximately

475 nm, giving a yellow color to the stones and adding to that contributed by the N3 clusters. When crystals are irradiated, either naturally or artificially, the N2 cluster can trap an electron to form a negatively charged $N2^-$ center that has an absorption peak at approximately 989 nm in the infrared. This absorption band can spill over into the red part of the visible spectrum, leading to stones with a blue tone and producing blue diamonds. When all these nitrogen-based color centers are present in roughly equal quantities, the stones take on a green hue.

Although nitrogen-linked color centers give rise to the highly valued yellow-hued diamonds, many prized blue diamonds are the result of boron impurities on normally occupied carbon atom sites (Fig. 4c). Boron, with an electron configuration $1s^2 2s^2 2p^1$, has only three outer bonding electrons instead of the four found on carbon. These three are used in fulfilling the bonding requirements of the structure, but one bond of the four is incomplete and lacks an electron, making the defect a hole excess center. In semiconductor physics terms, the center introduces a narrow band of energy levels approximately 0.4 eV above the valence band. The transition of

Color Centers, Fig. 5 (a)

An NV^- center in diamond, schematic; (b) approximate energy level diagram of an NV^- center



an electron from the valence band to this band gives rise to an absorption peak in the infrared with a high-energy tail encroaching into the red at 700 nm. The boron-doped diamonds therefore absorb some red light and leave the gemstone with an overall blue color.

Nitrogen-Vacancy Centers

The diamond color center that has been studied in most detail is that consisting of a single nitrogen substitutional impurity located next to a carbon vacancy, together with a trapped electron to form a *negatively charged nitrogen-vacancy center*, NV^- (often just called an NV center). These defect centers are readily created by the irradiation of nitrogen-containing natural or synthetic diamonds, diamond thin films, or diamond nanoparticles with high-energy protons. The proton irradiation results in the formation of carbon vacancies, and if the crystals are then annealed at above 600 °C, the temperature at which the vacancies become mobile, they diffuse through the structure until they encounter a nitrogen impurity. The strain around the nitrogen atom effectively traps the vacancy, preventing further migration. In the resultant NV^- centers, the tetrahedron surrounding the carbon vacancy is composed of three carbon atoms and one nitrogen atom (Fig. 5a).

These centers are being investigated for applications, including room temperature quantum computing, nanoscale magnetometers, and fluorescent markers. The applications follow from the unique features of the energy levels of the (NV^-)

center. The ground state term of the electronic structure is ³A and the first excited state term is ³E (Fig. 5b) [see ► [Transition-Metal Ion Colors](#), for a description of ³A and ³E terminology]. (Note that both the ground state and excited state terms are split into several levels. This splitting is of prime importance for many applications [6] but does not dominate the overall color aspects of the centers and can be ignored in the present context.) Excitation from the ³A ground state to the ³E excited state is by absorption of light over the range of approximately 514–560 nm, giving stones a pink hue. Emission falls in the range 630–800 nm, but the observed color is dominated by a particularly strong red fluorescence at 637 nm. Under suitable observing conditions, single bright red fluorescent (NV^-) centers can be observed, making nanodiamonds that incorporate these defects ideal probes to track vital processes in living cells.

Postscript

The electronic structure of the neutral NV and negative NV^- color centers in diamond have been intensively studied, and it has taken some 35 years to reach the current understanding of energy levels of this defect. It would seem reasonable to suspect that similarly detailed investigations of the other color centers described will also lead to significant revisions in their electronic structures and consequently a more precise description of their color-engendering abilities.

Cross-References

- ▶ [Lanthanoid Ion Color](#)
- ▶ [Transition-Metal Ion Colors](#)

References

1. Tilley, R.J.D.: *Defects in Solids*. Wiley, Hoboken (2008)
2. Tilley, R.J.D.: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
3. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
4. Hayes, W., Stoneham, A.M.: *Defects and Defect Processes in Nonmetallic Solids*. Wiley, New York (1985). Reprinted by Dover, Mineola (2004)
5. Fowler, W.B.: Chapter 2. The imperfect solid – color centers in ionic crystals. In: Hannay, N.B. (ed.) *Treatise on Solid State Chemistry. Defects in Solids*, vol. 2. Plenum, New York (1975)
6. Acosta, V., Hemmer, P (eds.): *Nitrogen-Vacancy Centres: Physics and Applications*. M. R. S. Bulletin, vol. 38, Materials Research Society, PA 15086-7573, USA (2013)

Color Circle

Paul Green-Armytage
School of Design and Art, Curtin University,
Perth, Australia

Synonyms

[Color wheel](#); [Hue sequence](#)

Definition

The color circle, as generally understood and widely used, is a diagram with a continuous sequence of hues arranged in the order of the spectrum. (The gap between spectral red and spectral violet is bridged with extra-spectral purples.) The color circle diagram is used as a guide to color mixing and color composition. It is also used in the classification of colors and is incorporated in all three-dimensional color order systems.

Introduction

Very many color circle diagrams have been designed and published. The essential feature is that the diagram must represent the sequence of hues in correct order and in a continuum: reds, oranges, yellows, greens, blues, purples, reds, oranges, yellows, greens, etc. The number of separate hues in the sequence can vary as can the hues themselves that are selected for inclusion. The starting point for some color circles is a choice of so-called primary colors. Other circles are organized so that colors opposite to each other are, in some sense, complementary. Many color circles are also organized so that the degree of difference between neighboring colors in the sequence appears to be the same all round the circle.

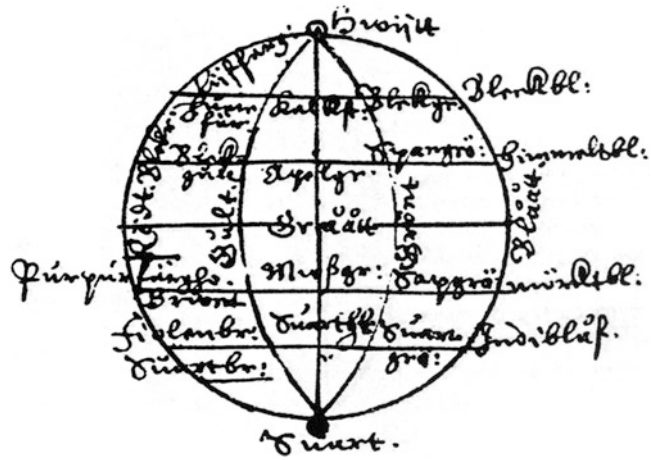
The Variety of Color Circles

There is no single “correct” design for a color circle. Different circles have been constructed on different principles. They are not necessarily presented in color; in some the colors are simply identified by name. Color circles can be grouped in three broad categories: those that represent colors as something physical, those that represent colors as visual phenomena – what people see –, and those where the colors can be understood as either or both, physical and/or visual. In the first category the colors represent lights, paints, inks, or dyes, and the position of the colors in relationship to each other is generally determined by their physical properties or by the way that these lights, paints, inks, or dyes can be mixed to produce a large range of other colors. In the second category the colors are simply themselves and it is their appearance that determines how they are related in the circle. In the third category it may not be clear whether the colors are to be understood as physical or visual or both. It could be that the designers and users of such circles confuse the physical and the visual aspects of color. For an account of the way that the physical and visual aspects of color can be confused, see the entry on “Primary Colors” in this encyclopedia.

As color circles fall into different categories, and are constructed according to different principles, the relationship between colors can vary from circle to circle. This does not mean that the

Color Circle,

Fig. 1 Circular diagram from the manuscript by Sigfrid Forsius (Reproduced from Ref. [6], with permission)



sequence of hues can vary – that must always be the same – but the distance between hues can vary. For example, it could be that two different circles have yellow in the 12:00 o’clock position, but red might be at 3:00 o’clock in one circle and at 4:00 o’clock in the other. The principles underlying the design of the color circles would have fixed the position of red in relation to yellow. Problems can arise when there is a mismatch between the intentions of the designer and the expectations of the user. A particular color circle can be criticized for not representing particular color relationships when it was never the intention that such relationships be embodied in the design.

Precursors to the Color Circle

The color circle, as defined above, was invented by Isaac Newton, whose color circle is illustrated in his book *Opticks*, first published in 1704 (see Fig. 4) [1, pp. 32–35]. Before Newton it was believed that there was a beginning and an end to the sequence of colors. Aristotle had claimed that all colors derived from mixtures of white and black which he placed at either end of a linear scale [2, pp. 31–32]. Aristotle’s prestige was so great that science was dominated by his ideas until the time of Galileo [3, p. 213]. Some circular color diagrams do predate Newton but they still reflect Aristotle’s ideas.

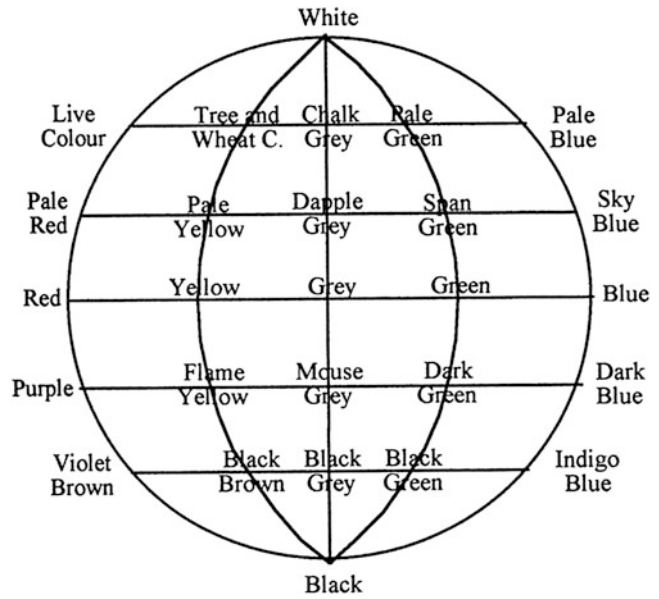
John Gage describes a circular diagram that survives in an illuminated manuscript from the fifteenth century. This shows the colors of urine, between white and black, that helped physicians

diagnose certain diseases [4, pp. 162 and 171]. White and black also appear in other circular diagrams, two in a manuscript by Sigfrid Forsius (1611) and another in a printed medical text by Robert Fludd (1629–31). Fludd’s circle has this sequence: white, yellow, orange, red, green, blue, and black [2, p. 42]. Forsius echoes Aristotle when he writes: “Among colours there are two principles, white and black, from which all others arise” [5, pp. 12–13]. The Forsius manuscript contains two circular diagrams. Each has white in the 12:00 o’clock position and black at 6:00 o’clock. Forsius explains that his first circle illustrates the way in which the “Ancients” arranged the colors. Down the left side, between white and black, are pale yellow, yellow, orange, red, purple, brown, and violet. Down the right side are ash gray, gray, sky blue, blue, pale green, green, and dark green. The second Forsius diagram is shown in Fig. 1 and recreated in Fig. 2 with color names translated from old Swedish.

Claims have been made that this second Forsius diagram represents a sphere [7, p. 224]. If this is correct then the central horizontal line must be read as a circle seen from the side with red, yellow, green, and blue on the circumference and gray in the middle. Werner Spillmann has applied the principles of graphic projection to this interpretation and points out that it would mean that the hues are in the wrong order [8, p. 7]. Yellow must be closest to the observer and green farthest away or vice versa. Either way, the

Color Circle,

Fig. 2 Reconstruction of the Forsius diagram with color names translated from old Swedish



hue sequence would be red, yellow, blue, and green, which is not the order in which they appear in the spectrum. However these early diagrams are read, not one of them, unequivocally, shows the sequence of hues as continuous and in the correct order.

Development of the Color Circle

The color circle did not evolve so much as develop a growing variety of uses. Newton's circle identifies colors with light of different wavelengths and he shows how the diagram can be used to illustrate the results of additive mixing. Some later circles illustrate the results of subtractive mixing from a set of "primaries." In other circles the positions of the colors are determined by how they appear. Such arrangements are used in systems of color identification. Color circles have been developed where the apparent difference between neighboring colors is the same all round the circle. Colors that are opposite to each other in some circles are described as being complementary. Even spacing and complementary relationships are seen as significant in theories of color harmony.

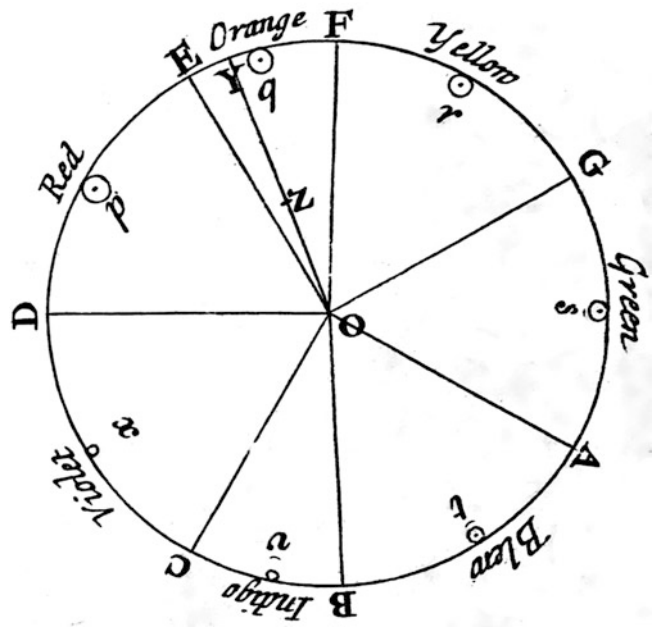


Color Circle, Fig. 3 Lindsay MacDonald demonstrating how a prism separates white light into different colors

Newton

In Fig. 3 Lindsay MacDonald is showing how white light is refracted by a prism to reveal the different colors. Although not easily visible here, the blue shades into a slightly reddish violet at the short-wave end of the spectrum. Perhaps it was the redness at each end of the spectrum that gave

Color Circle, Fig. 4 Isaac Newton's color circle representing colors as different wavelengths of light (Reproduced from Ref. [6], with permission)



Newton the idea of connecting the two ends to form his circle. Newton's color circle is shown in Fig. 4.

Strictly speaking, the sequence of hues in Newton's circle is incomplete. The circle is divided into seven segments, each identified by name. At first Newton refers to each segment as representing a single color – “Let the first part DE represent a red, the second EF orange . . .” – but he goes on to say that these represent “all the colors of uncompounded light gradually passing into one another as they do when made by Prisms . . .” [1, p. 32]. So the segment DE does not represent a single red but a range of the hues that would be identified as “red.” And the lines that separate the segments represent the borders where colors that would be identified by one name give way to those that would be identified by the next; the “red” segment would have a range of colors from reds to orange-reds. Because the colors in the circle represent the wavelengths of light in the visible spectrum, there is no place for the bluish reds and purples that are not visible in the spectrum. And if the hues shade into each other across the borders separating the other named segments, there would be a break in the sequence at point D with no spectral hues to shade from violets to reds.

A color circle which does show all the hues, spectral and non-spectral, shading into each other was produced by Michel-Eugène Chevreul and published in 1864 (Fig. 5). This can be set beside another of Chevreul's circles (Fig. 6) which is divided into 72 hues. Chevreul's color circles are described by Verena Schindler [9, pp. 66–68].

The colors in Chevreul's circle between the 5:00 o'clock and 7:00 o'clock positions are not visible in the spectrum and so have no place in Newton's circle. However, if Newton had intended to represent seven hues only, with no shading from one named hue to the next, the gap between violet and red would be no more noticeable than that between red and orange. The hues are in the correct order and the sequence is continuous.

The First Color Circles Published in Color

The first color circles to be published in color appear in an enlarged edition of a book on miniature painting. The author of the first edition of 1673 has been identified as Claude Boutet, but the color circles were only added in the enlarged edition of 1708 in a new section on pastel painting. The unknown author of this later section writes: “Here are two circles by which one will be able to see how the primitive colors, yellow,

Color Circle,

Fig. 5 Michel-Eugène Chevreul 1864. Color circle with hues shading into each other (Reproduced from Ref. [6], with permission)



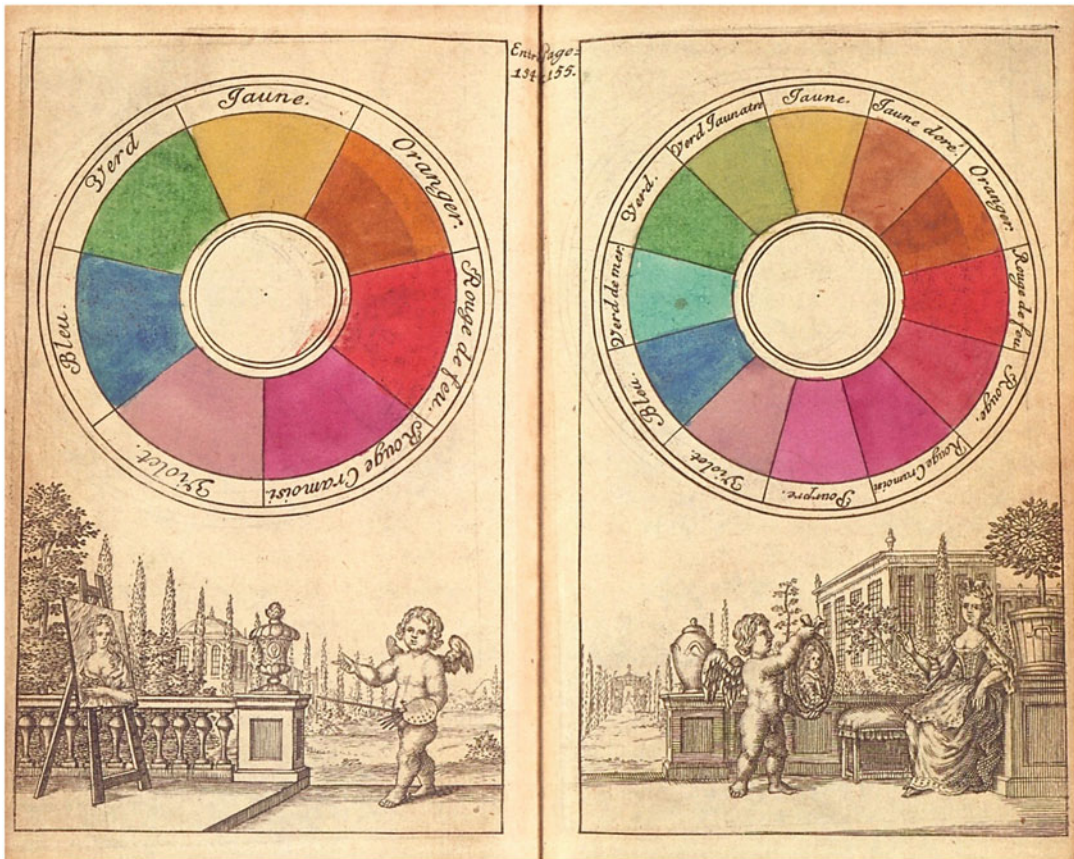
Color Circle,

Fig. 6 Michel-Eugène Chevreul 1864. Color circle with 72 discrete hues (Reproduced from Ref. [6], with permission)



fire red, crimson red and blue generate the other colors” [2, p. 57]. So these circles are demonstrations of subtractive mixing with paints (Fig. 7). The possibility of mixing a complete sequence of hues from just three “primitive colors” was

already known at that time so the use of two “primitive” reds is striking. Perhaps the author did not think that any available red pigment could qualify as “true red” so two reds had to be used, one a yellowish red and the other bluish.



Color Circle, Fig. 7 Color circles from the enlarged 1708 edition of the *Treatise on Miniature Painting* (Reproduced from Ref. [6], with permission)

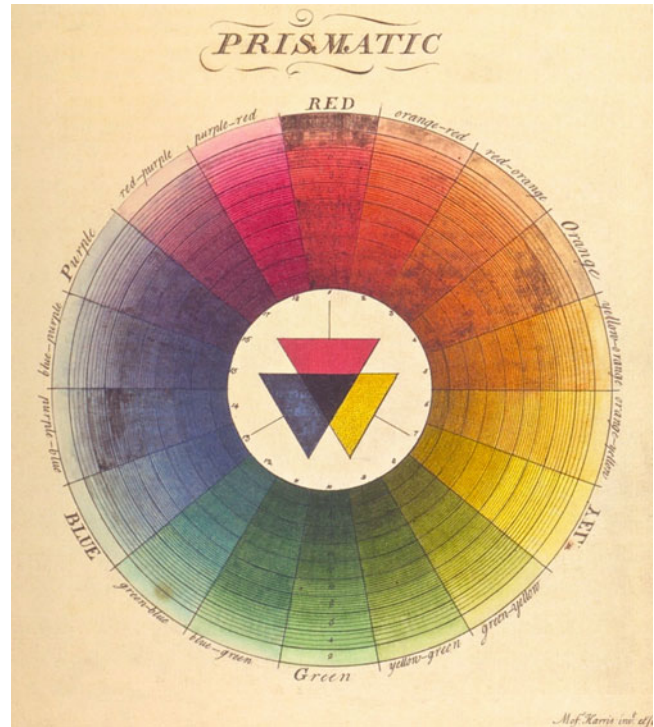
And perhaps no single red could deliver a satisfactory orange as well as a satisfactory purple. In this respect these early color circles foreshadow the Color Bias Wheel devised by Michael Wilcox [10, p. 15]. To make sure that an artist can achieve vivid colors all round the circle, Wilcox has two blues and two yellows as well as two reds in his Color Bias Wheel.

Harris

Moses Harris held to the belief that “all the variety of colours . . . can be formed from Red, Blue, and Yellow” [11, p. 3], but he seems to have recognized the shortcomings of available pigments. He explains that he “treats on colour in the abstract” [11, p. 7]. This suggests that he had in mind some kind of theoretical ideal. As he points out, “Colour which we may call material,

or artificial, are very imperfect in themselves, and being made of various substances . . . maketh the colouring part extremely difficult . . .” [11, p. 7]. For Harris, red, blue, and yellow were “primitives” which could be mixed to produce the “mediates” purple, green, and orange. He does list representative pigments: vermilion, ultramarine, and king’s yellow for the “primitives” and sap green and red orpiment for two of the “mediates” (no pigment is listed for purple). The fact that he lists separate pigments for his “mediates” suggests that these pigments would have been used, in addition to vermilion, ultramarine, and king’s yellow, to paint the color circles in his book – this is to make his demonstration of the theory more convincing with acceptably vivid colors all round the circle (Fig. 8).

Color Circle, Fig. 8 Color circle by Moses Harris 1772 (Reproduced from Ref. [6], with permission)



Harris was an entomologist as well as an accomplished artist. Reference to his color circles would have been helpful when identifying and recording the colors of butterflies and other insects, while the circles could also serve as a guide to mixing paints to match those colors. Furthermore, Harris may have been the first to point out that a color circle can reveal relationships between colors that would now be called complementary. He refers to “contrasting colors” that are “so frequently necessary in painting” [11, p. 6] and goes on to explain how one should “look for the colour . . . in the system, and directly opposite to it you will find the contrast wanted” [11, p. 6]. And he provides a kind of definition: “if the colours so mixed are possest of all their powers, they then compose a deep black” [11, p. 7]. One current definition of complementary colors is that they should mix to a neutral – white from additive mixture in the case of lights, gray from partitive mixing in the case of colored segments on a spinning disc, and near black from subtractive mixing in the case of paints. At the end of the book, Harris describes the phenomenon of colored shadows.

He explains that a stick placed in the orange light of a candle will cast a blue shadow, this result being predictable from the positions of orange and blue on opposite sides of his circle. A more extensive discussion of complementary colors can be found under that heading in this encyclopedia.

Goethe

Johann Wolfgang von Goethe is best known as a writer of novels, plays, and poetry, but he also wrote a book on color theory which remains influential today. Goethe was satisfied that “yellow, blue, and red, may be assumed as pure elementary colours, already existing; from these, violet, orange, and green, are the simplest combined results” [12, p. 224]. When these six colors are arranged in a circle, yellow is opposite violet, blue is opposite orange, and red is opposite green. Goethe places great emphasis on such relationships: “the colours diametrically opposed to each other in this diagram are those which reciprocally evoke each other in the eye” [12, p. 21]. This can be recognized in afterimages and Goethe describes his experience of a “beautiful

sea-green” when a girl wearing a scarlet bodice moved out of sight [12, p. 22]. He argues that such experiences show how the eye demands completeness. The red bodice gave way to green as a union of blue and yellow. Goethe saw in afterimages “a natural phenomenon immediately applicable to aesthetic purposes” [12, p. 320]. The afterimage phenomenon provides another way of defining complementary relationships with the added claim by Goethe that such relationships are beautiful. So Goethe is reinforcing the notion that the color circle can be seen as a tool for developing harmonious color combinations.

Chevreul

Michel-Eugène Chevreul developed a set of nine color circles, graded from full hue (Fig. 6) to almost black, and each with 72 hues, as the basis of a comprehensive color order system. In his introduction to a translation of Chevreul’s book *De la loi du contraste simultané des couleurs*, Faber Birren explains how “Chevreul devoted himself not only to color organization, color harmony, and contrast effects, but to methods of naming and designation of colors” [13, p. 29]. A special memorial edition of Chevreul’s book was published by the National Press of France in 1889 when Chevreul himself was 103 years old. The potential confusion between the physical and visual aspects of color is evident in a prefatory note to this edition: “In order to guarantee to the plates of this book the stability which their scientific nature requires . . . it was necessary to resort only to mineral colors whose stability was certain. . . . Since the three colors chosen by Chevreul as basic, *red, yellow, blue*, cannot be reproduced precisely by means of isolated materials, they were obtained by mixing” [13, p. 27]. If the “basic” colors red, yellow, and blue could only be obtained by mixing their status as “basic” would need to be clarified.

Hering

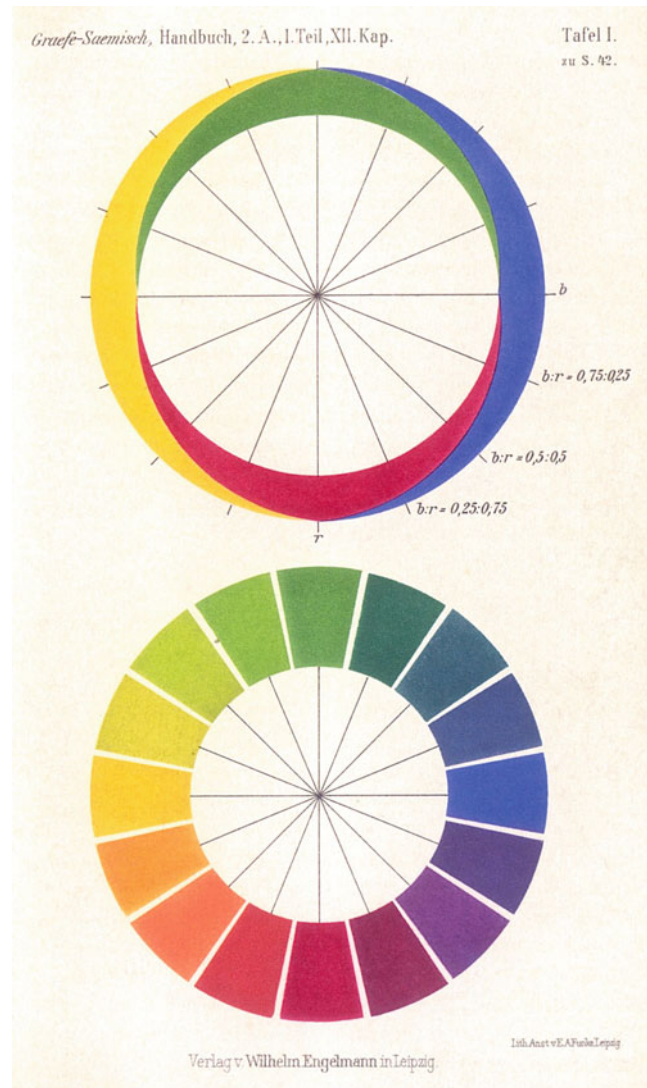
The primacy of red, yellow, and blue was challenged by Ewald Hering. The scientific orthodoxy that emerged during the nineteenth century was that three ► **primary colors** had their counterparts in the human eye in the form of three different

types of receptor cell, each tuned to one of these primaries. This notion was first proposed by George Palmer when he suggested that “the surface of the retina is compounded of particles of three different kinds, analogous to the three rays of light” [14, p. 41]. Thomas Young, working independently, came to a similar conclusion. Young suggested that the sensitive particles in the retina were associated with “the three principal colours, red, yellow, and blue” [15, p. 147]. In a subsequent lecture he referred to “three simple sensations . . . red, green and violet” [16, p. 440]. Hering could not reconcile any set of three primary colors with his own subjective experience. For Hering there are six basic color phenomena – six *ufarben* – white, black, yellow, red, blue, and green. Since it is possible to describe the hue of any color in relation to yellow, red, blue, and green, Hering proposed that “corresponding to the four hue variables . . . there are four physiological variables” [17, p. 48]. Hering, therefore, designed a color circle which represented colors simply as visual phenomena with yellow, red, blue, and green as primaries (Fig. 9).

The Natural Color System (NCS)

Hering’s theories were developed in Sweden and are the basis for the Natural Color System, NCS (Fig. 10). As with Hering’s color circle, the NCS has four “elementary colors” which are defined in visual terms as a yellow that is neither greenish nor reddish, a red that is neither yellowish nor bluish, a blue that is neither reddish nor greenish, and a green that is neither bluish nor yellowish [18, p. 132]. It is important to note that the NCS was designed just as a means of describing colors and showing how they are related as visual phenomena. The NCS is “value neutral in that it does not give rules for what is ugly and what is attractive” [19, p. 4]. There are no claims that it is to be used as a guide to color harmony. It has been criticized for not having the degree of difference between neighboring colors the same all round the circle. Although the colors are shown in a continuous circle, the four elementary colors are to be understood as the beginnings and ends of four separate hue sequences. The colors between red

Color Circle, Fig. 9 Color circles by Ewald Hering (Reproduced from Ref. [6], with permission)

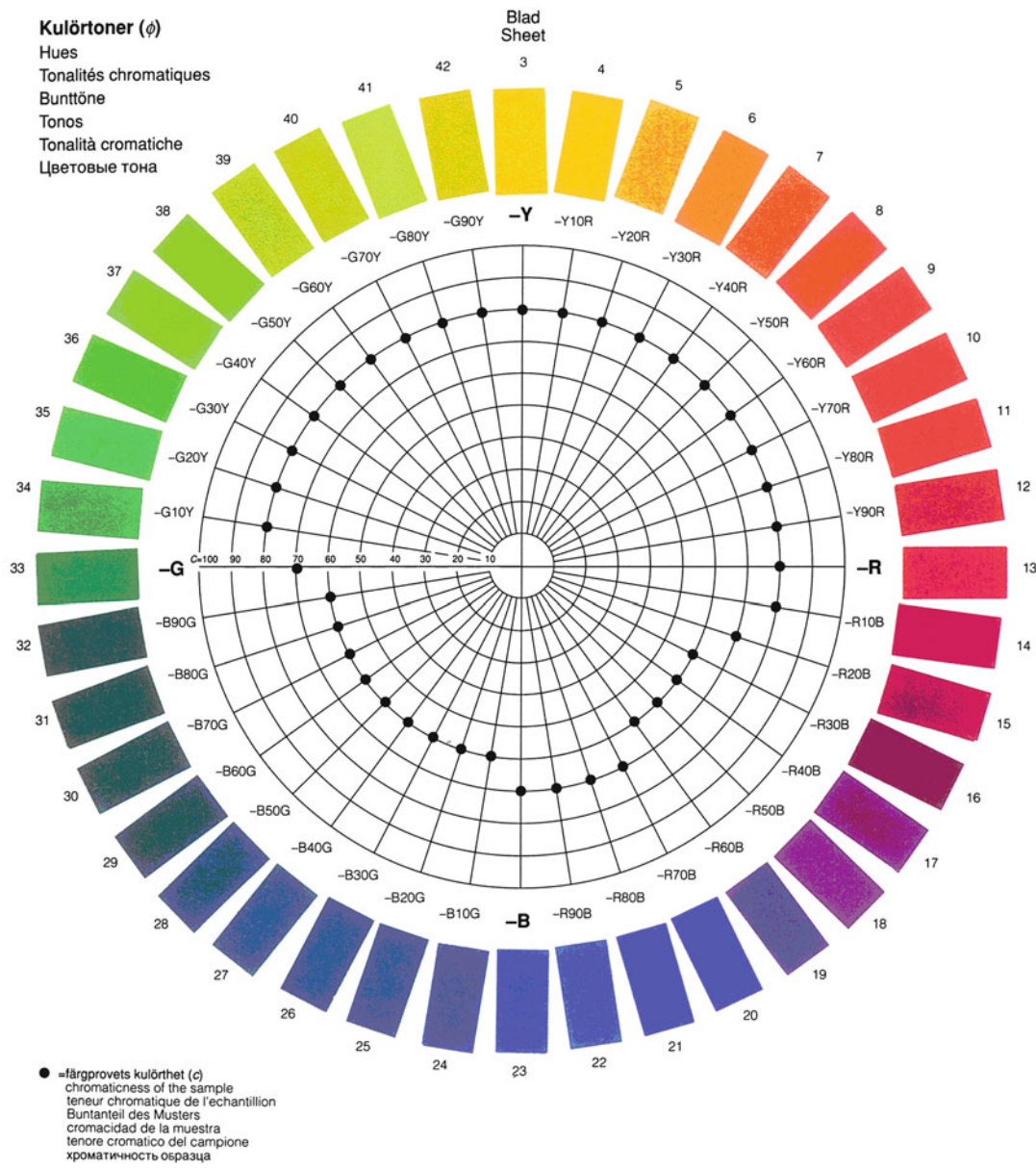


and blue are equally spaced visually as are the colors between blue and green, but there is a greater degree of visual difference in the red to blue sequence than in the blue to green.

Ostwald

Wilhelm Ostwald developed a color circle that is superficially similar to that of Hering and the NCS in that it is based on four rather than three “fundamental colors” (Fig. 11). However, these colors are not treated as the beginning and end points of four separate hue sequences but simply as landmarks in one continuous sequence. For Ostwald,

the color opposite to a given color should be the one that is “most different” so that “the entire circle is filled with such pairs of contrasting colors, which shall be called complementary colors” [20, p. 34]. The complementary relationship is established by optical mixture using a spinning disc. Segments of yellow and red on a disc would blend to a single color when the disc is spun, the blend in this case appearing orange. The complementary relationship is established when the blend appears a neutral gray. Ostwald chose “a pure yellow that is neither greenish nor reddish” [20, p. 33] as the starting point for his hue



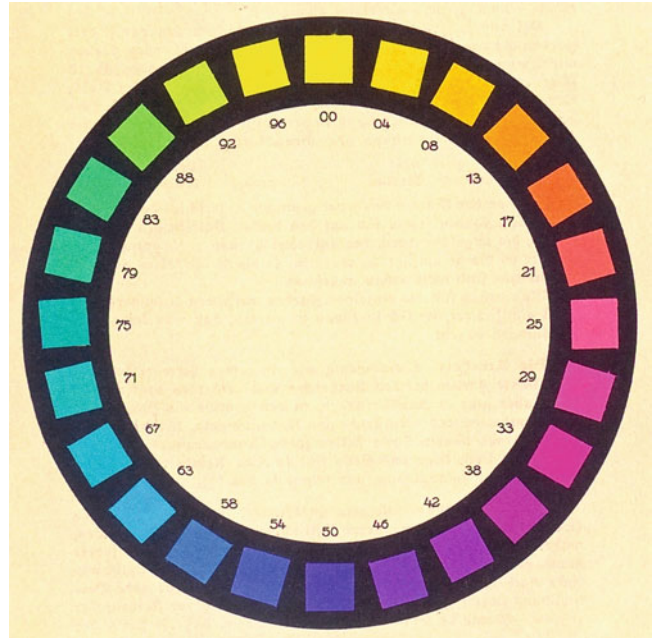
Color Circle, Fig. 10 Color circle of the Natural Color System, NCS (Reproduced from Ref. [6], with permission)

sequence. This would correspond to the elementary yellow of Hering and the NCS but the complementary of this yellow, as established by the spinning disc technique, is not quite a blue that is neither reddish nor greenish but one that is slightly reddish which Ostwald identifies as “ultramarine blue.” Similarly if Ostwald’s fundamental red is neither yellowish nor bluish, its complementary is

what Ostwald calls “sea green,” a green that is certainly bluish. So Ostwald’s four fundamental colors are not the exact equivalent of the Hering/NCS elementary colors. And Ostwald, unlike those who developed the NCS, did intend his system to be used for generating harmonious color combinations for application in the arts and design. For Ostwald his system represents “order”

Color Circle,

Fig. 11 Color circle by Wilhelm Ostwald (Reproduced from Ref. [6], with permission)



and he is famous for his “basic law”: “Harmony = Order” [20, p. 65].

Müller

Ostwald’s ideas were taken up and developed by Aemilius Müller who produced his *Swiss Colour Atlas* using dyes rather than pigments. Müller’s color circle has 60 hue steps (Fig. 12). Müller also produced a number of designs with beautiful color gradations to demonstrate Ostwald’s “basic law.” Müller’s work is described by Stephanie Wettstein [21, pp. 144–149].

The CIE Chromaticity Diagram

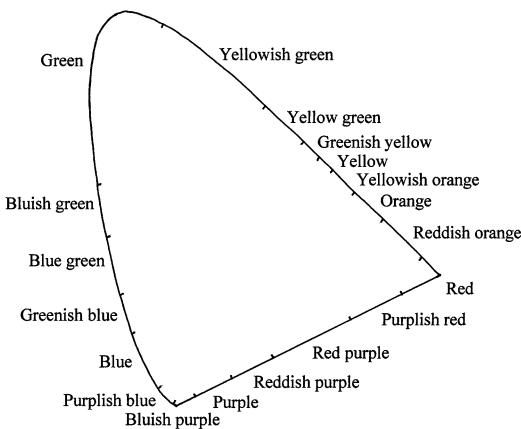
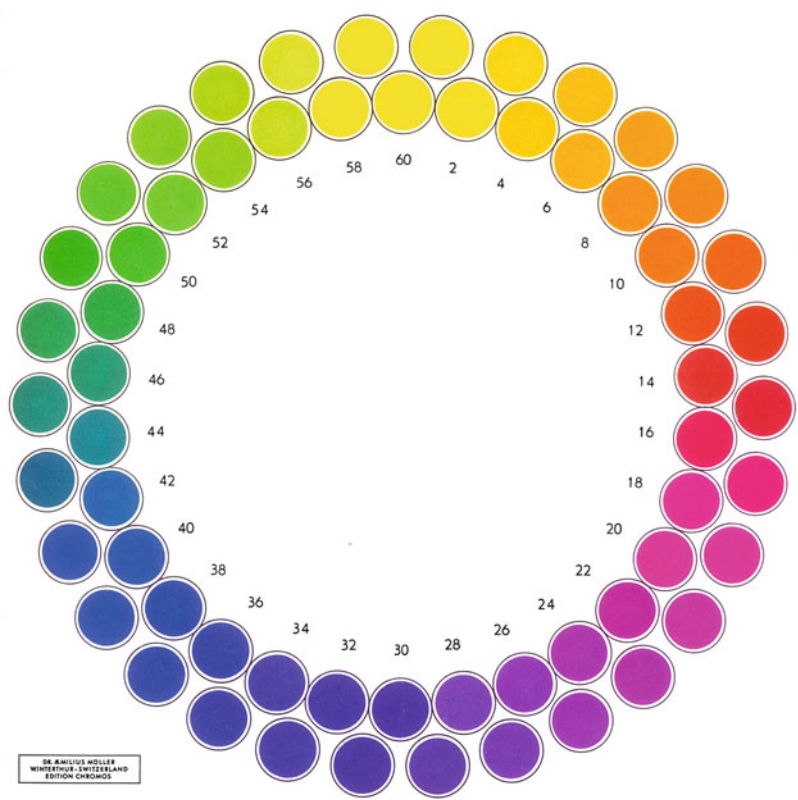
Although not geometrically circular, the 1931 CIE chromaticity diagram can be regarded as a color circle in that it represents the pure spectral hues and the extra-spectral purples in a continuous sequence. The diagram is used in the international system for measuring color stimuli. The diagram is often shown in color but Roy Berns warns against this as being misleading [22, p. 61]. No printing inks can match the purity and intensity of the spectral lights themselves. It is better simply to

show the line diagram marked out with color names, such as those proposed by Kenneth Kelly [23, p. 67], much as Newton did with the first color circle. The CIE diagram, with Kelly’s names, is shown in Fig. 13.

Munsell

Albert Henry Munsell developed his color system at the turn of the twentieth century and published *A Color Notation* in 1905. Munsell’s color circle has “five principal hues” [24, p. 20]. These are red, yellow, green, blue, and purple, which are spaced at equal intervals around the circle. In his sequence of hues, Munsell aimed at perceptual uniformity [2, p. 115]. Like Ostwald, Munsell believed in an ordered arrangement of colors as the key to harmony and suggested a number of paths through his system that would connect colors for a harmonious result. Figure 14 shows a page from the 1929 edition of the *Munsell Book of Color*. Twenty hues are included, each at several steps of increasing departure from neutral gray. Use of the CIE system to measure the Munsell color chips revealed a number of

Color Circle,
Fig. 12 Color circle by
Aemilius Müller
(Reproduced from Ref. [6],
with permission)



Color Circle, Fig. 13 CIE chromaticity diagram with
color names proposed by Kenneth Kelly

irregularities. A combination of instrumental measurement and visual judgments by members of an expert committee resulted in the Munsell renotations and the revised Munsell system that is widely used today.

The Color Circle Today

Today there are color circles in use that are based on three, four, and five primaries, but the three-primary model remains dominant with the primaries red, yellow, and blue; the secondaries orange, purple, and green; and 12 hues altogether. This is born out by an appeal to the Internet. Of the first 100 images from a Google search, made on December 7, 2013, more than half were twelve-hue circles and nine of these were the circle designed by Johannes Itten with primaries and secondaries identified in the center (Fig. 15).

Itten

Itten’s circle is attractive, clear, and memorable, but it needs to be viewed with caution. Itten was an artist writing for students of art. Artists work with pigments and so the “color classification must be constructed in terms of the mixing of pigments” [25, p. 21]. But Itten defines his primaries in terms of appearance: “a red that is

as used by printers. This is demonstrated in the entry on “Primary Colors” in this encyclopedia.

Itten’s color circle (Fig. 15) was recreated for the present entry by Lisa Hannaford using the

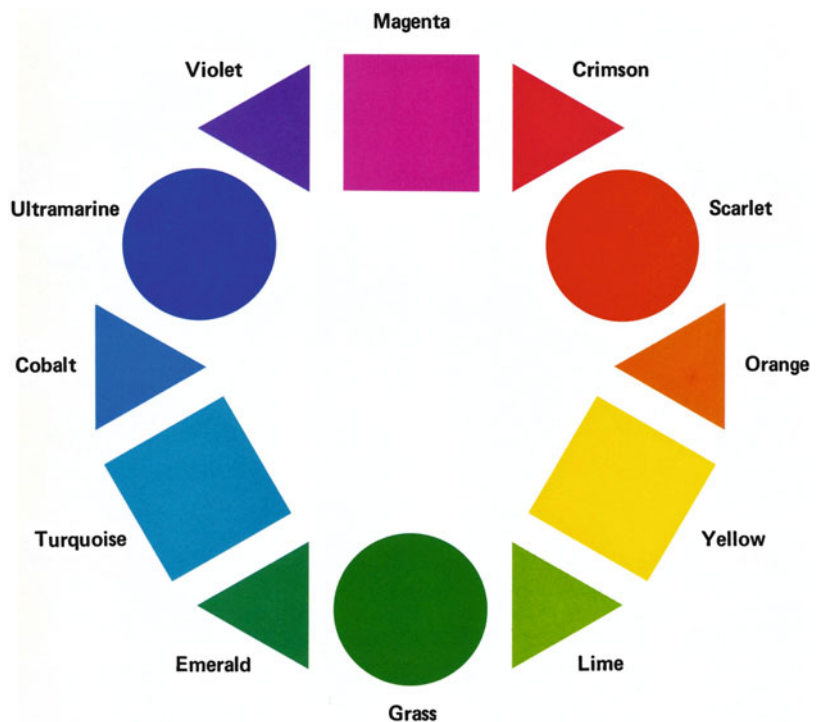
computer program *Illustrator*. As displayed on the computer screen, Itten’s diagram is created with the additive primaries red, green, and blue. When printed in hard copy, for Itten’s book as well as from the computer file, the inks used are the subtractive primaries cyan, magenta, and yellow. The relationship between the additive primaries and the subtractive primaries are shown in a color circle first presented at a conference in 1978 [27, p. 167, 28, p. 3 and cover] (Fig. 16). For this diagram, cyan is identified as turquoise, which is a more familiar color name. A shape code identifies the additive primaries as circles and the subtractive primaries as squares.

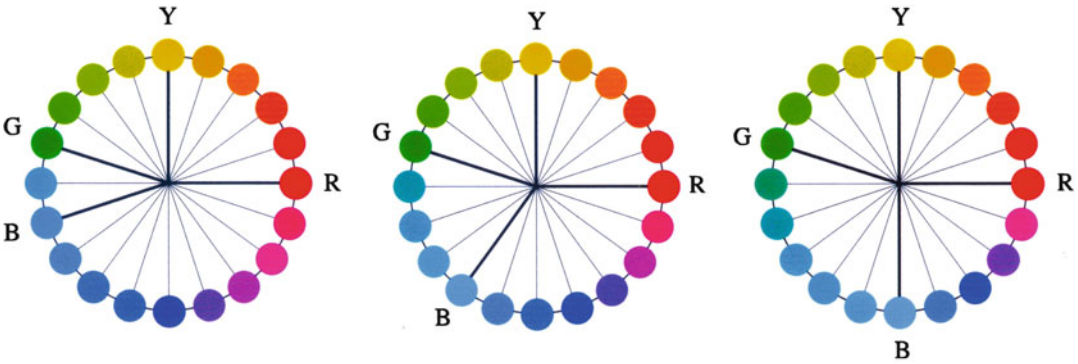
Perhaps Itten could have acknowledged the shortcomings of available pigments and followed the example of Moses Harris by explaining how he “treats on colour in the abstract.” If the diagram is misleading as a guide to mixing paints, it is reasonable to wonder what purpose it is intended to serve. Arnkil asks this question and suggests that it may have been “associated with his idea of the 12-colour circle as the basis of colour harmony” [26, p. 88].



Color Circle, Fig. 15 Color circle by Johannes Itten as reconstructed by Lisa Hannaford

Color Circle, Fig. 16 Color circle by Paul Green-Armytage showing the relationship between additive and subtractive primaries





Color Circle, Fig. 17 Color circles with complementary colors opposite to each other according to three definitions. From *left to right*: subtractive mixture to near black; the

generation of each other's hue in afterimages; optical mixture to neutral gray

Itten follows Goethe in asserting the significance of the afterimage phenomenon for color harmony, and he claims that opposite colors in his circle are complementary. Harmonious color combinations can supposedly be found by drawing a regular geometric figure inside the circle. Lines passing through the center of the circle, equilateral triangles, isosceles triangles, squares, and rectangles, as they touch the colors in the circle, all point to harmonious color combinations. As well as complementary pairs, Itten illustrates harmonious triads and tetrads – combinations of three and four colors. No doubt this is a useful starting point for people who lack confidence, but closer scrutiny reveals a problem. Itten defines complementary relationships in three ways, all of which have been encountered in the work of others: subtractive mixing to near black (Harris), afterimages (Goethe), and additive mixing to neutral gray (Ostwald). The theory would be more convincing if these different ways of defining complementary relationships yielded the same pairs, but this is not always the case. The most dramatic variation is with blue. The complementary of blue is red-orange from subtractive mixture, yellow-orange as an afterimage, and yellow from additive mixture. If these pairs are to be opposite to each other in different circles, the distribution of ► [hues](#) would need to be adjusted.

An Elastic Color Circle

Complementary relationships, as established in different ways, can be illustrated by stretching and compressing the color circle [29, p. 266]. With the elementary colors of the NCS as reference points, and by keeping yellow and red constantly at 12:00 o'clock and 3:00 o'clock, respectively, blue and green can be moved to new positions. The number of hue steps between the elementary colors will increase or decrease accordingly to bring the differently defined complementary pairs opposite to each other as shown in Fig. 17.

Conclusion

The color circle has a long history and is well established in color theory. It is used to represent colors as something physical, as lights, pigments, inks, or dyes, as well as colors as visual phenomena. It is used to illustrate relationships between colors, in systems of color classification and identification, as a guide to color mixing and as a tool in the search for harmonious color combinations. Although the sequence of hues is always the same, the intervals between hues can vary as the principles behind the construction of the circles varies. There is no single color circle that is "correct." Rather than try to establish a single color circle as some kind of standard, or to insist on a single

purpose for the color circle, it is more helpful to recognize that a color circle can embody a variety of information and that it may be necessary to stretch or compress the hues in the circle according to the information that is required.

Cross-References

- Appearance
- Chevreul, Michel-Eugène
- Color Combination
- Color Contrast
- Color Harmony
- Color Order Systems
- Complementary Colors
- Hering, Karl Ewald Konstantin
- Itten, Johannes
- Munsell, Albert Henry
- Ostwald, Friedrich Wilhelm
- Palmer, George
- Primary Colors
- Unique hues

References

1. Newton, I.: Opticks. In: MacAdam, D. (ed.) Sources of color science. The MIT Press, Cambridge, MA (1970 [1704])
2. Kuehni, R., Schwarz, A.: Color ordered. Oxford University Press, New York (2008)
3. Russell, B.: History of western philosophy. George Allen & Unwin, London (1961)
4. Gage, J.: Colour and culture. Thames and Hudson, London (1993)
5. Hård, A.: Quality attributes of colour perception. Fackskrift nr 8. The Swedish Color Centre Foundation, Stockholm (1969)
6. Spillmann, W. (ed.): Farb-Systeme 1611–2007: Farb-Dokumente in der Sammlung Werner Spillmann. Schwabe Verlag, Basel (2009)
7. Parkhurst, C., Feller, R.L.: Who invented the color wheel? Color Res. Appl. **7**(3), 217–230 (1982)
8. Spillmann, W.: Farbskalen – Farbkreise – Farbsysteme. Schweizerischer Maler- und Gipserunternehmer-Verband, Wallisellen (2001)
9. Schindler, V.M.: Michel Eugène Chevreul. In: Spillmann, W. (ed.) Farb-Systeme 1611–2007. Schwabe Verlag, Basel (2009)
10. Wilcox, M.: Blue and yellow don't make green. Artways, Perth (1987)
11. Harris, M.: The natural system of colours. Whitney Library of Design, New York (1963 [1766])
12. Goethe, J.W.: Theory of colours. The MIT Press, Cambridge, MA (1970 [1840])
13. Birren, F.: Introduction. In: Chevreul, M.E. (ed.) The principles of harmony and contrast of colors and their applications to the arts. Reinhold Publishing Corporation, New York (1967 [1854])
14. Palmer, G.: Theory of colors and vision. In: MacAdam, D. (ed.) Sources of color science. The MIT Press, Cambridge, MA (1970 [1777])
15. Young, T.: Theory of light and colours. In: Peacock, G. (ed.) Miscellaneous works of the late Thomas Young. John Murray, London (1855 [1801])
16. Young, T.: A course of lectures on natural philosophy and the mechanical arts. Johnson Reprint Corporation, New York (1971 [1807])
17. Hering, E.: Outlines of a theory of the light sense. Harvard University Press, Cambridge, MA (1964 [1920])
18. Hård, A., Sivik, L.: NCS – natural color system: a Swedish standard for color notation. Color Res. Appl. **6**(3), 129–138 (1981)
19. Smedal, G.: NCS – as a basis for colour training at the National College of Art, Craft and Design, Bergen, Norway. In: Hård, A., et al. (eds.) Colour report F 26. Scandinavian Colour Institute, Stockholm (1983)
20. Ostwald, W.: The color primer. Van Nostrand Reinhold, New York (1969)
21. Wettstein, S.: Aemilius Müller. In: Spillmann, W. (ed.) Farb-Systeme 1611–2007. Schwabe Verlag, Basel (2009)
22. Berns, R.: Billmeyer and Saltzman's principles of color technology. Wiley, New York (2000)
23. Agoston, G.A.: Color theory and its application in art and design. Springer, Berlin (1987)
24. Munsell, A.H.: A color notation. Munsell Color, Newburgh (1946)
25. Itten, J.: The elements of color. Van Nostrand Reinhold, New York (1970)
26. Arnkil, H.: Colours in the visual world. Aalto University, Helsinki (2013)
27. Green-Armytage, P.: Violets aren't blue: colour sensations and colour names. In: Condous, J., et al. (eds.) Arts in cultural diversity. Holt, Rinehart and Winston, Sydney (1980)
28. Green-Armytage, P.: Violets aren't blue, they're 'purple'. Gaz. Off. J. West. Aust. Inst. Technol. **12**(1), 2–6 (1979)
29. Green-Armytage, P.: The value of knowledge for colour design. Color Res. Appl. **31**(4), 253–269 (2006)

Color Combination

José Luis Caivano

Secretaria de Investigaciones FADU-UBA,
Universidad de Buenos Aires, and Conicet,
Buenos Aires, Argentina

Synonyms

Color coordination; Color harmony; Color mixture; Color syntax; Color union

Definition

Color combination is mainly an aspect of color syntax. To combine means to put one thing in relation to another, or several things together, so that the individuals lose significance and the meaning of the whole predominates. To combine also means to organize an ordered sequence. In some cases it acquires the sense of mixing or merging. However, mixing color pigments or lights normally yields just one color as the result of the mixture, and in this sense it cannot be termed a color combination, for which two or more colors in some relation must be perceived. Combination certainly is very closely connected to harmonization and coordination. Color combination, thus, is meant whenever there is more than one color associated, related, or harmonized with another: two colors already determine a certain kind of combination.

Overview

Color combination is, in principle, an aspect of color syntax. All perceptible colors can be organized in the so-called ► [color order systems](#) or models. This is usually made by means of three color variables or dimensions, for instance, hue, saturation, and lightness (HSL), or hue, blackness, and chromaticness (according to the Natural

Color System), or hue, value, and chroma (according to the Munsell color system), or some other similar triad of variables. These color order systems allow for the precise identification and notation of colors and their arrangement in a logical way. Some of the systems even allow to predict the results of color mixtures. It is possible to compare color order systems to dictionaries, which assemble all the words available in a language in alphabetical order. So, color order systems arrange and organize all the colors that humans can see, produce, and use, according to certain sequences determined by the mentioned color variables.

This possibility of having the repertoire of all perceptible colors orderly arranged facilitates the selection of colors, by following certain criteria, in order to use them in artistic compositions, architectural works, or pieces of design. Only in few rare occasions (for instance, in experimental situations) colors are seen isolated; in the great majority of cases (both in nature and in human productions), colors appear in a context where there are also other colors. That is to say, every color is combined in a certain way with other colors. Such as words (which in a dictionary appear isolated) are combined with other words in order to make sentences and phrases with some sense and give origin to poetry, narrative, essay, etc., and also in the same way as sounds are combined according to the criteria and invention of a music composer to give origin to musical pieces, so colors are grouped in larger compositional units. And it is the context, the particular combination, the way in which colors are grouped together and related to one another, what gives a sense, a certain kind of signification or meaning, some utility to the whole composition and to each of the involved colors.

Painters dispose and mix colors in their ► [palettes](#) with the final aim of combining them in a canvas. Architects combine materials with different natural colors in a building or either use paints to endorse different parts of their work with color. Filmmakers and directors of photography decide about the color sequences that appear in

successive scenes of the film. Clothing and fashion designers think about the chromatic combinations of the apparels they produce. Landscape designers choose and arrange the botanical species and other materials taking into account also color combinations. And it is possible to continue providing this kind of examples almost indefinitely, because there is practically no profession, discipline, or human activity in which color does not play a role.

If according to different authors and experiments, the number of perceptible colors may range, depending on various factors, from 2,000,000 to 7,000,000, the combinatorial possibilities rise to billions, even restricting them to the minimal expression of just two- or three-color combinations. Now, how are colors combined, with which kind of criteria, and in which type of contexts? At first, it is possible to talk about spatial and temporal contexts, depending on the colors being arranged simultaneously in an object or successively in a certain temporal sequence.

Syntactic Color Combinations

Spatial Color Combinations

The spatial color combinations have, at first, some basic and elementary rules. In terms of the abstract and logical possibilities and from the point of view of spatial arrangements, three possible cases can be pointed out for two-color combinations in a two-dimensional space:

1. That one color is applied over and inside another (interiority)
2. That both colors partially overlap each other (overlapping)
3. That they are juxtaposed one beside the other (juxtaposition)

The possibility of both colors being some distance apart is not considered here because in this case the color filling the separation, the background, appears as a third color. Also, there is no sense in considering a total superposition of both colors (both occupying exactly the same space), because in this case the result is just one

color, and hence, this cannot be termed a color combination.

These three possibilities produce different results or have different consequences for color light and for color-pigment combination and also differ if there is a mixture or blending of the involved colors or if opaque color surfaces that do not mix together are combined. Combining colors imply that in some cases the colors are mixed and give origin to new colors. However, if the result of the mixture is just one color, this will not be a color combination.

For instance, considering transparent color filters:

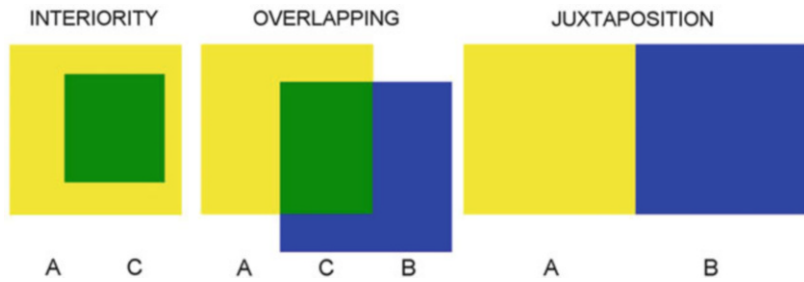
1. If over an area of a transparent color filter *A* another piece of color filter *B* is set in relation of interiority, the outcome is two colors: color *A* and a new color, *C*, which is the subtractive mixture of *A* and *B*, while color *B* is missed.
2. If the colors overlap, the result is three colors: *A*, *B*, and *C*.
3. In the case of juxtaposition, there is no color mixture, so that the result is color *A* just besides color *B*.

These combinations with their respective results are shown in Fig. 1.

Exactly the same happens with transparent inks and watercolors. Similar situations arise also when *A* and *B* are colored lights, but in this case color *C* is the result of an additive mixture. The situations are quite different with opaque color surfaces: in all cases, no new color *C* appears.

Additionally, in all situations – interiority, overlapping, and juxtaposition – phenomena of simultaneous contrast occur, so that, in reality, when considering colors *A* and *B* as seen in isolation, the perceptual result of the combination is, apart from the cases in which color *C* appears, colors *A1* and *B1*, because when being combined each color is tinged with the ► [complementary color](#) of the other, or with the other color, according to the principles of simultaneous contrast; i.e., color sensations change from seeing color stimuli in isolation to seeing color stimuli in combination.

Color Combination,
Fig. 1 Basic possibilities
 of two-color combinations
 in a two-dimensional space



Now, if the combined colors have relatively small areas to be perceived individually, an additive mixture is produced when they consist of color lights (as in the case of the TV screen), a partitive mixture occurs when they are small pigmented color surfaces (as with the pointillist technique of painting), and a mixed syntheses occur – partially subtractive and partitive – if the small dots are made of transparent inks that in some zones overlap each other and in some others are separated on the background (as in the case of color printing).

Temporal Color Combinations

Phenomena of contrast appear whenever two or more colors are combined in a certain relation, but if this is a temporal combination, where the colors appear in a sequence, with certain durations and intervals, what is produced is a successive contrast or the phenomenon of post-image.

When the time span of visualization of a color that is followed by another is long enough, an adaptation to the first stimulus occurs, and, as a consequence, the second stimulus is affected by the successive color contrast.

























When the frequency of appearance of two or more colors is fast enough to fall below the perceptible threshold (as in a flickering situation), an optical mixture of the colors involved in the sequence will be produced. It also happens here that two or more colors combined in these conditions give only one color as a result, the color that is the consequence of the optical mixture.

Color Selection as the First Step for Harmonic Color Combinations

If the specific chromatic relations among the combined colors are taken into account, the field of

► **color harmony** appears. There are a lot of proposals and theories about this. From a purely syntactical point of view, paying attention to the relations among the colors themselves and the quantity of colors combined (two or more colors), it is possible to mention, for instance, a combination of monochromatic colors, complementary colors, split complementaries, double complementaries, analogs, color triads or trichrome combinations, tetrachrome combinations, etc.

César Jannello had a logical way to face the issue of color harmony. He used to pose the aesthetic problem in design in terms of constancy or variation of perceptual variables: too much constancy produces boredom, too much variation generates visual chaos. Thus, it is in-between these two extremes that a fruitful field of harmonies in design can be found. Starting from the three color variables or dimensions – for instance, hue, saturation, and lightness – there are just eight possibilities for the selection of colors, whether these variables are kept constant or change. In Fig. 2, the sign plus (+) means constancy, and the sign minus (–) represents variation of the considered dimension. The first formula, the one in which everything is constant, is not of much use because it gives as a result the selection of just one and the same color (even when it may be boring, a color combination where the same color is repeated is possible, however). In the remaining formulas, where some type of variation appears, the interval of variation may be kept constant or may change according to some criterion, for instance, by modifying hue, lightness, or saturation in regular steps or intervals; by increasing intervals; by choosing opposite poles; etc. This model provides a logical basis for the selection of colors to be applied in a combination.

	3 CONSTANTS		2 CONSTANTS		1 CONSTANT			0 CONSTANTS
	1	2	3	4	5	6	7	8
HUE	+	+	+	—	+	—	—	—
SATURATION	+	+	—	+	—	—	+	—
LIGHTNESS	+	—	+	+	—	+	—	—
Notation: H/S/L	 40 / 240 / 120	 40 / 120 / 180	 40 / 240 / 120	 40 / 240 / 120	 40 / 120 / 180	 40 / 240 / 120	 40 / 120 / 180	 40 / 120 / 180
	 40 / 240 / 120	 40 / 120 / 120	 40 / 180 / 120	 120 / 240 / 120	 40 / 60 / 120	 120 / 180 / 120	 120 / 120 / 120	 120 / 60 / 120
	 40 / 240 / 120	 40 / 120 / 60	 40 / 120 / 120	 200 / 240 / 120	 40 / 0 / 60	 200 / 120 / 120	 200 / 120 / 60	 200 / 0 / 60
	COLORS IN POINT	COLORS IN LINE			COLORS IN SURFACE			COLORS IN VOLUME

Color Combination, Fig. 2 One example of Jannello’s logical scheme providing rules for selecting harmonious colors to be used in a color combination

A Theory for Colors in Combination

Anders Hård and Lars Sivik [1] have settled the basis for a theory of colors in combination. They have developed a structure that considers three dimensions or factors that are useful to describe or analyze color combinations: (1) *color interval* (dealing with color discrimination and having distinctness of border, interval kind, and interval size as subvariables), (2) *color chord* (dealing with color identification and having complexity, chord category, and chord type as subvariables), (3) *color tuning* (dealing with how color combinations can be varied and having surface relations, color relations, and order rhythm as subvariables). Figure 3 shows an outline of this model, published by Hård in 1997 [2].

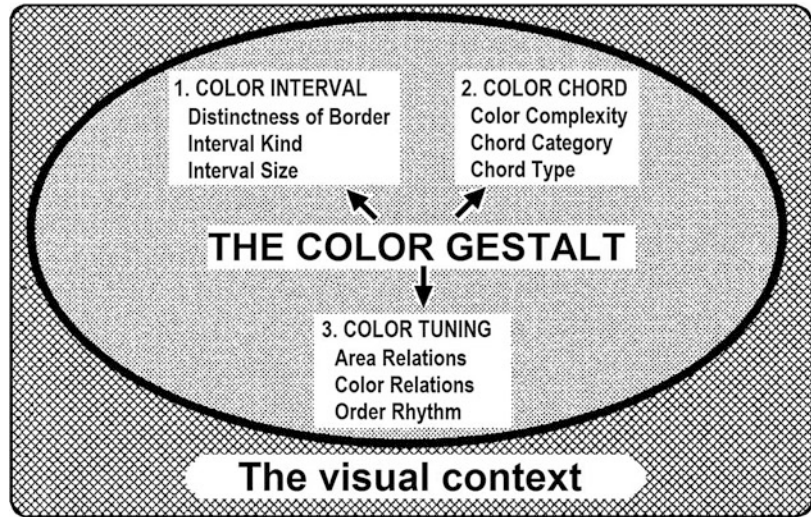
This model for color combination was worked out along various years, and during its development their concepts, dimensions, and subvariables were changing to some extent. Previous formulations of this theory were published by Hård and Sivik in 1985 and 1994 [3, 4]. In some of these, for instance, the visual context in which the color combination appears is considered as a fourth and very important factor.

Semantic and Pragmatic Aspects in Color Combinations

It has been said at the beginning of this entry that the combination of colors is mainly a syntactical aspect. But since colors have a semantic weight, produce emotions, have meanings, are used as signals, indicate situations, propose behaviors, communicate messages, etc., and all this can vary according to the way in which colors are combined and according to the context, it is also possible to consider color combinations from a semantic point of view.

Here is a simple example. The three colors of the traffic lights (red, yellow, green) are a spatial and temporal syntactical combination, on one side. They make a triad of separated color lights, displayed in a circular shape, that appear with a spatially codified vertical arrangement: red at top, yellow in the middle, and green below. The temporal sequence is also regulated and codified: yellow comes after green and red comes after yellow. The step from red back to green is normally direct, without intermediation of yellow, and the same sequence is repeated again: green, yellow, and red; green, yellow, and red; and so

Color Combination,
Fig. 3 Outline of a color
 combination model by
 Anders Hård [2]



on. The size of each color has also a specific relationship: the red circular light has a larger size than yellow and green. And the same happens with the time interval or duration in which each light is displayed: yellow appears for a brief instant, while red and green have longer durations. Now on, all these are purely syntactic aspects. Nothing has been said yet about what meaning this selection, arrangement, and sequence of combined colors conveys. By entering in the semantic domain, it is possible to talk about the codification of those three colors in that particular context of use: red means “stop,” and green means “go,” while yellow is a warning about the change of light from green to red that is coming soon and imply that the user has to take the necessary caution, either apply the brakes to stop the vehicle or speed up the march to make it through before the red light appears.

Thus, there are also semantic issues that are combined from a syntactical arrangement of colors and the context in which they are used and interpreted. The same red color used for the traffic lights may have very different meanings in other contexts: it may connote emotions such as passion, love, and rage; it may indicate something that is important to notice and deserves to stand out (a red typo in a context of black letters and words); it means expulsion from the field in the context of a football (soccer) match; it may also

connote speed or status in a car (a red Ferrari), etc. Hence, it is the context (either the social, cultural, geographic, or temporal context in which the colors appear, as well as the relationship with other colors that are in the same context or situation) what endorses colors and color combinations with a certain sense or meaning.

Color combinations have been studied from the semantic point of view by various authors. Elda Cerrato points out the basic concepts behind the idea of color combination, discusses some color order systems and color harmony theories related to this (mainly Ostwald, Munsell, and Arnheim), and addresses the issue of how culture conditions harmonies, preferences, and aesthetical principles of color combinations [5].

Shigenobu Kobayashi, working at the Nippon Color and Design Research Institute, has devised a method to classify single colors or three-color combinations by their associated images [6, 7]. Through the analysis using the axes warm/cool, soft/hard, and clear/grayish as coordinates, this method, which was also extended and developed with some collaborators [8, 9], can plot climatic and cultural differences in color semantics. In 1997, Kobayashi and Iwamatsu extended the color combination research to be able to include five-color combinations [10], even when from the countless number of possible five-color combinations they chose to make the survey by

selecting 20 pairs of contrasting combinations of five colors (40 color combinations in total). To justify that particular research, they point out that “a five-color combination makes it easier to gain a psychological understanding of scenic conditions and convey the sense of ambience than two- and three-color combinations.”

The book by Hideaki Chijjiwa, from the Musashino College of Art, at Musashino Art University, intends to be a manual for choosing color combinations for different purposes, taking into account meanings and mood [11]. For practical applications in art, design, industry, or everyday life, it provides a guide for selecting two-, three-, and four-color combinations associated to adjectives such as striking, tranquil, exciting, natural, warm, cold, young, feminine, and surprising. As for the quantity of colors to be employed in a combination, the author advises to limit them to two or three. A warning is made when using four-color combinations that should be selected very carefully, while five-color combinations are directly discouraged. The book by Bride M. Whelan continues in the same venue [12].

Lars Sivik (working sometimes in collaboration with Anders Hård and Charles Taft) carried out research on the meanings of color combinations [13–15]. The descriptive model uses the Natural Color System as a basis, and the methods are aimed at studying the stability and variability of color-meaning associations across time and cultures. These studies “literally mapped the world of color with respect to how associations to various words systematically vary across different parts of the color world.” In the research published in 1989, Sivik selected 130 words by a semantic differential scaling method, and the subjects judged color images as “to how well the different word went with the color composition in question” [15]. The main purpose was “to obtain a small number of variables that would be reasonably representative of all color describing variables.”

Other authors have also used the semantic differential method to study the meanings of color combinations, and more specifically their affective values, by applying this tool to two- and three-color harmonies [16, 17].

Colors, in general, and color combinations, in particular, can be considered as a system of signs; they certainly have syntactic aspects (which include both the elements and the combinatory), semantic aspects, and also pragmatic aspects that imply the use of these signs by the interpreters. In a previous publication, the author of this entry describes and illustrates various semiotic concepts with examples taken from color theory and provides an account of some of the advances of color theory within the framework of semiotic categories [18].

Some color theorists go even beyond these considerations, proposing that colors and color combinations can be taken as a language, for instance, Luckiesh in 1918 [19], Sanz in 1985 and 2009 [20, 21], Oberascher in 1993 [22], as well as Hård in the already mentioned article published in 1997 [2]. However, this should be taken perhaps with certain caution. Human languages (for instance, verbal languages, English, Spanish, French, German, etc.) serve not only for communicational purposes but also for cognitive and modeling purposes; they allow to build categories, models, and theories about the world, in order to understand it, explain it, and make it meaningful for the human species. If it is possible to demonstrate that color combinations can have a similar status, then the color-language idea will be more than just a metaphor.

Cross-References

- [Color Contrast](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Color Scheme](#)
- [Complementary Colors](#)
- [Palette](#)

References

1. Hård, A., Sivik, L.: A theory of colors in combination – A descriptive model related to the NCS color-order system. *Color Res. Appl.* **26**(1), 4–28 (2001)

2. Hård, A.: Colour as a language. Thoughts about communicating characteristics of colours. In: Sivik, L. (ed.) *Colour and Psychology. From AIC Interim Meeting 96 in Gothenburg*, pp. 6–12. Scandinavian Colour Institute, Stockholm (1997)
3. Hård, A.: A colour combination theory for a human environment. In: *Mondial Couleur 85, Proceedings of the 5th Congress of the AIC. Centre Français de la Couleur, Paris* (1985)
4. Sivik, L., Hård, A.: Some reflections on studying colour combinations. *Color Res. Appl.* **19**(4), 286–295 (1994)
5. Cerrato, E.: Cultura y combinatorias de color: cómo la cultura condiciona armonías, preferencias, recomendaciones, leyes estéticas en las combinatorias del color. In: Caivano, J., Amuchástegui, R., López, M. (comp.) *Argen Color 2002, Actas del 6° Congreso Argentino del Color*, pp. 27–34. Grupo Argentino del Color, Buenos Aires (2004)
6. Kobayashi, S. (comp.): *A Book of Colors*. Kodansha International, Tokyo (1987)
7. Kobayashi, S.: *Color Image Scale*. Kodansha International, Tokyo (1991)
8. Kobayashi, S., Sato, K.: The theory of the color image scale and its application. In: Billmeyer, F., Wyszecki, G. (eds.) *AIC Color 77, Proceedings of the 3rd Congress*, pp. 382–383. Adam Hilger, Bristol (1978)
9. Kobayashi, S., Suzuki, H., Horiguchi, S., Iwamatsu, K.: Classifying 3-color combinations by their associated images on the warm/cool and clear/grayish axes. In: Nemcsics, A., Schanda, J. (eds.) *AIC Colour 93, Proceedings of the 7th Congress*, vol. C, pp. 32–36. Hungarian National Color Committee, Budapest (1993)
10. Kobayashi, S., Iwamatsu, K.: Development of six methods of color psychological study. In: *AIC Color 97, Proceedings of the 8th Congress*, pp. 727–730. The Color Science Association of Japan, Kyoto (1997)
11. Chijiwa, H.: *Color Harmony. A Guide to Creative Color Combinations*. Rockport Publishers, Rockport (1987)
12. Whelan, B.M.: *Color Harmony 2. A Guide to Creative Color Combinations*. Rockport Publishers, Rockport (1994)
13. Sivik, L.: Evaluation of colour combinations. In: *Mondial Couleur 85, Proceedings of the 5th Congress of the AIC. Centre Français de la Couleur, Paris* (1985)
14. Sivik, L.: Dimensions of meaning associated to color combination. In: *AIC Symposium 1988, Colour in Environmental Design*, pp. 10.1–10.5. Winterthur Polytechnic, Winterthur (1988)
15. Sivik, L.: Research on the meanings of color combinations. In: *AIC Color 89, Proceedings of the 6th Congress*, vol. II, pp. 130–133. Grupo Argentino del Color, Buenos Aires (1989)
16. Kansaku, J.: The analytic study of affective values of color combinations: a study of color pairs. *Jap. J. Psychol.* **34**, 11–12 (1963)
17. Nayatani, Y.: An analysis of affective values on three-color harmony by the semantic differential method. In: Richter, M. (ed.) *AIC Color 69, Proceedings of the 1st Congress*, pp. 1073–1081. Muster-Schmidt, Göttingen (1970)
18. Caivano, J.: Color and semiotics: a two-way street. *Color Res. Appl.* **23**(6), 390–401 (1998)
19. Luckiesh, M.: *The Language of Color*. Dodd, Mead and Company, New York (1918)
20. Sanz, J.C.: *El lenguaje del color*. Hermann Blume, Madrid (1985)
21. Sanz, J.C.: *Lenguaje del color: sinestesia cromática en poesía y arte visual*. H. Blume, Madrid (2009)
22. Oberascher, L.: The language of colour. In: Nemcsics, A., Schanda, J. (eds.) *AIC Colour 93, Proceedings of the 7th Congress*, vol. A, pp. 137–140. Hungarian National Color Committee, Budapest (1993)

Color Constancy

Bei Xiao

Department of Computer Science, American University, Washington, DC, USA

Synonyms

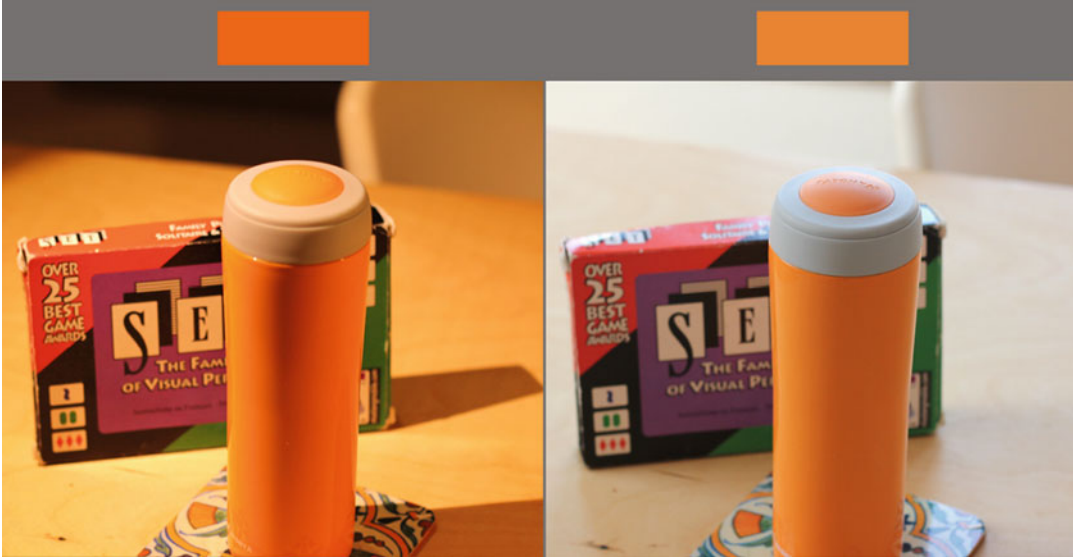
[Chromatic adaptation](#); [White balancing](#)

Definition

Color constancy refers to the ability of the human visual system to perceive stable object color, despite significant variation of illumination. Color constancy is also a desired algorithm in machine vision. In image processing, it is widely used in white balancing algorithms. Color constancy has been an active research topic in the past 100 years. For a thorough understanding of this subject, please see the following recent reviews [4, 14, 16, 18, 36, 35, 58, 72].

Overview

Figure 1 shows photos of the same scene under two illumination conditions. The right photo was



Color Constancy, Fig. 1 Images of the same thermal mug lighted under two illuminants (tungsten illumination and window daylight). *Top*: the patches show rectangular regions filled with the color from roughly the same locations of the mug in the two images. *Bottom*: the images from which the patches were extracted. The image on the

left was taken under two tungsten lamps. The image on the *right* was taken under the daylight from the windows. The photographs were taken by the author using a Canon EOS Rebel T2 digital SLR camera with a 50 mm fixed lens and the automatic white balance function of the camera disabled

taken in the morning when the dominant light source is the daylight, while the left photo was taken in the evening when the dominant light source is tungsten light. If one were in the scene, one would perceive the mug under both illumination conditions to have the same orange color, though individual pixels from the same location from the mug in two photos appear different (see the patches above the photos). Humans exhibit very good color constancy under natural viewing conditions (see a recent review [16]). However, constancy can also be poor under certain conditions. Figure 2 illustrated a scenario when color constancy fails for most observers. Color constancy is important in real-world tasks such as object and scene recognition and visual search [67, 81].

Unlike the human visual system, the image captured by a digital camera has the issue that the color of the scene may be shifted by the change in the external illumination, even though the intrinsic spectral property of the object in the scene (e.g., the mug) stays the same. The goal of a color constancy algorithm is to correct the color shift caused by the illumination change and to extract

reliable color features that are invariant to the change in illumination [25, 43]. The method of correcting image color shift caused by changes in the scene illumination in a camera is called white balance.

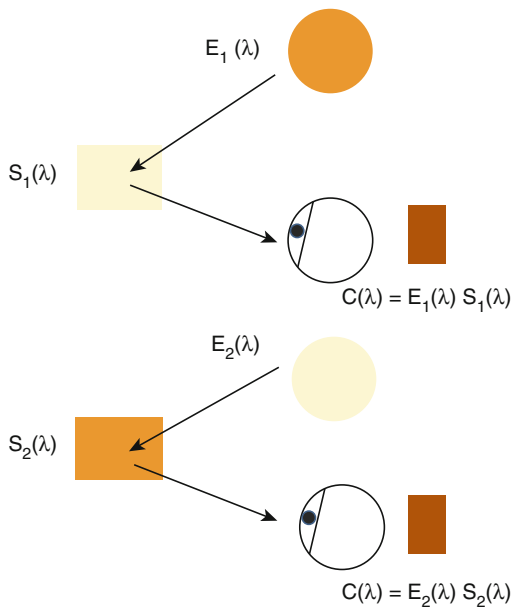
The Problem of Color Constancy

Figure 3 illustrates the problem of color constancy. The color signal reaching the eye, $C(\lambda)$, is a wavelength-by-wavelength product of the spectral power distribution of the illumination $I(\lambda)$ and the surface reflectance function $S(\lambda)$. Different light sources have different spectral power distributions; for example, daylight has different spectral power distribution from that of a tungsten light source. Surface reflectance function $S(\lambda)$ is an intrinsic property of a surface, and it is determined by how the surface absorbs and reflects light. Under a neutral light source, objects with different surface reflectance functions appear to have different colors. The goal of color constancy is to extract the intrinsic surface reflectance



Color Constancy, Fig. 2 Failures of color constancy. *Left:* a photograph of fruits taken under a monochromatic low-pressure sodium light. *Right:* the same scene was illuminated by normal broadband light sources. Most observers

won't be able to tell the color of the bell pepper from the *left* image (the images were downloaded from http://www.soxlamps.com/advantages_sub.htm). But such monochromatic light source is rare in the real world



Color Constancy, Fig. 3 Illustration of color constancy. The spectrum of reflected light reaching the eye, $C(\lambda)$, is wavelength-by-wavelength product of surface reflectance S and the illumination E . The problem of color constancy is challenging because different combination of illumination and surface reflectance can result in the same color signal (This illustration is adapted from David Brainard [16])

function from the color signal. Color constancy is an ambiguous problem because different combinations of illuminant and surface can give rise to the same color signal (see Fig. 3). Many

computational models in the past suggest that the visual system first makes an estimate of the illuminant and uses it to recover the surface reflectance function {Brainard [15, 40, 43, 68]}.

What Do We Know About Human Color Constancy?

How Human Color Constancy Is Measured?

After establishing the problem, now the question is how good the human visual system is at color constancy. To answer this, we need to measure color constancy in a controlled way. Three common methods have been used in the past to measure color constancy in a laboratory setting: color naming where observers name colors of surfaces under different illuminations [46, 47], asymmetric matching where observers adjust a match surface under one illuminant to match the color appearance of a reference surface under another illuminant [2, 21, 80], and achromatic adjustment where observers adjust the chromaticity of a test surface so that it appears achromatic and then repeat the task when the test is embedded in scenes with different illuminants [13]. It was found that asymmetric matching and achromatic adjustment reach similar conclusions of constancy when the two tasks were compared using the same scenes [21].

How Good Is Human Color Constancy?

Overall, these methods show that human constancy is not perfect but generally very good. We can compute a color constancy index from either asymmetric matching or achromatic adjustment experiments, where 0 % means no constancy and 100 % means perfect constancy [2, 21, 20, 73]. Most studies in color constancy use simplified laboratory stimuli that consist of flat and matte surfaces under diffuse lighting conditions (for reviews, see [2, 13, 14, 16, 56]).

For such flat-matte-diffuse surfaces in simple scenes, especially when only the illuminant is varied in the scene, color constancy can be very good with an average constancy index around 0.85 for real scenes [13] and around 0.75 for graphical simulated scenes [28]. However, constancy can be decreased significantly when both the illuminants and the surfaces are varied in a scene. Figure 4b shows an example of such manipulations. The surface in the scene was manipulated so that the color of the background wall reaching the eye under illuminant A is the same as when a neutral colored wall is illuminated under illuminant B (the middle and the leftmost images). In this condition, constancy index drops to 0.2 [80]. Similar effects have also been explored in previous studies [52, 53]. In these cases, constancy is reduced but not completely diminished.

Natural scenes are rarely composed of flat-matte-diffuse surfaces. First, objects usually have 3D shapes and are made of non-matte material such as metal, plastic, or wax, which looks glossy and translucent. Second, the objects in the three-dimensional space are arranged at different depths from the viewing point. Lastly, the lighting condition often has complex spatial and spectral distribution. How good is human color constancy in natural scenes? The factors in natural scenes that have been studied include the effect of three-dimensional pose of surfaces ([8–11], Boyaci et al. 2003, [29, 30, 44, 65, 66]), the effect of lighting geometry on constancy [1, 60–62], the stereo depth on constancy [79], the 3D shape, and the material that an object is made of on color constancy [33, 59, 61, 80].

Cognitive factors have also been considered in studying color constancy [63, 71]. Some objects

have characteristic color, such as bananas are yellow and cucumbers are green. How do we take the prior knowledge of objects' color into account in achieving color constancy? A recent work by Kanematsu [51] suggests the effect of familiar contextual object on color constancy is small.

Figure 4 depicts several scenes used previously in experiments on color constancy. The associated stimuli effectively challenge existing theories of color constancy.

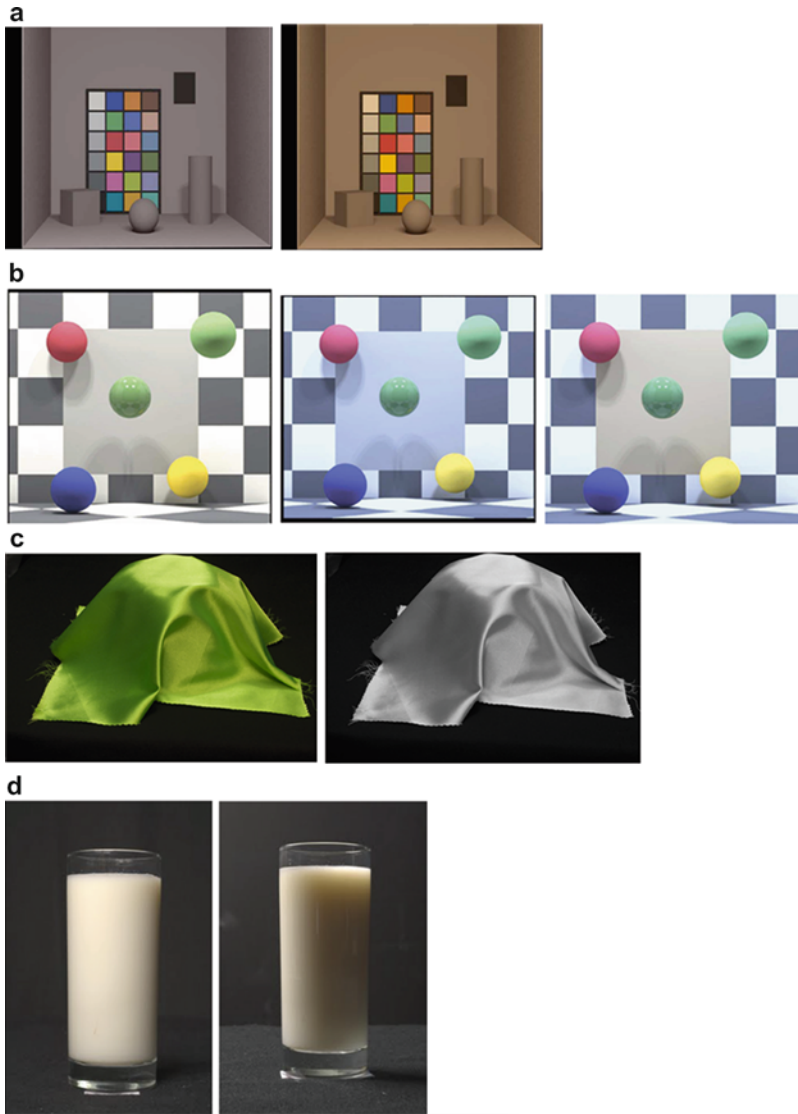
Theories of Color Constancy

How does the visual system achieve color constancy? One approach to understand constancy is to explain it using low-level visual mechanisms such as chromatic adaptation. The color signal reaching the eye, $C(\lambda)$, is encoded by the responses of three classes of light-sensitive photoreceptors in the retina, which are referred to as long (L)-, medium (M)-, and short (S)-wavelength-sensitive cones [17, 50]. Let us represent the spectral properties of the reflected light reaching the eye by the quantal absorption rates for the three classes of cones. The light signal, r , can be represented by a three-dimensional column vector (Eq. 1):

$$r = \begin{bmatrix} r_L \\ r_M \\ r_S \end{bmatrix} \quad (1)$$

The cone signals are subjected by adaptation. Von Kries proposed that the LMS cone signals are scaled by a multiplicative factor, and at each retinal location the gains are scaled independently [76]. For each cone class, the gain is set in inverse proportion to the spatial mean of the signals from the cones of the same class. This algorithm is called von Kries adaptation. The adapted cone signals, a , can be obtained by multiplying the cone signals r by a diagonal matrix D , where the elements g_L , g_M , and g_S represent the three gains:

$$a = \begin{bmatrix} a_L \\ a_M \\ a_S \end{bmatrix} = \begin{bmatrix} g_L & 0 & 0 \\ 0 & g_M & 0 \\ 0 & 0 & g_S \end{bmatrix} r = Dr \quad (2)$$



Color Constancy, Fig. 4 Various stimuli used in color constancy and color perception experiments. (a) Stimuli used to study color constancy by [28]. Synthetic images contain a flat test surface embedded in a relative complex scene (b). The rendered images were used in a study of color constancy of 3D object by Xiao and Brainard [80]. The leftmost scene was illuminated by a neutral illuminant, the middle scene was illuminated by a bluish illuminant, the rightmost scene was illuminated by the same bluish illuminant as the middle scene, but the reflectance of the background surface has been changed so that the light reflected from the background is the same as in the leftmost image. (c) Photograph of a piece of fabric draped over an object. The left image shows the original photograph of the fabric. There is a significant variation of color

across fabric's surface. Such color variation is important for material perception. The right image shows the same photograph in gray scale. A recent study on tactile and visual matching of fabric properties shows that observers make more mistakes predicting fabric's tactile properties with grayscale images than color images (Photos taken by the author). (d) Translucent objects, such as liquid, stone, skin, and wax, represent new challenges for color constancy research. The photo on the left shows a glass of fat-free milk illuminated from the front, and the photo on the right shows the same glass of fat-free milk illuminated from the back. One can observe that the color appears to be slightly different when the illumination direction is varied (The photos were taken by Ioannis Gkioulekas from Harvard University [45])

The von Kries adaptation inspired Land's famous retinex theory [54], which has very wide application in camera color balance and can be used to explain human color constancy for the flat-matte-diffuse stimuli. The central principle of the retinex theory is that the lightness values at each pixel are calculated independently for each cone class. For an analysis of the retinex theory and color constancy, see a study by Brainard and Wandell [19].

Further along the visual pathway, in addition to cone adaptation, secondary adaptation is also proposed. The effects include gain control after the combination of the cone signals and also subtractive modulation instead of multiplicative modulation [48, 49, 64, 69, 77, 78].

The adaptation models predict color constancy quite well for flat-matte-diffuse scenes. However, it often fails to predict constancy for rich scenes [52]. We do now know how to obtain the values of the multiplicative gain from images, which contain spatially rich information. That being said, the mechanistic approach sometimes can inspire new algorithms for color constancy [39].

Another approach is to use a computational method developed from a computer vision perspective. As described above, color constancy is an ill-posed problem. Bayesian methods combine the information contained in the observed scene with information given a priori about the likely physical configuration of the world [15]. In the case of color constancy, some prior knowledge about the illuminants and surface reflectance can resolve ambiguity. Earlier work has used statistical constraints of illuminants and surface reflectance on solving color constancy such as the gray-world, subspace, and physical realizability algorithms [23, 27, 34, 57]. Brainard and Freeman [15] constructed prior distribution describing the probability of illuminants and surface in the world and then estimate the illuminant from the posterior distribution conditioned on the image intensity data. Brainard et al. [22] applied the similar Bayesian model to predicate the degree of human color constancy across different manipulations and connect the variation in constancy to the prior distribution of the illuminant.

What Do We Know About Machine Color Constancy?

While human visual system is equipped with good color constancy, the digital camera has to rely on color balancing algorithm to discount the illumination effect and extract the invariable object color. This process is also called color balance or white balance. The most popular method is based on adaptation such as the von Kries coefficient rule and Land's retinex theory discussed above [31, 54, 76]. But this of this type of model is restricted to simple scenes.

In some sense, there is a significant overlap between the development of algorithms for machine color constancy and the modeling of human color constancy. However, one distinction is the choice of stimuli. To understand human color constancy, simple synthetic stimuli that allow systematic manipulation of scene parameters are often used as experimental stimuli. A successful machine color constancy algorithm, on the other hand, should aim at correcting illumination effects for real-world complex images (see a recent review by [43]).

The best-known statistical method is the gray-world theory, which assumes that the average reflectance of a scene is gray [23]. Some other similar algorithms include white patch and max-RGB [37, 38, 54] and shade of gray [32]. The gray-world algorithm will fail if the average reflectance is not achromatic or if there is a large uniform colored surface in the scene. The incorporation of higher-order statistics in terms of image derivatives is proposed, where a framework called gray edge is presented [74]. Chakrabarti et al. [24] go beyond statistics of per-pixel colors and model the spatial dependencies between pixels by decomposing the input images into spatial sub-bands and then model the color statistics separately in each sub-band.

Forsyth [34] introduced the gamut-mapping method. It is based on the assumption that only a limited set of colors (canonical gamut) can occur under a given illuminant. The model learns a model based on training images (the canonical gamut) and estimate the illuminant based on the input features.

Inspired by Brainard [15], Rosenberg et al. [68] introduced a Bayesian model of color constancy utilizing a non-Gaussian probabilistic model of the image formation process and demonstrated that it can outperform the gamut-mapping algorithm. Gehler et al. [40] extended the Bayesian algorithm using new datasets, which allows the algorithm to learn more precise priors of the illuminations.

A new thread of algorithms estimates the illuminant using high-level features [6, 41, 55, 75]. For example, Gijsenij and Gervse [42] proposed to dynamically determine which color constancy algorithm to be used for a specific image based on the scene category. Bianco and Schettini [5] proposed a model to estimate illuminant from color statistics extracted from faces automatically detected in the image. It takes advantage that the skin color tends to cluster in the color space, which provides a valid cue to estimate the illuminant. Inspired by human visual system mechanisms, a recent study by Gao et al. [39] built a physiologically based color constancy model that imitated the double-opponent mechanism of the visual system. The illuminant is estimated by searching for the maxima of the separate RGB channels of the responses of double-opponent cells in the RGB space.

As the technology of camera sensor develops, several datasets captured by high-quality SLR camera in RAW format have been used in the community. The popular datasets include Ciurea's dataset [26], Middlebury color database [24], and Barnard's datasets [3, 70].

Future Directions

Color constancy is an important and practical problem in both human and computer vision. Color constancy provides an excellent model system to understand how the visual system solves ambiguity. Significant progress has been made to understand color constancy of simple scenes. A big challenge is how to extend the theoretical model of color constancy for simple scenes to predict human performance in rich scenes.

Even though many computer vision algorithms are successful at correcting color bias caused by illumination, whether or not the human visual system uses similar algorithms to achieve constancy is poorly understood. A major challenge to reconcile the latter with the former is a matter of scene and stimulus complexity. Often, images used in computer vision algorithms are real photos or videos, which contain rich scenes. To model color constancy in humans, however, experiments require more simplified scenes and conditions. How to bridge the two fields is a promising endeavor for future research.

Cross-References

- [Color Scene Statistics, Chromatic Scene Statistics](#)
- [Environmental Influences on Color Vision](#)

References

1. Adelson, E.H.: Lightness perception and lightness illusions. In: *The New Cognitive Neurosciences*, p. 339. MIT Press, Cambridge, MA (2000). Retrieved from http://www.cs.tau.ac.il/~hezy/Vision_Seminar/Lightness_Perception_and_Lightness_Illusions.htm
2. Arend, L., Reeves, A.: Simultaneous color constancy. *J. Opt. Soc. Am. A.* **3**(10), 1743–1751 (1986). Retrieved from <http://www.opticsinfobase.org/abstract.cfm?id=2483>
3. Barnard, K., Martin, L., Funt, B., Coath, A.: A data set for color research. *Color Res. Appl.* **27**(3), 147–151 (2002). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.4.123&rep=rep1&type=pdf>
4. Bianco, S., Schettini, R.: Computational color constancy. In: *Visual Information Processing (EUVIP)*, 2011 3rd European Workshop on, Paris, 4–6 July 2011, pp. 1–7. IEEE (2011). doi:10.1109/EuVIP.2011.6045557
5. Bianco, S., Schettini, R.: Color constancy using faces. In: *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on Biometrics Compendium, IEEE (2012)
6. Bianco, S., Ciocca, G., Cusano, C., Schettini, R.: Improving color constancy using indoor – outdoor image classification. *IEEE Trans. Image Process.* **17**(12), 2381–2392 (2008). Retrieved from <http://www.ivl.disco.unimib.it/publications/pdf/bianco2008improving-color.pdf>
7. Bloj, M.G., Kersten, D., Hurlbert, A.C.: Perception of three-dimensional shape influences colour perception

- through mutual illumination. *Nature* **402**(6764), 877–879 (1999). Retrieved from <http://www.nature.com/nature/journal/v402/n6764/abs/402877a0.html>
8. Bloj, M., Ripamonti, C., Mitha, K., Hauck, R., Greenwald, S., Brainard, D.H.: An equivalent illuminant model for the effect of surface slant on perceived lightness. *J. Vis.* **4**(9), 735–746 (2004). doi:10.1167/4.9.6
 9. Bloj, M.G., Hurlbert, A.C.: An empirical study of the traditional Mach card effect. *Perception Lond.* **31**(2), 233–246 (2002). Retrieved from <http://www.perceptionjournal.com/perception/fulltext/p31/p01sp.pdf>
 10. Boyaci H, Maloney LT, Hersh S. The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *J Vis.* 2003;3(8):541–553. Epub 2003 Sep 25
 11. Boyaci, H., Doerschner, K., Maloney, L.T.: Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *J. Vis.* **4**(9) (2004). Retrieved from <http://www.journalofvision.org/content/4/9/1.full>
 12. Boyaci, H., Doerschner, K., Snyder, J., Maloney, L.: Surface color perception in three-dimensional scenes. *Vis. Neurosci.* **23**(3/4), 311 (2006). Retrieved from http://www.bilkent.edu.tr/~hboyaci/Vision/Boyaci_Doerschner_Snyder_Maloney_VisNeuro_2006.pdf
 13. Brainard, D.H.: Color constancy in the nearly natural image. 2. Achromatic loci. *J. Opt. Soc. Am. A* **15**(2), 307–325 (1998). Retrieved from <http://www.opticsinfobase.org/viewmedia.cfm?URI=josaa-15-2-307&seq=0>
 14. Brainard, D.H.: Color constancy. In: *The Visual Neurosciences*, vol. 1, pp. 948–961. MIT Press, Cambridge, MA (2004). Retrieved from <http://www.cns.nyu.edu/csh04/Articles/Brainard-02.pdf>
 15. Brainard, D.H., Freeman, W.T.: Bayesian color constancy. *J. Opt. Soc. Am. A* **14**(7), 1393–1411 (1997). Retrieved from <http://www.opticsinfobase.org/viewmedia.cfm?uri=josaa-14-7-1393&seq=0>
 16. Brainard, D.H., Radonjić, A.: Color constancy. In: Werner, J.S., Chalupa, L.M. (eds.) *The New Visual Neuroscience*, pp. 545–556. MIT Press, Cambridge (2013)
 17. Brainard, D.H., Stockman, A.: Colorimetry. (1995). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.3027>
 18. Brainard, D.H., Stockman, A.: Colorimetry. In: Bass, M. (ed.) *OSA Handbook of Optics*. McGraw-Hill, New York (2010)
 19. Brainard, D.H., Wandell, B.A.: Analysis of the retinex theory of color vision. *J. Opt. Soc. Am. A.* **3**(10), 1651–1661 (1986). Retrieved from <http://www.opticsinfobase.org/viewmedia.cfm?URI=josaa-3-10-1651&seq=0>
 20. Brainard, D.H., Wandell, B.A.: A bilinear model of the illuminant's effect on color appearance. In: *Computational Models of Visual Processing*, pp. 171–186. MIT Press, Cambridge, MA (1991)
 21. Brainard, D.H., Brunt, W.A., Speigle, J.M.: Color constancy in the nearly natural image I. Asymmetric matches. *J Opt Soc Am A Opt Image Sci Vis* **14**(9), 2091–2110 (1997). Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9291602
 22. Brainard, D.H., Longère, P., Delahunt, P.B., Freeman, W.T., Kraft, J.M., Xiao, B.: Bayesian model of human color constancy. *J. Vis.* **6**(11) (2006). Retrieved from <http://www.journalofvision.org/content/6/11/10.full>
 23. Buchsbaum, G.: A spatial processor model for object colour perception. *J. Franklin Inst.* **310**(1), 1–26 (1980). Retrieved from <http://www.sciencedirect.com/science/article/pii/0016003280900587>
 24. Chakrabarti, A., Scharstein, D., Zickler, T.: Color datasets. An empirical camera model for Internet color vision. In: *Proceedings of the British Machine Vision Conference (BMVC)* (2009)
 25. Chakrabarti, A., Hirakawa, K., Zickler, T.: Color constancy with spatio-spectral statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(8), 1509–1519 (2012). <http://cilab.knu.ac.kr/seminar/Seminar/2013/20130330/ColorConstancywithSpatio-SpectralStatistics.pdf>
 26. Ciurea, F., Funt, B. A large image database for color constancy research. In: *Proceedings of the Eleventh Color Imaging Conference* (2003)
 27. D'Zmura, M., Iverson, G., Singer, B.: Probabilistic color constancy. In: Luce R.D., D'Zmura M., Hoffman D.D., Iverson G., Romney, K. (eds.) *Geometric Representations of Perceptual Phenomena*, pp. 187–202. Lawrence Erlbaum Associates, Mahwah (1995)
 28. Delahunt, P.B., Brainard, D.H.: Does human color constancy incorporate the statistical regularity of natural daylight? *J. Vis.* **4**(2) (2004). Retrieved from <http://www.journalofvision.org/content/4/2/1.full>
 29. Doerschner, K., Boyaci, H., Maloney, L.T.: Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *J. Vis.* **4**(2) (2004). Retrieved from <http://www.journalofvision.org/content/4/2/3.full>
 30. Epstein, W.: Phenomenal orientation and perceived achromatic color. *J. Psychol.* **52**(1), 51–53 (1961). Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/00223980.1961.9916503>
 31. Finlayson, G.D., Drew, M.S., Funt, B.V.: Color constancy: generalized diagonal transforms suffice. *J. Opt. Soc. Am. A* **11**(11), 3011–3019 (1994). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.872&rep=rep1&type=pdf>
 32. Finlayson, G.D., Hordley, S.D., Hubel, P.M.: Color by correlation: a simple, unifying framework for color constancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1209–1221 (2001). Retrieved from http://th.physik.uni-frankfurt.de/~triesch/courses/275vision/papers/finlayson_et_al_pami_2001.pdf
 33. Fleming, R.W., Bülthoff, H.H.: Low-level image cues in the perception of translucent materials. *ACM Trans. Appl. Percept. (TAP)* **2**(3), 346–382 (2005). Retrieved from <http://dl.acm.org/citation.cfm?id=1077409>

34. Forsyth, D.A.: A novel algorithm for color constancy. *Int. J. Comp. Vis.* **5**(1), 5–35 (1990). Retrieved from <http://link.springer.com/article/10.1007/BF00056770>
35. Foster, D.H.: Does colour constancy exist? *Trends Cogn. Sci.* **7**(10), 439–443 (2003). Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14550490
36. Foster, D.H.: Color constancy. *Vision Res.* **51**(7), 674–700 (2011). Retrieved from <http://www.sciencedirect.com/science/article/pii/S0042698910004402>
37. Funt, B., Shi, L.: The effect of exposure on MaxRGB color constancy. In: *Proceedings of the SPIE, San Jose. Human Vision and Electronic Imaging XV*, vol. 7527 (2010)
38. Funt, B., Shi, L.: The rehabilitation of MaxRGB. In: *Proceedings of the IS&T Eighteenth Color Imaging Conference, San Antonio* (2010)
39. Gao, S., Yang, K., Li, C., Li, Y.: A color constancy model with double-opponency mechanisms. In: *Computer Vision (ICCV), 2013 IEEE International Conference on, Sydney, 1–8 Dec 2013*, pp. 929–936. IEEE (2013). doi:10.1109/ICCV.2013.119
40. Gehler, P. V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited (2008)
41. Gijsenij, A., Gevers, T.: Color constancy using natural image statistics. *IEEE Trans. Patt. Anal. Mach. Intell.* **33**(4), 687–698 (2007). doi:10.1109/TPAMI.2010.93
42. Gijsenij, A., Gevers, T.: Color constancy using natural image statistics and scene semantics. *IEE Trans. Pattern Anal. Mach. Intell.* **33**(4), 687–698 (2011). Retrieved from <http://staff.science.uva.nl/~gevers/pub/GeversPAMI11.pdf>
43. Gijsenij, A., Gevers, T., van de Weijer, J.: Computational color constancy: survey and experiments. *IEEE Trans. Image Process.* **20**(9), 2475–2489 (2011). doi:10.1109/TIP.2011.2118224
44. Gilchrist, A.L.: When does perceived lightness depend on perceived spatial arrangement? *Percept. Psychophys.* **28**(6), 527–538 (1980). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.211.7842&rep=rep1&type=pdf>
45. Gkioulekas I, Zhao S, Bala K, Zickler T, Levin A. Inverse volume rendering with material dictionaries. *ACM Trans Graphics (TOG)*. 2013;32(6). doi:10.1145/2508363.2508377
46. Helson, H.: Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J Exp. Psychol.* **23**(5), 439 (1938). Retrieved from <http://psycnet.apa.org/journals/xge/23/5/439/>
47. Helson, H., Jeffers, V.B.: Fundamental problems in color vision. II. Hue, lightness, and saturation of selective samples in chromatic illumination. *J. Exp. Psychol.* **26**(1), 1 (1940). Retrieved from <http://psycnet.apa.org/journals/xge/26/1/1/>
48. Hurvich, L.M., Jameson, D.: An opponent-process theory of color vision. *Psychol. Rev.* **64**(6p1), 384 (1957). Retrieved from <http://cogsci.bme.hu/~gkovacs/letoltes/HurvichJameson1957.pdf>
49. Jameson, D., Hurvich, L.: Opponent-response functions related to measured cone photopigments*. *J. Opt. Soc. Am.* **58**(3), 429–430 (1968). Retrieved from http://www.opticsinfobase.org/abstract.cfm?URI=josa-58-3-429_1
50. Kaiser, P.K., Boynton, R.M.: *Human Color Vision* (287). Optical Society of America, Washington, DC (1996). Retrieved from <http://www.getcited.org/pub/100154932>
51. Kanematsu, E., Brainard, D.H.: No measured effect of a familiar contextual object on color constancy. *Color Res. Appl.* **39**(4), 347–359 (2013). Retrieved from http://color.psych.upenn.edu/brainard/papers/Kanematsu_Brainard_13.pdf
52. Kraft, J.M., Brainard, D.H.: Mechanisms of color constancy under nearly natural viewing. *Proc. Natl. Acad. Sci. U. S. A.* **96**(1), 307–312 (1999). Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9874814
53. Kraft, J.M., Maloney, S.I., Brainard, D.H.: Surface-illuminant ambiguity and color constancy: effects of scene complexity and depth cues. *Perception* **31**(2), 247–263 (2002). Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11922136
54. Land, E.H.: The retinex theory of color vision. *Sci. Am.* **237**, 108–128 (1977). Retrieved from http://xayimg.com/kq/groups/18365325/470399326/name/E.Land_-_Retinex_Theory%255B1%255D.pdf
55. Li, B., Xu, D., Lang, C.: Colour constancy based on texture similarity for natural images. *Color. Technol.* **125**(6), 328–333 (2009). Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1478-4408.2009.00214.x/full>
56. Maloney, L.T.: Physics-based approaches to modeling surface color perception. In: *Color Vision: From Genes to Perception*, pp. 387–416. Cambridge University Press, Cambridge, New York (1999). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.211.8602&rep=rep1&type=pdf>
57. Maloney, L.T., Wandell, B.A.: Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A.* **3**(1), 29–33 (1986). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.4745&rep=rep1&type=pdf>
58. Maloney, L. T., Brainard, D. H.: Color and material perception: achievements and challenges. *J. Vis.* **10**(9). doi:10.1167/10.9.19 (2010)
59. Motoyoshi, I., Nishida, S., Sharan, L., Adelson, E.H.: Image statistics and the perception of surface qualities. *Nature* **447**(7141), 206–209 (2007). Retrieved from <http://www.cns.nyu.edu/~msl/courses/2223/Readings/MotoyoshiNishidaSharanAdelson.Nature.2007.pdf>
60. Obein, G., Knoblauch, K., Viéot, F.: Difference scaling of gloss: nonlinearity, binocularity, and constancy.

- J. Vision. **4**(9), 711–20 (2004). Retrieved from <http://www.journalofvision.org/content/4/9/4.full>
61. Olkkonen, M., Brainard, D.H.: Perceived glossiness and lightness under real-world illumination. *J. Vis.* **10**(9), 5 (2010). doi:10.1167/10.9.5
 62. Olkkonen, M., Brainard, D.H.: Joint effects of illumination geometry and object shape in the perception of surface reflectance. *Iperception* **2**(9), 1014–1034 (2011). doi:10.1068/i0480
 63. Olkkonen, M., Hansen, T., Gegenfurtner, K.R.: Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *J. Vis.* **8**(5) (2008). Retrieved from <http://www.journalofvision.org/journalofvision.org/content/8/5/13.full>
 64. Poirson, A.B., Wandell, B.A.: Appearance of colored patterns: pattern—color separability. *J. Opt. Soc. Am. A.* **10**(12), 2458–2470 (1993). Retrieved from <http://white.stanford.edu/~brian/papers/color/smacth.pdf>
 65. Radonjić, A., Todorović, D., Gilchrist, A.: Adjacency and surroundedness in the depth effect on lightness. *J. Vis.* **10**(9) (2010). Retrieved from <http://171.67.113.220/content/10/9/12.full>
 66. Ripamonti, C., Bloj, M., Hauck, R., Mitha, K., Greenwald, S., Maloney, S.I., Brainard, D.H.: Measurements of the effect of surface slant on perceived lightness. *J. Vis.* **4**(9) (2004). Retrieved from <http://www.journalofvision.org/journalofvision.org/content/4/9/7.full>
 67. Robilotto, R., Zaidi, Q.: Limits of lightness identification for real objects under natural viewing conditions. *J. Vis.* **4**(9) (2004). Retrieved from <http://www.journalofvision.org/content/4/9/9.full>
 68. Rosenberg, C., Ladsariya, A., Minka, T.: Bayesian color constancy with non-Gaussian models. In: *Advances in Neural Information Processing Systems* (2003)
 69. Shevell, S.K.: The dual role of chromatic backgrounds in color perception. *Vision Res.* **18**(12), 1649–1661 (1978). Retrieved from <http://deepblue.lib.umich.edu/bitstream/2027.42/22782/1/0000337.pdf>
 70. Shi, L., Funt, B. V.: Re-processed version of the Gehler color constancy database of 568 images. Simon Fraser University (2010)
 71. Smet, K., Ryckaert, W.R., Pointer, M.R., Deconinck, G., Hanselaer, P.: Colour appearance rating of familiar real objects. *Color Res. Appl.* **36**(3), 192–200 (2011). Retrieved from http://www.esat.kuleuven.be/electa/publications/fulltexts/pub_2070.pdf
 72. Smithson, H.E.: Sensory, computational and cognitive components of human colour constancy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**(1458), 1329–1346 (2005). doi:10.1098/rstb.2005.1633
 73. Troost, J.M., De Weert, C.M.M.: Naming versus matching in color constancy. *Percept. Psychophys.* **50**(6), 591–602 (1991). Retrieved from <http://link.springer.com/article/10.3758/BF03207545>
 74. Van De Weijer, J., Gevers, T., Gijsenij, A.: Edge-based color constancy. *IEEE Trans. Image Process.* **16**(9), 2207–2214 (2007). Retrieved from http://hal.archives-ouvertes.fr/docs/00/54/86/86/PDF/IP07_yandeweijer.pdf
 75. Van De Weijer, J., Schmid, C., Verbeek, J.: Using high-level visual information for color constancy. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE (2007)
 76. von Kries, J.: Chromatic adaptation. *Festschrift der Albrecht-Ludwigs-Universität*, pp. 145–158. (1902)
 77. Walraven, J.: Discounting the background – the missing link in the explanation of chromatic induction. *Vision Res.* **16**(3), 289–295 (1976). Retrieved from <http://www.sciencedirect.com/science/article/pii/0042698976901127>
 78. Webster, M.A., Mollon, J.D.: Colour constancy influenced by contrast adaptation. *Nature* **373**(6516), 694–698 (1995). Retrieved from <http://www.nature.com/nature/journal/v373/n6516/abs/373694a0.html>
 79. Werner, A.: Color constancy improves, when an object moves: high-level motion influences color perception. *J. Vis.* **7**(14) (2007). Retrieved from <http://www.journalofvision.org/content/7/14/19.full>
 80. Xiao, B., Hurst, B., MacIntyre, L., Brainard, D.H.: The color constancy of three-dimensional objects. *J. Vis.* **12**(4), 6 (2012). doi:10.1167/12.4.6
 81. Zaidi, Q., Bostic, M.: Color strategies for object identification. *Vision Res.* **48**(26), 2673–2681 (2008). doi:10.1016/j.visres.2008.06.026

Color Contrast

Lois Swirioff¹, Nilgün Olguntürk² and Gertrud Olsson³

¹The Cooper Union for the Advancement of Science and Art Cooper Square, New York, NY, USA

²Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Ankara, Turkey

³School of Architecture, KTH, Royal Institute of Technology, Stockholm, Sweden

Synonyms

Chromatic contrast

Definition

Color contrast describes the perceptual effects of colors' adjacency in contexts, whether they occur

and are observed in two- or three-dimensional space. It is the relationship between the color of a stimulus and that of its immediate surround.

Overview

The concept of color contrast is concerned with the perception of color itself and is based upon the idea that the eye – the visual system – evaluates, a function of thinking. Light and color are therefore not merely surface phenomena but are intrinsic to human vision, to sight and insight. Light and color contrasts are primary attributes of vision. They signal the reactive eye to initiate and activate the complex visual system to respond to the visual world.

Fundamentally, the perceptions of the visual world, in its enormous diversity – the environment, man made or natural; people; places; animals; and vegetal life-space itself – are distinguished and contextualized by the glorious capacity to *see*. The primary function of the visual system evolved to react to this diversity of features as complex contrasts of shapes and colors. These formal attributes originate and are embodied in the response of the human eye to gradients of light and its absence, shadow. Gibson links the concept of gradient to both formal and spatial effects, as well as other incremental changes [1]. In the context of color contrast, it is possible to use this concept to express increments of light and color. It is possible to think of this as a “quantification

without number.” Color originates in the environment and in the visual system, as a complex and selective response to light.

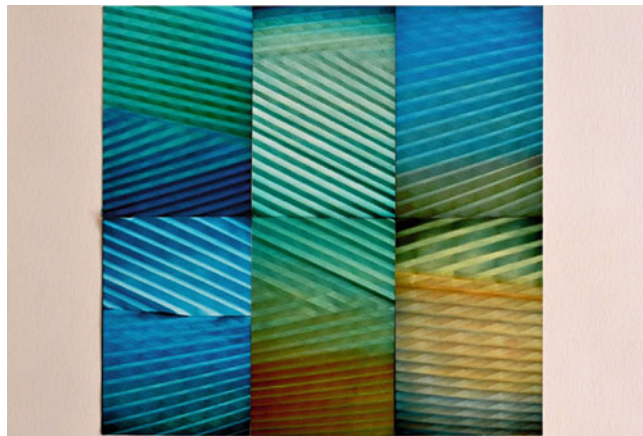
Visual artists, as abstract or figurative painters, are experts in defining the images they create on two-dimensional surfaces by analyzing and expressing them as shapes and areas of light and color. All visual languages require this ability. Figurative painting depends upon the narrative orchestration of these components; for abstract works of art, shapes and colors, with their strong associative attributes, suffice as expressive forms. Now, how does the visual system make sense in the first place of the complexity of the visual world? By responding to its gradients of light and colors as areas of contrast (Fig. 1).

But color, as Josef Albers asserted, is the most changeable component in art. And the painter Delacroix boasted he could turn a mud color into gold. It is all a matter of context. Chevreul, the chemist, in the nineteenth century defined “simultaneous contrast” as the influence of colors in backgrounds to affect the appearance of colors within their boundaries [2]. Thus, for example, a middle gray against a black background looks light, while that same gray against a light background appears dark. Lights and darks are particularly susceptible to change (Figs. 2 and 3).

Josef Albers, in the twentieth century, explored and exhaustively expanded the simultaneous contrast law, as well as the Bezold effect, the Liebmann effect, Fechner’s psychophysical laws, and Goethe’s perceptual concepts, by his and his

Color Contrast,

Fig. 1 Lois Swirnoff:
Santa Fe Rhythm, acrylic on
folded paper



Color Contrast,

Fig. 2 The same gray appears as two

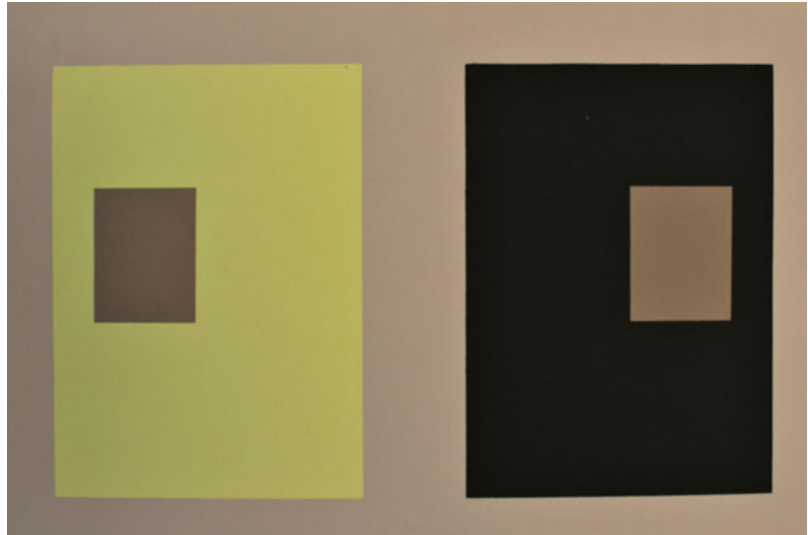
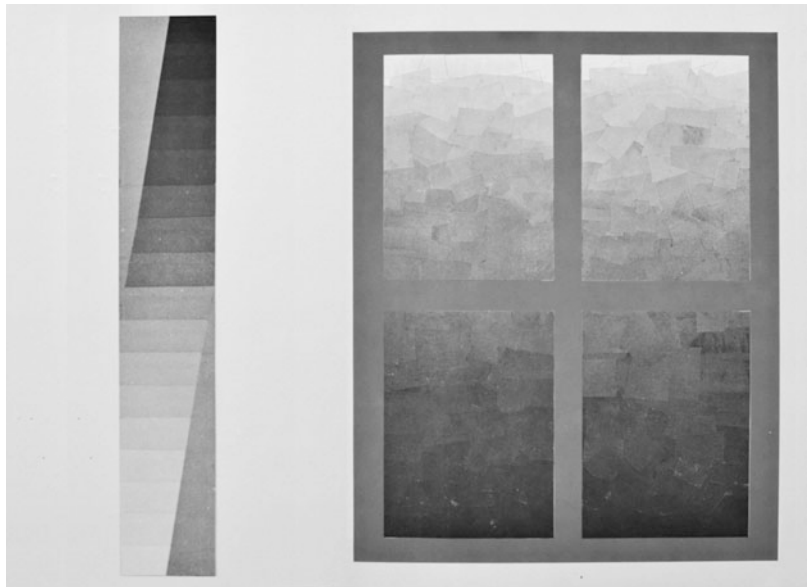
**Color Contrast,**

Fig. 3 Scales of gray



students' experiments and in his great work, *The Interaction of Color* [3]. In it, the empirical studies which entailed figure/ground relationships were explored far beyond Chevreul's. To the extent that contrasts between color complements, like a yellow and a violet, against opposing backgrounds, could appear to be similar (Figs. 4, 5, 6, and 7).

Underlying Albers' discoveries is the function of the visual system – the afterimage. The stimulus of a bright red area, after 30–60 seconds of exposure to the eye, will appear bright green. Just as staring at the bright light of a window in a darkened room appears as a dark window, when the eyes are diverted or shut. Another was the contrast boundary, the thin line at the edge of a

color figure which appears against a contrasting background, initiating the reaction of a color change. Application of gradients of colors as linear planes can cause mixtures between adjacent

color boundaries to appear as an additive third, i.e., a combined effect which regenerates the appearance of light. Changing the quantity of this gradient between two colors in a painting can cause many additive mixtures to be created. The work of Julian Stanczak and Richard Anuskiewicz, former graduate students of Albers, explore these phenomena and created the movement of Op Art in the 1960s in the United States (Fig. 8). The researchers Leo Hurvich and Dorothea Jameson describe this phenomenon in their article “From contrast to assimilation; in art and in the eye” [4] and attribute it to an opponent response system in human vision, beginning at the retina. They are proponents of the theories of Ewald Hering, whose book *Outlines of a Theory of the Light Sense* they translated into English [5].

Swirnoff’s experimental work with color began with a question: What happens to color interactions lifted from the two-dimensional plane when they interact in space? Will contrast boundaries prevail in adjacent colors when they appear as sequences of planes in space, as they do on the 2-D surface? Do the size and/or placement of volumes appear different when their surfaces reflect contrasting hues? How are clues to size and distance influenced by color? What is the influence of reds (long wavelength) to blues (short wavelengths) on the appearance of volumes in space?

Beginning with Gestalt principles of form and organization, size and placement, proximity,



Color Contrast, Fig. 4 Reverse ground: each X is the same color

Color Contrast, Fig. 5 Reverse ground: disc



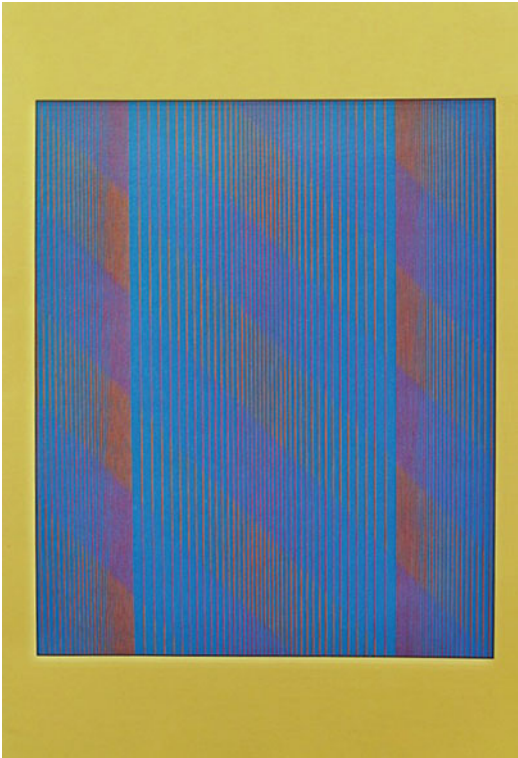
Color Contrast,**Fig. 6** Three colors appear as two**Color Contrast,****Fig. 7** Complementary colors violet and yellow appear equal

clustering, and grouping, Swirnoff applied primary surface colors (Color-Aid) to cubes or floating planes, differing in size and placement in a space frame, and observed their effects through a series of experiments. These findings are the subject of the paper titled “Spatial Aspects of Color,” a thesis written for her Master of Fine Arts at Yale University in 1956.

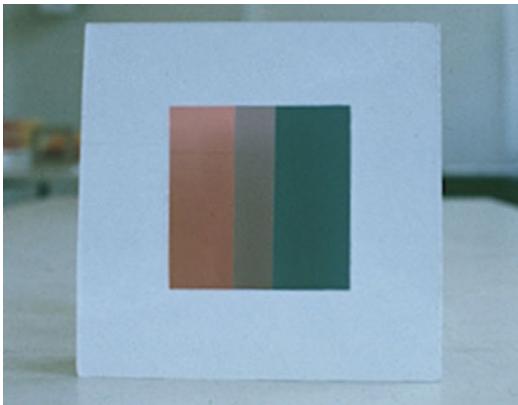
Since then, Swirnoff’s experiments with her students of design, over more than three decades, have explored contrast effects spatially, from tabletop experiments to collaborations of environmental scale, published in her book *Dimensional*

Color [6]. These experiments have resulted in the assertion that color can be defined as a nonlinear dimension.

A new phenomenon was discovered: color stereopsis (Figs. 9, 10, 11, 12, 13, and 14). This occurs when three interrelated colors, separated sequentially in a space/frame, are observed frontally through a square opening. In this study, the contrast boundary between an intermediary color, seen as adjacent to two, one placed behind, the other in front, disappears. With the absence of the middle intermediary, the visual system prefers to see the two color planes, which are frontally



Color Contrast, Fig. 8 Julian Stanczak: *Blue Squeeze*, acrylic on canvas



Color Contrast, Fig. 9 Color stereopsis: three colors fuse as two, *red/green*

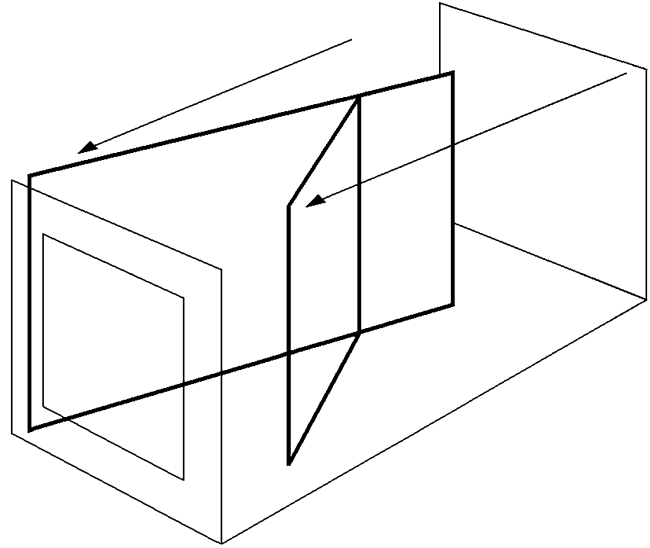
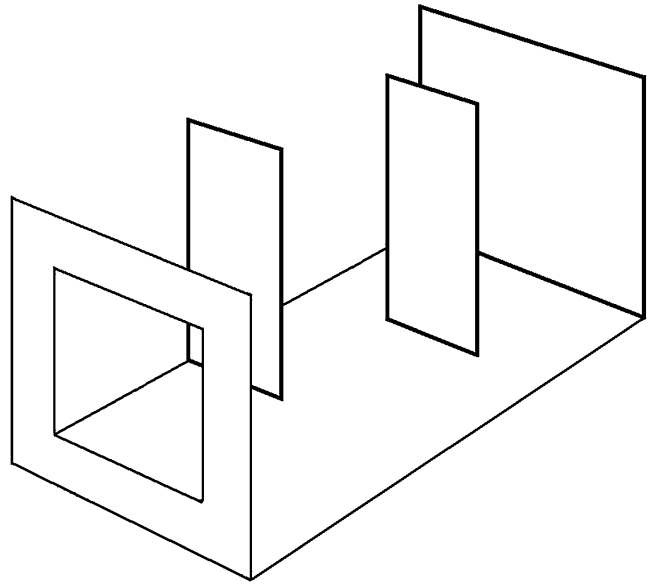
aligned in the space frame, appear to rotate to a diagonal position, and as they appear to intersect, trigger a strong stereoptic effect, analogous to the experience of depth when two flat images of the same subject, seen through a

“stereopticon” – differing slightly in their position – fuse.

In a problem devised to test the degree of light and shadow that defines a square pyramid, models were placed against the walls of the teaching studio to observe the patterns of light and shadow which reflect from the four white triangles of its faces (Figs. 15 and 16). The challenge – to see what degrees of contrast produced the visual effect of a pyramid – entailed placing gray papers ranging from white to black against the model in reversed order, to cancel the pattern of light and dark, causing it to appear flat, without altitude. It was found that the degree of change in the contrast values perceived, with scales of gray papers, between white to black was *geometric* – far greater than anticipated. Once found, the contrasting grays, applied to the four triangles in opposed order to their original appearance, matched. The contrast boundaries disappeared, and the model took the appearance of a flattened surface, a bisected figure or a pyramid of increased altitude, when rotated against the wall or in the hand. Thus, the response of the visual system to contrasts of light and dark depends upon *ratios* rather than measurable reflectances.

With a grant from the IIDA, Swirnoff was able to conduct a room-sized experiment, with students in the lighting studio of Parson’s School of Design in New York. The issue was to test the effect of colored light, projected on interior surfaces. The room’s interior was constructed of four nine-foot squares, three comprised the walls – two lateral ones were connected by 90° to the third back wall – all three joined to the fourth square as the floor, an open cube. Using theatrical gels, colored light was projected onto these surfaces, and it was observed how they affected the spatial appearance of the room (Figs. 17 and 18).

Color generated by light greatly intensifies and enhances the experience of color contrasts. Both ends of the visual spectrum were tried (long wave, reds and yellows, then short wave, blues) sequentially, in separate experiments, and the differences in spatial experience by the contrasts each presented, as anticipated, were obvious. The “red room,” however, provided more phenomena [6].

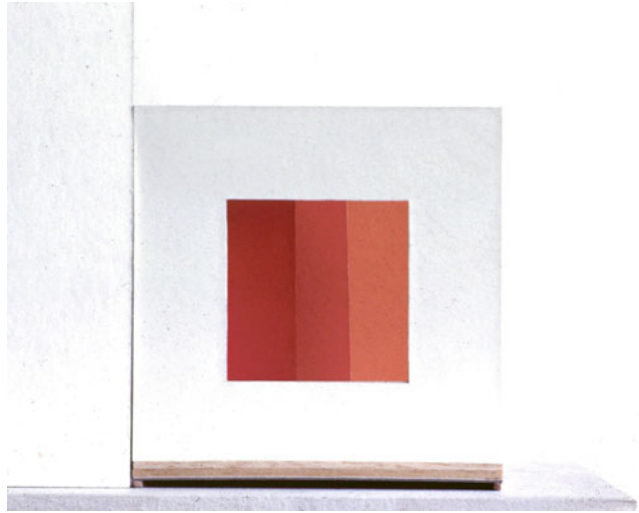
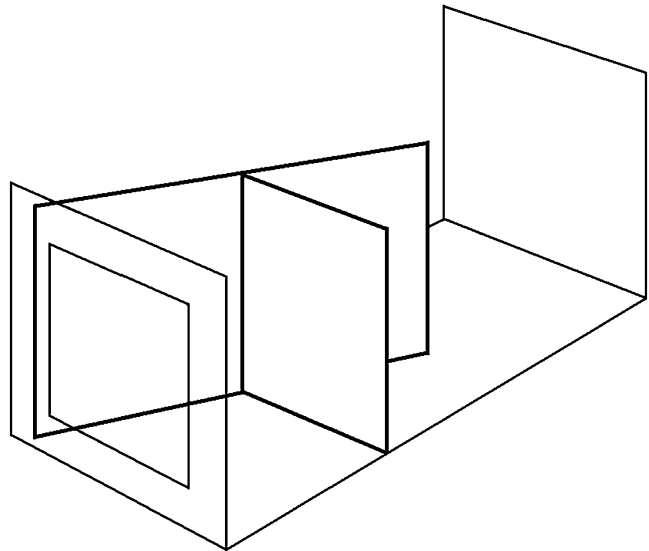
Color Contrast,**Fig. 10** Diagram: their spatial appearance**Color Contrast,****Fig. 11** Diagram: spatial placement of three planes

Against the back wall adjacent to the lateral wall planes, illuminated with red, a neutralized yellow light was projected, to fill its square surface. At the 90° juncture color contrasts intensified; when observed over time, a neutralized yellow light was seen, reflected from the back wall, adjacent to the red lateral wall, change at their mutual boundaries, at first to a green edge and then followed by the green filling in the entire square! While the adjacent red wall increasingly

appeared warmer, and at the contrast boundary, the 90° juncture with the green began to reflect a yellowish light.

These changes were observed as they occurred over a period of about 8–10 min, when surprisingly, the green at its contrast boundary with the yellowing red changed to blue/violet.

There had been a complete change in the colors of the contrast boundaries, beginning originally as a yellow-projected light, which changed the entire

Color Contrast,**Fig. 12** Color stereopsis value sequence of three reds**Color Contrast,****Fig. 13** Diagram: their spatial appearance

plane to green, followed by a contrast boundary appearing gradually as blue/violet, now adjacent to the red wall, whose contrast boundary appeared yellow – a complete reversal of contrast from the original stimuli of color.

Clearly, the intensity of the colored light provided a much greater stimulus to the visual system than that of a surface color. The color of surfaces induced to change by their adjacency on two-dimensional surfaces contrasts mutually

(simultaneous contrast); once perceived, they remain stable. This phenomenon, engendered by the intense light projected in space, seems to have extended to a sequence of perceptual changes which occurred after a much longer time – 8–10 min of observation. The experience all the observers had was *perceptual*. The effect eluded all attempts to record it, by film or digitally. To account for these changes, elicited over a period of time, Swinoff named the effect “sequential contrast.”

Color Contrast,

Fig. 14 Diagram:
alternative appearance as
zig/zag

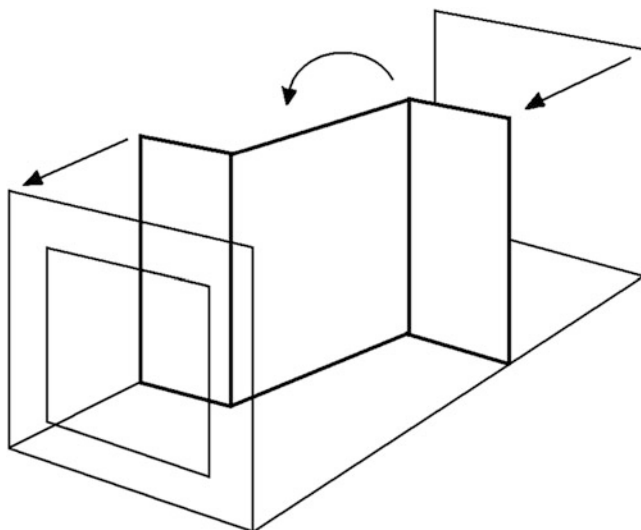
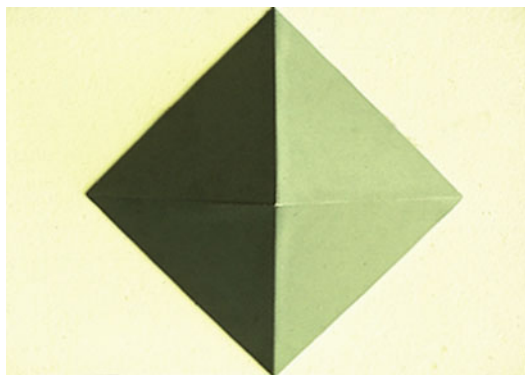
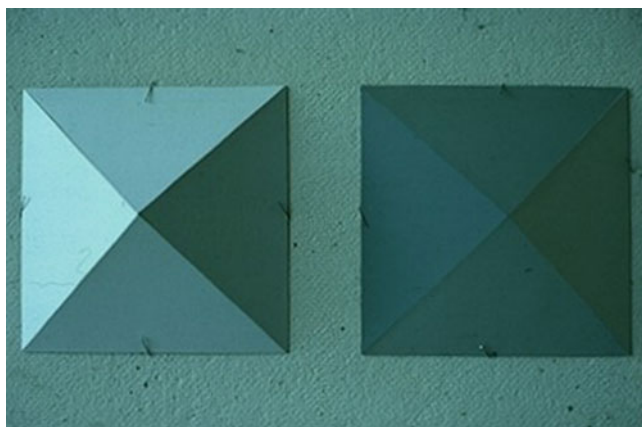
**Color Contrast,**

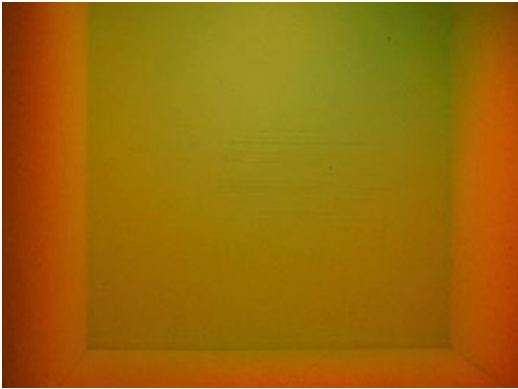
Fig. 15 Pyramids: *left*, the
white model; *right*, model
flattened by grays



Color Contrast, Fig. 16 Pyramid rotated 90° appears
bisected



Color Contrast, Fig. 17 The room illuminated by LED



Color Contrast, Fig. 18 The room's back wall changing to green

To conclude, the evidence raised by these experiments suggests that the issue of color contrast is fundamental to and an essential part of the process of sight. Far beyond the purview of design science to explain, however, they offer a challenge now to neurology and brain science.

Effects in Perception

Contrast detection is the basic visual task from which all other visual behaviors are derived. The human visual system gives virtually no useful information unless there is a contrast in the retina (thus, also in the environment that is being viewed). A small object, or a patch, can only be seen on a larger one if the two differ in color. These differences are known as contrast [7].

Sensation of color, or interpretation of color in the brain, is not only effected by adjacent areas of the stimulus, but also by the light under which the stimulus is seen. Two examples of this are the Helson-Judd effect and the Bezold-Brücke hue shift. The Helson-Judd effect is the tendency of lighter achromatic surfaces to take on the hue of the illuminant under which they are viewed and darker achromatic surfaces to take on the complementary hue. The Bezold-Brücke hue shift is a shift in the apparent color of a stimulus toward yellow or blue with the increasing intensity of light. If a pair of long-wavelength lights differing only in intensity is compared, the higher intensity

stimulus will look more yellow and less red than the lower intensity light. For shorter wavelengths, higher intensity lights look more blue and less green than lower intensity lights [7].

Contrast perception also causes visual effects that lead to variance in color sensation. This may be caused by either the psychophysics of the eye or by the interpretation of the brain. These effects are: successive contrast, simultaneous contrast, edge contrast, and assimilation (reversed contrast).

Successive contrast is the visual effect which occurs when eyes are fixed on a colored patch for a sufficient period of time and then moved on to another patch of a different color. It is likely that the image of the first patch will be perceived upon the image of the second with its afterimage [7]. An afterimage is the visual effect that occurs after light stimulus has been removed. In the case of successive contrast, afterimage complementary color of the initially viewed image will be imposed on the lately viewed image.

Simultaneous contrast is the visual effect which occurs when two different color patches are viewed together, where both will exhibit changes of appearance. Simultaneous contrast is affected by the distance between two colors. If two different color patches of equal size are placed side by side, both will exhibit changes in color appearance. When the patches are separated, the changes decrease and eventually diminish, as the distance between them is increased. Simultaneous contrast effects the hue changes by roughly superimposing the complementary color of the background on the foreground patch. This hue change is accompanied by changes in the perceived saturation depending on the afterimage complementary color produced. For example, if a blue-green patch is viewed against a red background, the patch will be perceived as blue-green again (the afterimage of red being a blue-green), but with an increased saturation [8].

If two areas of uniform colors having the same hue but slightly different luminance factor are viewed, adjacent to the boundary between the two areas, there is a relative enhancement of lightness of the lighter area and a corresponding darkening of the adjoining area [8]. This is called edge

contrast. If a black line is drawn along the edge at which two areas join, the effect of edge contrast is lost. The edge contrast phenomenon is caused psychophysically by interactions among nerve cells in the retina. It is also referred to as the Mach-band effect, the Mach contrast (named after the physicist Ernst Mach), or the border contrast [7].

Assimilation (reversed contrast) is the visual effect which occurs when two distinctly perceived colors seem to shift in appearance toward each other. An example would be strips of colors, namely, red, yellow, and blue. The red strips on the yellow background appear yellowish, and on the blue background, the same red strips appear bluish. Assimilation is also known as the Bezold spreading effect. Assimilation should not be confused with simultaneous contrast. In simultaneous contrast, a red area surrounded by a yellow area would tend to look more bluish (not yellowish); surrounded by a blue area, it would tend to look more yellowish (not bluish) [8]. It is also important to introduce a relation between the viewing distance and the effect that will be perceived. As the viewing distance is increased, producing retinal images of finer detail would be more possible, and a transition from simultaneous contrast to assimilation will occur. Finally, when a distance is reached beyond which the pattern is distinguishable, the visual experience of pointillism takes over where colors put on a surface will be perceived as a spatial mixture (optical mixture) [7].

Sonia Delaunay's Simultaneous Contrast

This concluding section focuses on Sonia Delaunay's and the early modernism's application of simultaneous contrast. For Sonia, and also for Robert Delaunay, and the poets Apollinaire and Cendrars, simultaneous contrast was a tool used to express the new era's speed, velocity, and movement. In Paris, the Russian-French artist Sonia Delaunay (1885–1979) worked with strong contrasts of color in her early expressionist painting period around 1906. A few years later, she started to work with simultaneous contrast in a

non-figurative way in several projects: paintings, fashion, ballet costumes, textiles, prints, books, and interior designs. As early as 1912, she created one of the first painting techniques in an abstract form with colors applied to large areas with the aim of achieving an interaction of color contrast [9]. She picked up the concept of simultaneous contrast from Chevreul and called her style of painting *simultanée*. Principally, she worked with color contrast to add power and chromatic strength to the tints.

Sonia – and her husband, the painter Robert Delaunay – first observed Chevreul's law in nature. In Spain and Portugal, she writes: the diffusion of light is the purest. The light is so strong that the colors themselves become distinct and their hues become robust. No haze or tones of gray interfere and mix the colors; no achromatic grayed effects appear. The quality of this light allowed the artists “to go even further than Chevreul in finding dissonances in colored light” [10]. She explains “dissonances” as “rapid vibrations, which provoked greater color exaltation by the juxtaposition of specific hot and cold colors” [10]. To create the color vibrations, the Delaunays began to divide the shades of colors into hot and cold. This meant working with complementary colors and, to an even greater degree, with *cold* and *warm* color contrasts. Sonia Delaunay states that colors “agitated by hot and cold dissonances provoke a stimulating response” to the viewer [10].

Simultaneous Contrasts in Patterns

Sonia Delaunay's patterns in pure colors and in new color relationships are varied and have abstract forms: arcs and circles, rectangles, and triangles. For her, the circle was a symbol of the sun and also of simultaneous action. Her patterns change in rhythmic movements corresponding to their movements in value (lightness) and hue. Dresses and costumes are also inspired by the natural movement of the body. She began from the four basic colors – red, yellow, green, and blue – and from black and white. The red hue is often mixed with yellow to form an orange-red.

The green is mixed with yellow to form a hue similar to the color of a mimosa flower. The blue is a medium blue, and the yellow is strong. In addition, she uses gray [11].

Around 1912, standard fabrics usually had large flowers against black or strongly colored backgrounds [9]. Sonia Delaunay's simultaneous fabrics changed this custom; she manufactured the same elements as those used in her paintings. Her abstract patterns are simple and have clear motifs in composition and color (quite similar to some African patterns). The colors are often vivid, but they harmonize most agreeably. Characteristic of her textile design is that the forms and patterns may well appear geometric, but the color surfaces are "characterized by *rhythm*" [10]. Due to Sonia Delaunay's simultaneous placement of the colors, they produce new and original effects "right before your eyes" [10]. They are thus responsive to the architecture of modern life, to the new active way of living. In connection with this textile creation, she began to work with the first "simultaneous automobile," a *Citroën B12* (1925), painted in the colors of the rainbow. In this way, she was at the forefront in the showing of art outside of the salons.

Simultaneous Contrast in Costume and in Fashion

During the summer of 1913, Sonia Delaunay began to design simultaneous dresses. She made and mounted them in collages made of textile. These dresses caught a new wave in fashion corresponding with the latest popular dances of the time, foxtrot and tango. In her abstract forms of arcs, circles, rectangles, and triangles, she created a movement of color. But the forms and the contrasting colors also enhanced the natural movement of the forms of the body, matching with the rhythms of Latin music. In the dancehall the Bal Bullier in Montparnasse, one could see the action of the dancers united with the action of color and light. Sonia Delaunay was commissioned to design the costume for the ballet *Cléopâtre* (1918). Working for the theater, she could experiment with successive designs for

lengths of fabric: textiles wrapped around the human form, the body set into action in dance, all visual movements of the costume. Cleopatra's costume was built up of discs in pure colors decorated with sequins and pearls. The ballet established Sonia Delaunay's name as an innovator in both costume and fashion [12]. In 1922, a textile manufacturer in Lyon, France, asked her for a set of fabric designs and promptly ordered 50 designs for silk. For the commission, she began to study color relations and introduced abstract geometrical designs in printed silk. The subject of textile studies refined her control of the interaction of colors [11]. A few years later, on Boulevard Malesherbes in Paris 1925, Sonia Delaunay opened her own shop, the *Boutique Simultanée*. There she offered simultaneous design in the form of coats, dresses, handbags, and even interior furnishings.

Cross-References

- ▶ [Anchoring Theory of Lightness](#)
- ▶ [Appearance](#)
- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [Chromostereopsis](#)
- ▶ [Color Harmony](#)
- ▶ [Color Vision, Opponent Theory](#)
- ▶ [Fechner, Gustav Theodor](#)
- ▶ [Mach Bands](#)
- ▶ [Optical Art](#)
- ▶ [Unique Hues](#)

References

1. Gibson, J.J.: *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston (1966)
2. Chevreul, M.-E.: *The Principles of Harmony and Contrast of Colors* (English translation by Faber Birren). Reinhold, New York (1967)
3. Albers, J.: *The Interaction of Color*. Yale University Press, New Haven (1963)
4. Hurvich, L.M., Jameson, D.: From contrast to assimilation; in art and in the eye. *Leonardo* **8**, 125–131 (1975)
5. Hering, E.: *Outlines of a Theory of the Light Sense* (Translated by Leo M. Hurvich and Dorothea Jameson). Harvard University Press, Cambridge, MA (1964)

6. Swinoff, L.: *Dimensional Color*. Birkhauser, Boston (1989). Second revised edition: W. W. Norton, New York (2003)
7. Camgöz, N.: *Effects of Hue, Saturation, and Brightness on Attention and Preference*. Dissertation. Bilkent University, Ankara (2000). Also available by UMI, Bell & Howell Co., Ann Arbor (2001)
8. Agoston, G.A.: *Color Theory and Its Application in Art and Design*. Springer, Berlin (1987)
9. Damase, J.: *Sonia Delaunay. Fashion and Fabrics*. Thames and Hudson, London (1991)
10. Delaunay, R., Delaunay, S.: *The New Art of Color. The Writings of Robert and Sonia Delaunay*. The Viking Press, New York (1978)
11. Olsson, G.: *The Visible and the Invisible: Color Contrast Phenomena in Space*. Axl Books, Stockholm (2009)
12. Delaunay, S.: *A Retrospective*. Albright-Knox Art Gallery, New York (1980)

Color Coordination

► Color Combination

Color Dictionaries and Corpora

Angela M. Brown

College of Optometry, Department of Optometry,
Ohio State University, Columbus, OH, USA

Definition

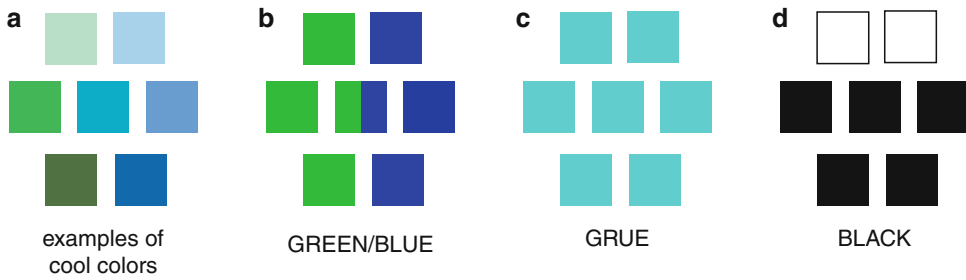
In the study of linguistics, a **corpus** is a data set of naturally occurring language (speech or writing) that can be used to generate or test linguistic hypotheses. The study of color naming worldwide has been carried out using three types of data sets: (1) corpora of empirical color-naming data collected from native speakers of many languages; (2) scholarly data sets where the color terms are obtained from dictionaries, wordlists, and other secondary sources; and (3) philological data sets based on analysis of ancient texts.

History of Color Name Corpora and Scholarly Data Sets

In the middle of the nineteenth century, color-name data sets were primarily from philological analyses of ancient texts [1, 2]. Analyses of living languages soon followed, based on the reports of European missionaries and colonialists [3, 4]. In the twentieth century, influential data sets were elicited directly from native speakers [5], finally culminating in full-fledged empirical corpora of color terms elicited using physical color samples, reported by Paul Kay and his collaborators [6, 7]. Subsequently, scholarly data sets were published based on analyses of secondary sources [8, 9]. These data sets have been used to test specific hypotheses about the causes of variation in color naming across languages.

From the study of corpora and scholarly data sets, it has been known for over 150 years that languages differ in the number of color terms in common use. Particularly, languages differ greatly in how they name the cool colors that are called “blue” and “green” in English (Fig. 1). Some languages, such as English, use a word *BLUE* that means only blue, in conjunction with a word *GREEN* that means only green. Other languages use a single term (here and elsewhere, “*GRUE*”) that means green or blue, and still other languages use a word (here, “*BLACK*”) that means both black and blue, to name the cool colors, in conjunction with *WHITE*, which names the light and warm colors.

Scholars in the nineteenth century established the two general explanations for this diversity of color terms across languages, which still guide much of the research on the topic today. The first explanation was that the people who spoke languages with few color terms had deficient color vision. This speculation was at first based on the philological analysis of extinct languages and arose in part because of general interest in the theory of evolution in the latter half of the nineteenth century. Proponents of this view speculated that humans and their color vision had evolved since ancient times. The second explanation was that people living at different times and in different cultures need to differentiate between different



Color Dictionaries and Corpora, Fig. 1 (a) Examples of cool colors; (b–d) false color coding of the corresponding color terms. The center sample of the

diagram is called *BLUE* by some informants and *GREEN* by others in (b), but it is called *GRUE* by all informants in (c), and it is called *BLACK* by all informants in (d)

colors, so their languages have different numbers of color terms. Particularly, ancient languages lived in simpler times and consequently had fewer color terms in their lexicons.

The Color Deficiency Explanation

The earliest scholar to study this variability across languages was **William Ewart Gladstone**, prime minister of England over the latter half of the nineteenth century and scholar of ancient Greek. Gladstone reported that Homer’s epic poems used a “paucity” of color terms, mostly relating to dark and light, with a few instances of terms that may have corresponded to *YELLOW*, *RED*, *VIOLET*, and *INDIGO*, but not *GREEN* and not *BLUE* [1]. The German philologist **Lazarus Geiger** [2] reviewed evidence from even older sources: the Hindu Veda hymns of India, the Zend-Avesta books of the Parsees, and the Old Testament of the Bible, as well as ancient Greek and Roman sources. Geiger argued that color lexicons progressed over time from a *BLACK*-and-*WHITE* system to a *BLACK*-and-*RED* system (where *RED* was his term for white or warm colors), then differentiating *YELLOW*, then adding *GREEN*, then *BLUE*. “In the earliest mental productions that are preserved to us of the various peoples of the earth . . . notwithstanding a thousand obvious and often urgently pressing occasions that presented themselves, the colour blue is not mentioned at all. . . . Of the words that in any language that are used for blue, a smaller number originally signified green; the

greater number in the earliest time signified black” [Ref. 2, pp. 49, 52].

Gladstone speculated that the Greek of the heroic age “had a less-evolved color sense that prevented him from seeing and distinguishing the many colors that modern people can see easily” [1] [p. 496]. Geiger came to a similar conclusion: “Were the organs of man’s senses thousands of years ago in the same condition as now. . . ? [p. 60] The circumstance that the colour-terms originate according to a definite succession, and originate so everywhere, must have a common cause. This cause cannot consist in the primarily defective distinction merely. . . . [W]e must assume a gradually and regularly rising sensibility to impressions of colour.”

The Cultural Explanation

Under the influence of the Darwinian thinking of the day, the English writer **Grant Allen** and the German ophthalmologist **Hugo Magnus** [23] thought that “primitive tribes” who lived in modern times could provide information on the color naming and sensory color capabilities of ancient humans. Therefore, they sent questionnaires to Christian missionaries, explorers, and diplomats around the world, asking them about the color capabilities of the people they encountered and the color terms in their informants’ native languages. Based on their responses, Allen wrote that “the colour-sense is, as a whole, absolutely identical throughout all branches of the human race” [Ref. 3, pp. 205] and afforded “a reasonable

presumption in favour of a colour-sense in the earliest members of the human race.” He therefore rejected the view, espoused by Gladstone and Geiger, that the reduced color vocabulary observed in many ancient and modern languages was due to a color vision defect. Magnus partly agreed, with qualification: “While some groups confirmed an awareness of colour, which rated in no way below that of the achievements of highly developed nations, others again gave proof of the lack of ability in identifying colours of middle- or short-wavelengths, and this was noted particularly in relationship to ‘blue’” [Ref. 4, p. 145].

Based on the results of his surveys, Grant Allen proposed the second, cultural explanation of the diversity of color naming worldwide. “Words arise just in proportion to the necessity which exists for conveying their meaning. . . . Primitive man in his very earliest stage will have no colour terms whatsoever. . . . But when man comes to employ a pigment, the name of the pigment will easily glide into an adjectival sense. . . . [p. 259]. The further differentiation of the colour-vocabulary . . . is most developed among . . . dyers, drapers, milliners, and others who have to deal with coloured articles of clothing. . . .” “How then are we to explain the singular fact, which Mr. Gladstone undoubtedly succeeds in proving, that the Homeric ballads contain few actual colour-epithets? In the following manner, it seems to me. Language is at any time an index of the needs of intercommunication, not of the abstract perceptions, of those who use it.”

Hugo Magnus became interested in the discrepancy between the excellent color awareness of many of the peoples in his survey and their difficult color naming. His summary of how color terms co-occur in languages is reminiscent of the results of Gladstone and Geiger: “. . . while in some . . . communities the known terminology begins and ends with ‘red’, it stretches in other ones well beyond the ‘yellow,’ and with yet others, even beyond the ‘green’.” Magnus began his research under the influence of Gladstone and Geiger, but in the end he was also influenced by Grant Allen’s work. Magnus concluded, “one might be tempted to formulate a . . . natural law of awareness – be that linguistically engendered

or physiologically-anatomically conditioned as part of the natural growth of man.”

Empirical Studies

The empirical tradition in the study of color terminology began with **W. H. R. Rivers**, a medical doctor and anthropologist, who traveled as an explorer to several parts of the world on behalf of the Royal Anthropological Institute. In his book *Reports of the Cambridge Anthropological Expedition to Torres Straits* [5], he compared the color vocabularies of three languages spoken by the Kiwai, Murray Islanders, and Western Tribes of the Torres Straits. “. . . As regards blue, the three languages may be taken as representatives of three stages in the evolution of a nomenclature for this colour. In Kiwai there is no word for blue; may blues are called names which mean black. . . while other blues are called by the same word which is used for green. In Murray Island there is no proper native term used for blue. Some of the natives, especially the older men, use [a native term], which means black, but the great majority use a term borrowed from English. . . . The language of the Western Tribe of Torres Straits presents a more developed stage. . . [a native term]. . . is used definitely for blue, but is also used for green. . . however, traces of the tendency to confuse blue and black still persist. . . .” Rivers also reviewed scholarly evidence from ancient and contemporary sources, including his own work in Egypt and the Andaman Islands. All the empirical evidence he reviewed supported his view that the naming of blue is highly variable across cultures: some call blue things *BLACK*, some call them *GREEN*, and some call them with a particular word for *BLUE*. He believed that *BLACK* was the most ancient term, *GREEN* was used in more developed societies, and *BLUE* was the most advanced color term.

In the twentieth century, the large-scale study of color naming across many languages was dormant until 1969, when **Brent Berlin** and **Paul Kay** published their monograph *Basic Color Terms: Their Universality and Evolution* [6]. Berlin and Kay collected a corpus of empirical color-

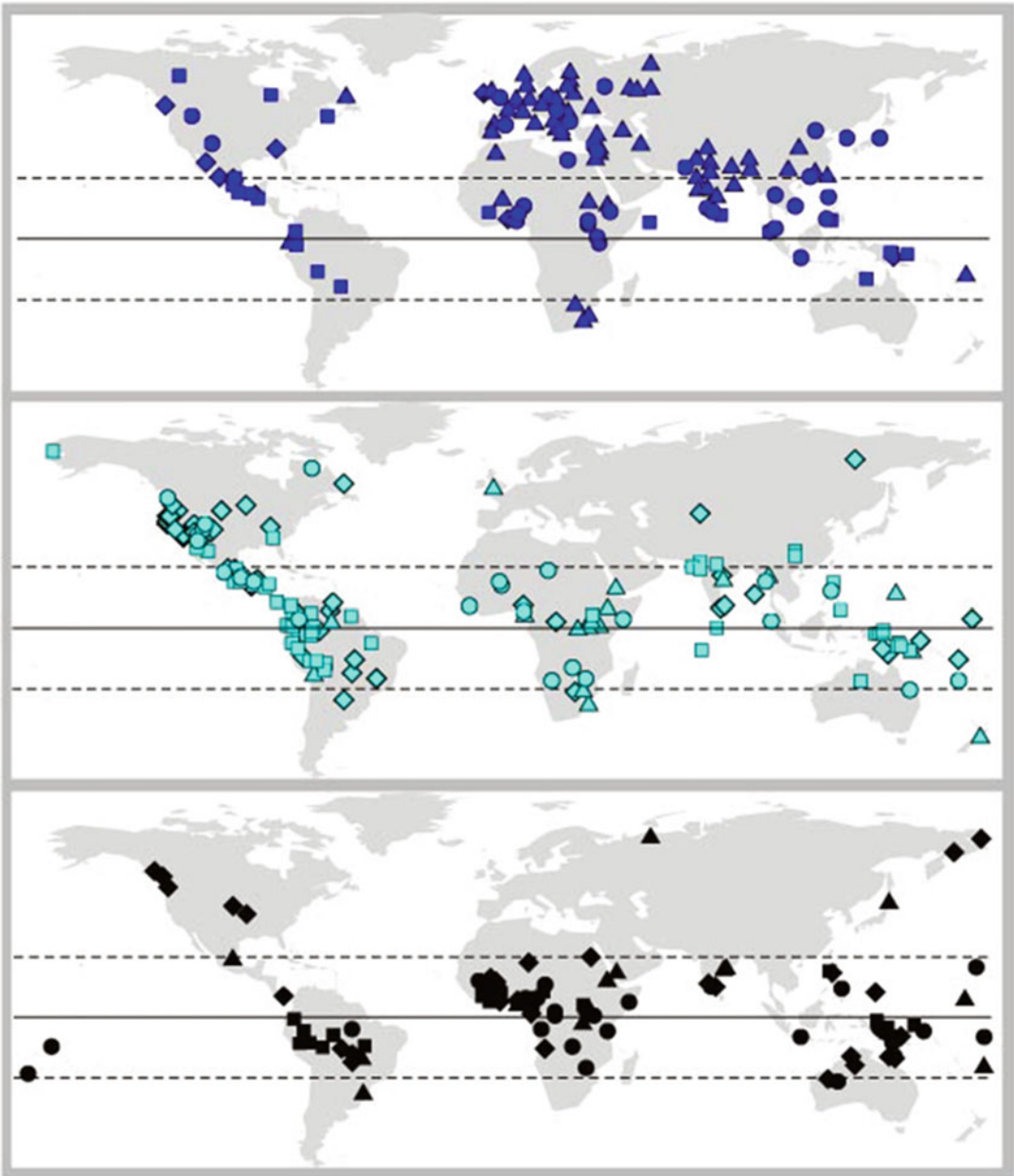
naming data on 20 languages on individual speakers who lived in the San Francisco Bay area in the mid-1960s. They showed each subject an array of Munsell color samples and asked them to indicate the range of colors they assigned to each color term in his or her native language. Berlin and Kay augmented their corpus with scholarly data on the color lexicons of 78 additional languages, which were obtained from dictionaries and other scholarly sources. Berlin and Kay observed that all the color terms in all the color lexicons in their data set were drawn on a superset of only 11 universal **basic color terms**: *BLACK*, *WHITE*, *RED*, *YELLOW*, *GREEN*, *BLUE*, *BROWN*, *ORANGE*, *PINK*, *PURPLE*, and *GRAY*. They also observed that these color terms occurred together in only about seven different combinations. They speculated that these seven combinations of basic color terms represented seven ordered **stages** along an **evolutionary sequence** whereby the most primitive languages distinguish only *BLACK* and *WHITE*, and other color terms are added in a fixed sequence, until all 11 color terms are present. Berlin and Kay assigned each language in their data set to one of their seven stages of color term evolution. Their idea about the evolution of color terms was in line with the ideas advanced by Gladstone, Geiger, Allen, Magnus, and Rivers, although their explanation for the evolution of color terms was more in line with Allen's.

The methodology of Berlin and Kay and their theoretical interpretation of their data were criticized by others [e.g., Ref. 10]. Therefore, in the 1970s, Kay, Berlin, and their colleagues collected a new corpus of data on 110 world languages: the **World Color Survey**. The languages in the World Color Survey (WCS) were mostly unwritten and were spoken in traditional societies with limited contact with Western industrialized culture. The geographical distribution of the WCS languages was generally quite similar to the worldwide distribution of all living languages (www.ethnologue.com/show_map.asp?name=World&seq=10). The WCS data set was made up of empirical color-naming data provided in face-to-face interviews by about 24 speakers of each language. Each subject viewed 330 color samples,

one at a time, and provided the color term they used in everyday life. Kay, Berlin, Maffi, Merrifield, and Cook published the *World Color Survey* [7], a book-length analysis of this corpus in which they identified each color term in each language with 1 of the 11 basic color terms of Berlin and Kay and updated their theory of color term evolution. They assigned each language to one of five stages, with two stages having three versions each, in their revised theory.

The **World Color Survey** corpus of color terms is available online and has been analyzed by Paul Kay and his colleagues [11, 12], who found evidence of universal color categories across the WCS languages. Independently, Lindsey and Brown [13] performed a cluster analysis of the color-naming patterns in the WCS corpus and discovered about eight distinct clusters of chromatic color terms, which, with the addition of the three achromatic terms *BLACK*, *WHITE*, and *GRAY*, corresponded approximately to the 11 basic color terms of Berlin and Kay. They further found that these color terms fell into about four color-naming systems ("motifs") [14], which corresponded only loosely to the seven stages of Berlin and Kay or the five stages of the WCS. Similarly to all previous scholars and investigators, Lindsey and Brown found that the motifs differed most prominently in the color terms used to name cool colors that speakers of English call "*blue*." In correspondence with the four motifs, some informants called blue samples *DARK*, some called them *GRUE*, some called them *BLUE*, and a few individuals called blue samples *GRAY* (a color term that was also used for middle-value neutral samples). Almost every World Color Survey language revealed individual differences among its speakers, and at least three of the four motifs were represented among the speakers of most languages [see also Ref. 15]. Previous empirical and scholarly work that sought the color terms in each language as a whole, including the data sets in Fig. 2, could not reveal this prominent variation among individuals.

In the tradition of Allen, Kay and his colleagues [e.g., Ref. 16] argued that larger color lexicons, at a later stage along their evolutionary sequence, occur in technologically advanced,



Color Dictionaries and Corpora, Fig. 2 The geographical distribution of languages that use BLUE/GREEN (top), GRUE (middle), or BLACK (bottom) to name the cool colors is known to be uneven worldwide. Maps show the localities of 371 living languages from color term corpora and scholarly data sets. Circles, 20 corpus languages and 77 living languages from scholarly sources from Berlin and Kay [6]; squares, 106 additional corpus languages from the World Color Survey [7], excluding two

already included in [6] and two that could not be typed; diamonds, 75 additional scholarly sources from Bornstein [8]; triangles, 93 additional scholarly sources from Brown and Lindsey [9]. Colonial languages (e.g., English, French, Spanish) are plotted where they were spoken in 1492 CE. Assignment of languages to BLUE/GREEN, GRUE, or BLACK types was made by the authors of the original data sets

economically developed cultures, where the presence of colored artifacts and trade with other cultures requires a larger, more nuanced color vocabulary. Several authors [17, 18] have examined this hypothesis by comparing the number of terms that Berlin and Kay ascribed to each language to its level of development as published by [19].

Modern investigators have examined this diversity across languages in the naming of blue, which was common to the analyses of Gladstone, Geiger, Magnus, Berlin and Kay, and the WC-S. Marc H. Bornstein [8] assembled a set of data on the presence or absence of *BLUE* in 145 languages from published sources and showed them on a world map. His analysis revealed a pronounced latitude effect: *BLACK* and *GRUE* languages tended to be spoken near the equator, and *BLUE* languages tended to be spoken at temperate latitudes (Fig. 2). Bornstein attributed this to the possible geographical variation in intraocular pigmentation, including the amount of melanin in the eye (the pigment epithelium and the iris) and the tint of the ocular media (the lens and macular pigment). Lindsey and Brown [20] reported that the yellow tint of the ocular lens, caused by the intense UVB in equatorial sunlight, could produce changes in color-naming behavior that were similar to those observed in *GRUE* languages. However, other work suggests that a causal link between the tint of the ocular lens and the naming of colors is at least partly modified by long-term chromatic adaptation [16, 21]. Brown and Lindsey [9] performed a geographical analysis of data on 118 ethnolinguistic groups for which both red-green color deficiency data (protan and deutan defects, not related to blue) and scholarly or dictionary data were available, also from published sources. The geographical results of Brown and Lindsey generally agreed with the results of Bornstein.

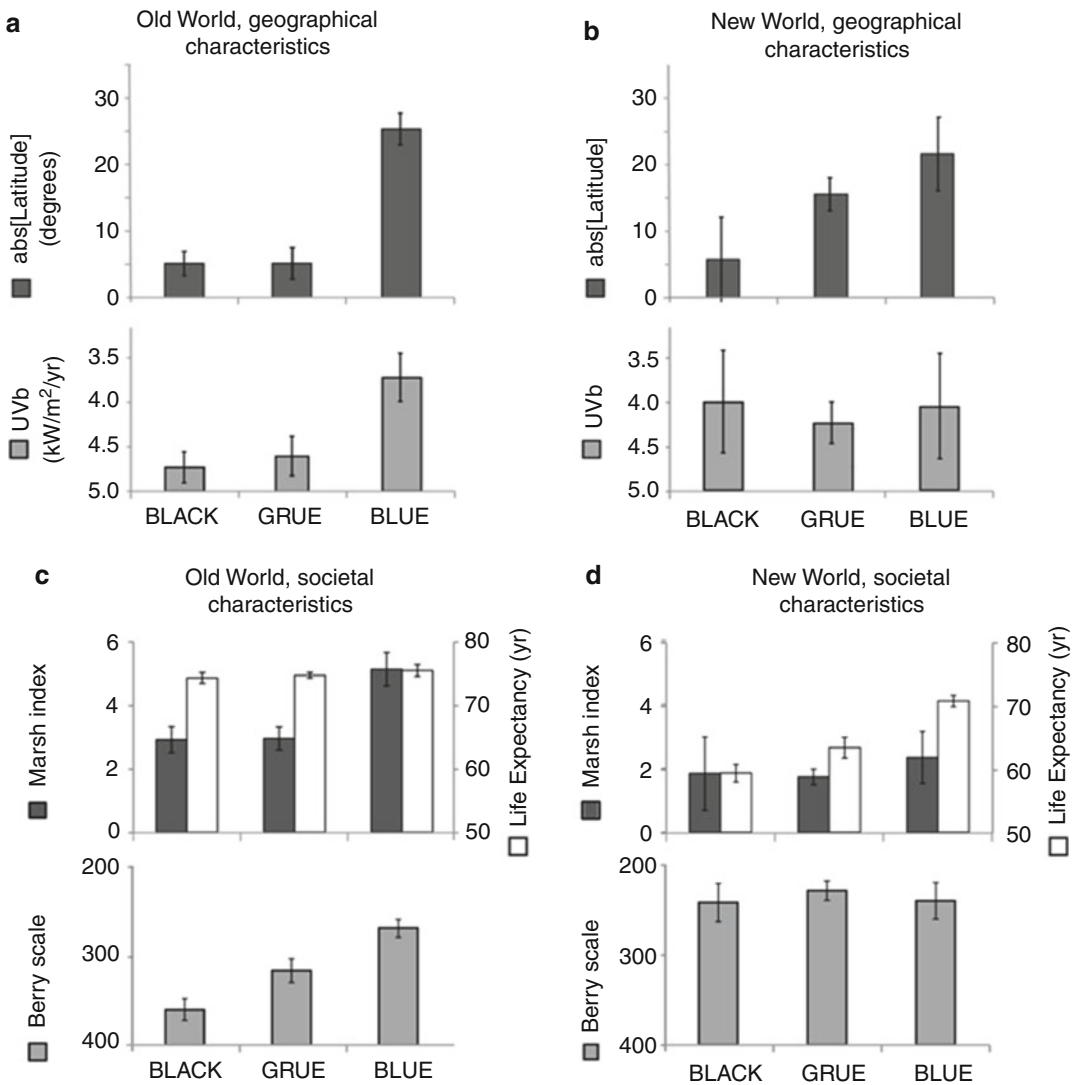
Color Naming Worldwide

There is still no single well-accepted explanation for the differences between languages in the use of *BLUE*, *GRUE*, and *BLACK*. Does *BLUE* vary across languages because of physiological differences among people, perhaps due to their different

exposure to the sun, as Bornstein and Lindsey and Brown (and Gladstone and Geiger before them) suggested? If so, there might be a correlation between *BLUE* and the physical geography of the localities where these languages are spoken. Is the variation in *BLUE* due to the superior economic and cultural development of advanced nations, as Kay and his colleagues (and Allen and Magnus before them) suggested? If so, there might be a correlation between *BLUE* and the societal characteristics of the cultures where these languages are spoken. Or, is it a historical linguistic phenomenon, with the geographical patterns in the Old World being caused by the predominance of Indo-European languages in Europe? If so, then the non-Indo-European languages spoken in Europe should not show the predominance of *BLUE* observed in the Indo-European languages.

Figure 3, panels a, b, shows two physical geography characteristics of the individual cultures from Fig. 2, latitude and the annual dose of UVB. If the physiological hypothesis is correct, both of these graphs should correlate (with positive slope on the graph) with the use of *BLUE* (*BLACK* vs. *GRUE* vs. *BLUE*). Only latitude correlates with the use of *BLUE* in both the Old World and the New World. The idea that latitude has its effect through its effect on the annual dosage of UVB from sunlight has the difficulty that UVB dosage is correlated with *BLUE* in the Old World, but not the New World. The exposure of an individual person to UVB will be modulated by the amount of time he/she spends outdoors.

Three societal characteristics are shown in Fig. 3, panels c, d. These are Marsh's "Index of Societal Differentiation," which is a measure of the societal development of an individual culture [19]; life expectancy, which is a measure of the human development of nations; and Berry's "Technological Scale" of the economic development of nations [22]. Marsh's Index and Berry's Scale correlate with *BLUE* in the Old World but not the New World, whereas life expectancy was correlated with *BLUE* in the New World but not the Old World. These three societal effects are imperfect indicators of development. Marsh's Index is only available for the individual cultures where 166 of the languages in Fig. 2 are spoken. It



Color Dictionaries and Corpora, Fig. 3 Characteristics of the cultures or nations where the languages from Fig. 2 are spoken. Each bar indicates the value of the characteristic, averaged (± 1 s.e.m) across cultures or nations having *BLACK*, *GRUE*, or *BLUE* as the word for blue in their languages' color lexicons. (a, b) Geographical characteristics of the localities where the 384 languages are spoken. Dark bars, absolute value of latitude, in degrees from the

equator; light gray bars, annual dose of UVB. (c, d) Societal characteristics. Dark bars, Marsh's "Index of Societal Differentiation," for 166 of the cultures shown in Fig. 2 [17, 19]. White bars, life expectancy; light gray bars, Berry's "Technological Scale" of economic development [22]. Life expectancy and Barry's Scale are shown for the nations where the languages are spoken

is based on the per capita annual energy consumption and the fraction of males engaged in agriculture. Therefore, it will be correlated with latitude (it takes more energy to keep warm in Alaska than in Cameroon, and the short growing season makes agriculture in Siberia an unprofitable occupation). Life expectancy and Berry's Scale are available

only for the nations within which the cultures are embedded. Life expectancy may show a ceiling effect in the Old World. Berry's Scale was the first principal component of an overall assessment of the economic advancement of nations, with contributions from transportation and trade, energy production and consumption, national product,

communications, and urbanization. All of these societal characteristic correlates have the difficulty that none of them correlate with *BLUE* in both the Old World and the New World.

Languages that include *BLUE* greatly predominate Europe, the Mediterranean, and the Near East (Fig. 2, top panel). However, this is not entirely due to the predominance of Indo-European languages in that area. The data sets in Fig. 2 include 46 Old World languages that are spoken north of the Tropic of Cancer and west of the Ural Mountains. Of the 28 Indo-European languages and 18 non-Indo-European languages in this group, all but two use *BLUE*; one Indo-European language (Gaelic) uses *GRUE*, and one non-Indo-European language (Nenets) uses *BLACK*. Whatever is responsible for the predominance of *BLUE* in this geographical region, it is not entirely a question of linguistic heritage.

In spite of over 150 years of research, involving empirical corpora of color-naming data, scholarly data sets of color lexicons, and philological analyses of ancient texts, it is not well understood why there is such great worldwide variation in the terms in the color lexicons of world languages. This topic continues to be the subject of much contemporary research.

Cross-References

- [Berlin and Kay Theory](#)
- [Evolution](#)
- [Unique Hues](#)
- [World Color Survey](#)

References

1. Gladstone, W.E.: Studies on homer and the Homeric age, vol. III. Oxford University Press, London (1858)
2. Geiger, L.: Contributions to the history of the development of the human race. Trubner & Col, London (1880)
3. Allen, G.: The colour-sense: its origin and development: an essay in comparative psychology. Houghton, Boston (1879)
4. Saunders, B., Marth, I.-T.: The debate about colour-naming in 19th century German philology. Leuven University Press, Leuven (2007)
5. Rivers, W.H.R.: Vision. In: Haddon, A.C. (ed.) Reports on the Cambridge anthropological expedition

- of the Torres straits, pp. 1–132. Cambridge University Press, Cambridge (1901)
6. Berlin, B., Kay, P.: Basic color terms: their universality and evolution. University of California Press, Berkeley/Los Angeles (1969)
7. Kay, P., et al.: The world color survey. CSLI, Stanford (2009)
8. Bornstein, M.H.: Color vision and color naming: a psychophysiological hypothesis of cultural difference. Psychol. Bull. **80**(4), 257–285 (1973)
9. Brown, A.M., Lindsey, D.T.: Color and language: worldwide distribution of Daltonism and distinct words for “blue”. Vis. Neurosci. **21**, 409–412 (2004)
10. Hickerson, N.R.: Basic color terms: their universality and evolution by Brent Berlin; Paul Kay. Int. J. Am. Linguist. **37**(4), 257–270 (1971)
11. Kay, P., Regier, T.: Resolving the question of color naming universals. Proc. Natl. Acad. Sci. U. S. A. **100**, 9085–9089 (2003)
12. Regier, T., Kay, P., Khetarpal, N.: Color naming reflects optimal partitions of color space. Proc. Natl. Acad. Sci. U. S. A. **204**, 1436–1441 (2007)
13. Lindsey, D.T., Brown, A.M.: Universality of color names. Proc. Natl. Acad. Sci. U. S. A. **103**, 16608–16613 (2006)
14. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within-language diversity. Proc. Natl. Acad. Sci. U. S. A. **206**, 19785–19790 (2009)
15. Webster, M.A., Kay, P.: Variations in color naming within and across populations. Behav. Brain Sci. **28**(4), 512 (2005)
16. Hardy, J.L., et al.: Color naming, lens aging, and Grue: what the optics of the aging eye can teach us about color language. Psychol. Sci. **16**, 321–327 (2005)
17. Hays, D.G., Margolis, E., Naroll, R., Perkins, D.R.: Color term salience. Am. Anthropol. **74**, 1107–1121 (1972)
18. Ember, M.: Size of color lexicon: interaction of cultural and biological factors. Am. Anthropol. **80**(2), 364–367 (1978)
19. Marsh, R.M.: Comparative sociology: a codification of cross-societal analysis. Harcourt, Brace & World, New York (1967)
20. Lindsey, D.T., Brown, A.M.: Color naming and the phototoxic effects of sunlight on the eye. Psychol. Sci. **13**(6), 506–512 (2002)
21. Delahunt, P.B., et al.: Long-term renormalization of chromatic mechanisms following cataract surgery. Vis. Neurosci. **21**, 301–307 (2004)
22. Berry, B.J.L.: An inductive approach to the regionalization of economic development. In: Ginsberg, N. (ed.) Essays on geography and economic development, pp. 78–107. University of Chicago, Chicago (1960)
23. Schöntag, R., Schäfer-Prieß, B.: Color term research of Hugo Magnus. In: MacLaury, R.E., Paramei, G.V., Don Dedrick, D., (eds) Anthropology of Color: Interdisciplinary Multilevel Modeling, pp. 107–122. John Benjamins, Amsterdam (2007)

Color Directions

► Color Trends

Color Dynamics

► Environmental Color Design

Color Effects to Humans and the Environment

► Functionality of Color

Color Environment

► Color Scene Statistics, Chromatic Scene Statistics

Color Experience

► Color Phenomenology

Color Field Painting

Zena O'Connor

Architecture, Design and Planning, University of Sydney, Sydney, Australia

Synonyms

[Hard-edge painting](#)

Definition

Color Field painting is an art movement that emerged in the mid-twentieth century. Artists whose works can be categorized as Color Field chose to focus predominantly on the use of color in their works almost to the exclusion of other

visual elements. Color Field paintings of the twentieth century are mostly works on canvas, and some artists applied color in a formal, hard-edge manner, while others chose to apply color in a more organic, free-form manner. The focus on color as demonstrated by the Color Field artists continues to influence contemporary artists; however, contemporary artists often explore color across a wider variety of media.

Overview

Of the three main currents in art that emerged in the twentieth century, Expressionism, Abstraction, and Fantasy, Color Field painters were inspired by the developments in Expressionism and Abstraction [1]. The focus of Expressionism was the artist's feelings and emotional responses via their conceptual content, subject matter, and painterly technique; and the focus of Abstraction was a more conceptual approach to the partial or complete nonrepresentational depiction of subject matter depicted with more formally structured painterly techniques.

Color Field painting emerged in the 1940s, and painters such as Mark Rothko (1902–1970), Helen Frankenthaler (1928–2011), and Morris Louis (1912–1962) focused on expressing emotion through painting, while Clyfford Still (1904–1980), Barnett Newman (1905–1970), and Ellsworth Kelly (born 1923) applied a more formal, structured approach to their conceptual content and composition. The predominant focus that all Color Field painters shared was the use of color as the key conduit for conveying emotional or conceptual content.

Art critic Harold Rosenberg suggested that abstract painting represented a new function whereby the canvas became an “arena in which to act. . . What was to go on the canvas was not a picture but an event” [2, p. 22]. From Rosenberg's perspective, “a painting is inseparable from the biography of the artist” [2, p. 23]. These views were countered at the time by art critic Clement Greenberg who preferred to focus on the formal qualities such as shape, color, and line rather than the act of painting. Greenberg applauded the

works of Still, Newman, and Rothko with their primary focus on color and marveled that their paintings “exhale color with an enveloping effect that is enhanced by size” [3, p. 226].

The approach of the Color Field painters reflected the nineteenth-century impressionist painter’s tradition of using patches of color to capture a scene’s ambience and convey a sense of movement plus capture a fleeting moment in time. This approach came about when Charles Baudelaire suggested that painters should “evoke reality, not by detailing its forms, but by using a line or patch of tone to stimulate the spectator to recreate reality through the act of imagination” (Baudelaire 1863, cited in [4, p. 9]).

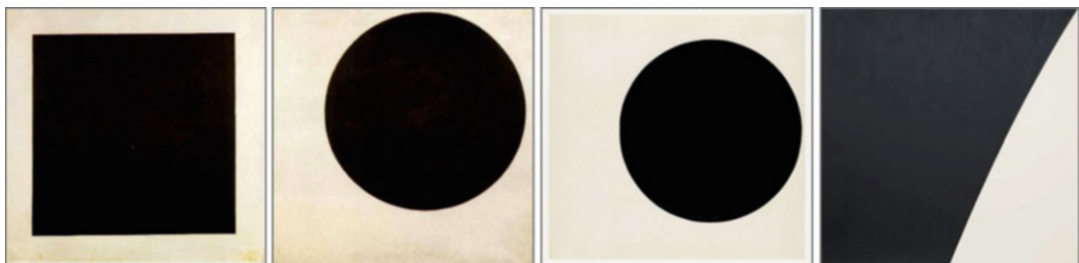
Color Field painters translated the impressionists small-scale color patches to much larger fields of color, encouraging their viewers to engage with their paintings on a more intimate but still static basis. Color was used to convey ambience as well as universal emotions such as *joie de vivre*, tension, tragedy, and tranquility. Rothko suggested that painting represented a “portal . . . into the vast recesses of the human psyche” and the role of color was to generate an emotional response: “I’m interested in eliciting basic human emotions – tragedy, ecstasy, doom and so on” (Rothko, cited in [5, p. 6]).

Using color as a form of communication or code, Color Field painters followed a long tradition. Vincent van Gogh (1853–1890) believed that color had the capacity to connect with the human condition: “I have tried to express the terrible passions of the human heart. . .” (Van Gogh, cited in ([6], p. 72)). In letters to his brother

Theo, Van Gogh indicated that he attached emotional meanings to various hues and he clearly preferred strong, saturated color [7]. Similarly, Paul Gauguin (1848–1903) and the symbolists believed that color could “act like words; that it [color] held an exact counterpart for every sensation, every nuance of feeling” [8, p. 129]. Likewise, Edvard Munch (1863–1944) used color and darker tones to help convey universal emotions such as anxiety, fear, or sorrow in paintings like *The Dance of Life* (1899) and *Death in the Sick-room* (1895). Wassily Kandinsky (1866–1944) also believed that color as well as form had the capacity to communicate, and he assigned certain connotations to specific colors; yellow, for example, represented warmth [9].

The conceptual and compositional approach of Suprematism and Russian Constructivism also had some influence on Color Field painting as evidenced by the patterns of similarity in the works of Kasimir Malevich (1879–1935) and El Lissitzky (1890–1941) and those of Rothko, Kelly, and Newman. While there may be some conceptual or philosophic differences between the movements, there are strong similarities in the use of simplified geometric form as per Malevich’s *Black Square* (1915) and *Black Circle* (1915) and Kelly’s *Circle Form* (1961) and *White Curve* (1976) (see Fig. 1).

Mark Rothko (1903–1970) aimed to engage the viewer on a deeper, more personal level and “relied on large fields of color to produce solemn and elevated works” that had the power to convey something about the human condition [8, p. 314]. Eschewing other visual elements, Rothko



Color Field Painting, Fig. 1 Patterns of similarity: Suprematism and Color Field painting



Color Field Painting, Fig. 2 Color Field paintings, Mark Rothko

suggested: “We may start with color,” which became the primary focus of his works (Rothko, cited in [10, p. 65]).

Rothko said: “I’m interested only in expressing basic human emotions: tragedy, ecstasy, doom, and so on” [11]. The size of Rothko’s canvases, which often feature blocks of saturated color horizontally stacked, reveals his desire to create an intense experience for the viewer, and he suggested that the ideal viewing distance for his works was a close arm’s length distance from the canvas [5]. Rothko noted, “I want to be very intimate and human. To paint a small picture is to paint yourself outside your experience, to look upon an experience as a stereopticon view or with a reducing glass. . . . However, to paint the larger picture, you are in it. It isn’t something you command. . . .” [8, p. 320]. Rothko’s paintings are a presence in themselves, so large and intensely colored that one is expected to feel their “spiritual vibration” [6, p. 72].

For Rothko, “A painting is not about experience. It *is* an experience”; an experience that transcends time and space conveying universal primal emotional experiences irrespective of gender, age, and cultural experience (Rothko, cited in [12, p. 57]). Rothko noted “My art is not abstract, it lives and breathes” (Rothko, cited in [12, p. 50]). Key works by Rothko include *Untitled [Blue, Green, and Brown]* (1952), *Ochre and Red on Red* (1957), *Light Red over Black* (1957), *Untitled* (1968), and *Black Form No. 8* (1964), as per Fig. 2.

Barnett Newman (1905–1970) used the color palette of the De Stijl movement and the Bauhaus in a completely different way: “Why give in to

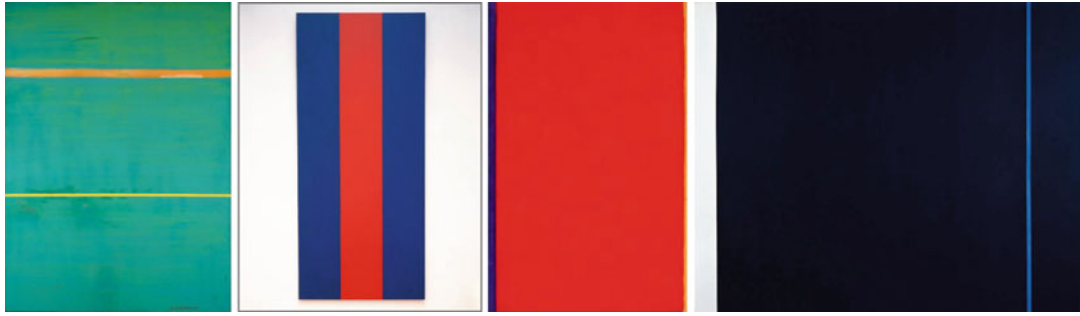
these purists and formalists who have put a mortgage on red, yellow and blue, transforming these colors into an idea that destroys them as colors?” (Newman, cited in [10, p. 27]).

Newman commented that “The central issue of painting is the subject matter . . . my subject is anti-anecdotal” whereby a painting is more self-sufficient and independent with color and shape standing alone and prominent without reference to anything else [13].

Newman’s paintings often feature vertical lines that serve as “an act of division and creation” [between one color plane or reality and another] . . . the “zip” has become the single most dramatic event in the composition” [14, p. 77]. Key paintings by Newman include *Dionysius* (1949), *Voice of Fire* (1967), *Who’s Afraid of Red, Yellow and Blue?* (1966), and *Midnight Blue* (1970), as per Fig. 3.

Helen Frankenthaler’s (1928–2011) work became synonymous with a freer, more sensuous approach to Color Field painting [15]. Frankenthaler’s works are large, and she applied color in a technique known as soak stain, wherein oil painted was diluted with turpentine so that the color soaked into the canvas creating halos of color.

The free-form and more sensuous nature of her paintings prompted Hughes to suggest that “Landscape, imagined as Arcadia, remained the governing image in Frankenthaler’s work, and her titles often invoked the idea of Paradise or Eden” [8, p. 154]. Key works by Frankenthaler include *Mountains and Sea* (1952), *Yellow Span* (1968), and *Nature Abhors a Vacuum* (1973), illustrated in Fig. 4.



Color Field Painting, Fig. 3 Color Field paintings, Barnett Newman

Color Field Painting, Fig. 4 Color Field paintings, Helen Frankenthaler



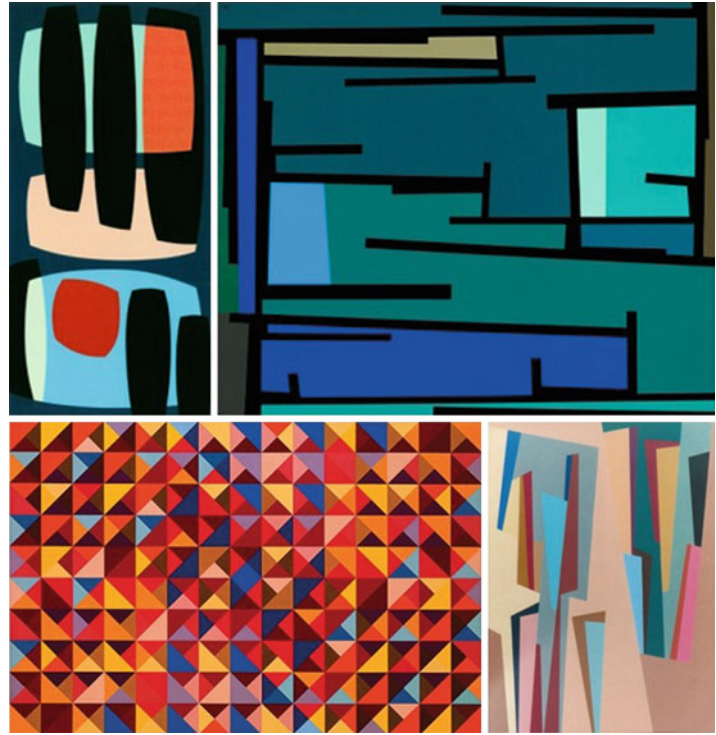
The use of color by the New York-based Color Field painters influenced the US West Coast painters as well as artists further afield, some of whom used new materials and techniques.

Karl Benjamin (1925–2012), whose work featured in the exhibition *Four Abstract Classicists* (1959–1960), contributed to the US West Coast response to New York Abstract Expressionists. Benjamin’s work shared similarities with the Color Field painters, and his works feature a sense of movement and vibrant optimism characteristic of the cool aesthetic of California during the 1950s and early 1960s.

The use of color in Benjamin’s work in conjunction with the repetition of shape and line creates works that are “fixed and stable” but in a state of “continuous flux” and convey a strong sense of movement and vitality [16]. Key works include *Black Pillars* (1957), #38 (1960), #7 (1966), and #7 (1986), illustrated in Fig. 5.

Lorser Feitelson (1898–1978), whose works also featured in the exhibition *Four Abstract Classicists* (1959–1960), painted planes of color interspersed with a fluid use of line that imbued his layered abstract works with a strong sense of graceful movement, vitality, and dynamism. Key

Color Field Painting,
Fig. 5 Paintings, Karl
 Benjamin



works by Feitelson include *Hardedge Line Painting* (1963), *Untitled* (1965), *Untitled* (1967), and *Coral and Blue Abstract* (1967), illustrated in Fig. 6.

Richard Paul Lohse (1902–1988) created color investigations that “deploy the full range of spectral colors in endlessly inventive series” [10, p. 43]. Lohse painted individual colors in large grid formations to illustrate the notion that “the crowd contains the possibility of the individual” and also to highlight that color juxtaposition had the capacity to annul the “sovereignty of the [individual] color square” (Lohse, cited in [10, p. 45]). Lohse’s works vary in scale from relatively small to very large, to enable an in-depth investigation of color and color juxtaposition [17]. Key works include *Progressive Reduction* (1942–1943), *Thirty Vertical Systematic Color Series in a Yellow Rhombic Form* (1943, 1970), and *Thematic Series in 18 Colors* (1982), see Fig. 7.

In the twenty-first century, artists have translated the idea of patches of color and the approach taken by Color Field painters into a variety of new

and different formats and mediums, often incorporating light, automated movement, as well as the mechanics of human perception.

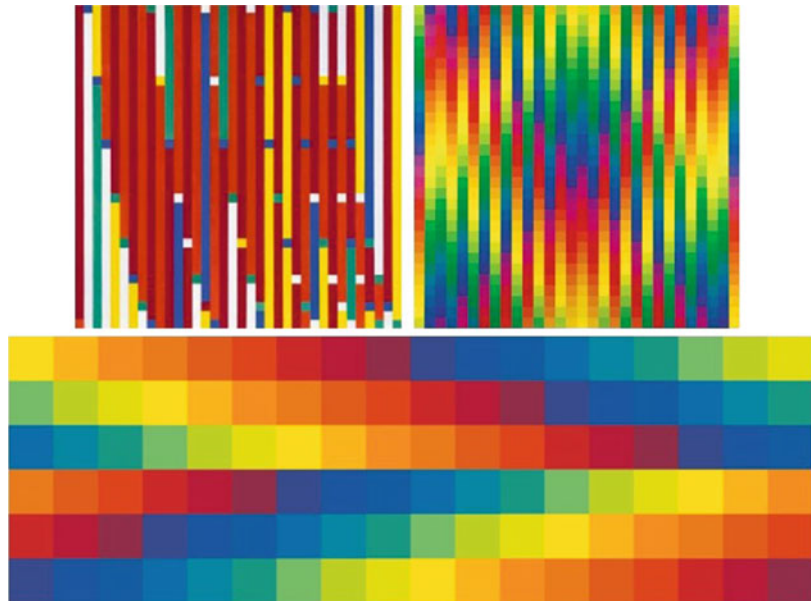
Gerhard Richter (born 1932) adopted the grid format of the early modernists and imbued this with an abundance of color that could be grouped at random, as per the series of large-format paintings: *256 Colors* and *4900 Colors* [10]. Richter’s random grouping of color mimics the complexity of color in nature, and he painted several versions of *256 Colors* (1974) plus recolored them in the 1980s. The work *4900 Colors* (2007) comprises 196 equal-size panels each containing 25 squares of color, as depicted in Fig. 8.

Robert Owen (born 1937) also prefers depicting color via a grid format as it is “capable of providing infinite variations” and suggests that his paintings are “about levels of feelings, orders of sensation, shifting sequences of time and rhythm” [18, p. 9]. In *Cadence* (2003), Owen has depicted the range of his emotions measured using a color tabulation. Responding to a competition for a public work by the Bureau of Meteorology, Owen said: “I thought,

Color Field Painting,
Fig. 6 Paintings, Lorser
Feitelson



Color Field Painting,
Fig. 7 Paintings, Richard
Paul Lohse



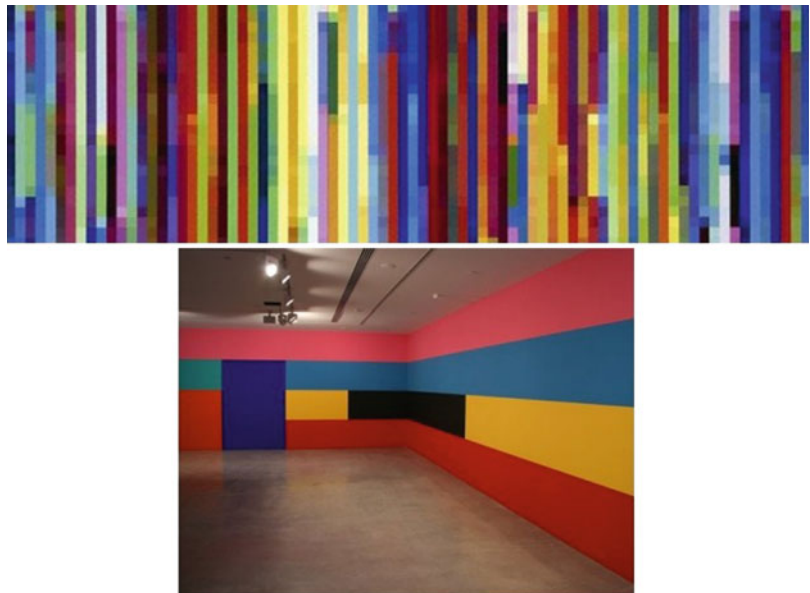
if they can measure atmosphere, I must be able to measure emotions. So, using a color tabulation, I intuitively picked how I felt every half hour during the day” [19]. The juxtaposition of color

and constant variation triggers a strong sense of movement across the work. *Sunrise #3* (2005), painted in situ in the Museum of Contemporary Art, Sydney, explores the impact of color by taking

Color Field Painting,
Fig. 8 Paintings, Gerhard
 Richter



Color Field Painting,
Fig. 9 Paintings,
 Robert Owen



the Impressionists notion of patches of color and translating this into a much larger work (Fig. 9).

Dion Johnson (born 1975), whose works are reminiscent of the works of Benjamin and

Feitelson, focuses on the interplay of color (Glier, 2011). Colors converge and sideswipe each other, setting off color juxtapositions that add a sense of dynamism and vitality to Los

Color Field Painting,
Fig. 10 Paintings, Dion
Johnson



Color Field Painting,
Fig. 11 Works by Rebecca
Baumann (*left*) and
Brendan Van Hek (*right*)



Angeles-based Johnson's large-scale canvases. Works such as *Velocity* (2012) and *Wild* (2012) feature hard-edge yet malleable color, as depicted in Fig. 10.

Rebecca Baumann (born 1983) has adopted the Color Field painters approach to color but added mechanized movement. In *Automated Color Field* (2011), the automated movement brings

Color Field Painting,
Fig. 12 Film posters
 designed by Saul Bass



ever-changing random juxtaposition to Baumann patches of color. Baumann is interested in the way “color is both universal and subjective” with the capacity to “move people beyond cognitive and conscious thought” (Baumann, cited in [20]). Baumann acknowledges the changeable nature of human response, and her artworks feature “apparently random change(s) across the field,

suggesting the flux in both our inner emotions and the outside world” (Baumann, cited in [21]).

Brendan Van Hek (born 1968) uses found neon such as remnants of advertising signs and symbols to create works like *Color Composition #3* (2013). In doing so, Van Hek subverts the original neon messages to create works that reflect the visual landscape and clutter of the twenty-first



Color Field Painting, Fig. 13 Poster and CD cover design inspired by Color Field painting

century: “*Color Composition #3* started as an idea about landscape – a horizon line and the movement of forms along, below and above it. What the work has developed into is one that describes a landscape of noises, activity, words, streaking taillights, constant action and movement – the city” (Van Hek, cited in [22]) (Fig. 11).

Color Field Painting: The Influence on Design

The unique approach to color by the Color Field painters influenced graphic design in the late twentieth century and early twenty-first century. Film posters designed by Saul Bass feature a hard-edge, Color Field painting aesthetic: *The Man with the*

Golden Arm (1955), *Anatomy of a Murder*, by Saul Bass (1959), *The Cardinal* (1963), and *The Human Factor* (1979), illustrated in Fig. 12.

Similarly, posters for *Before Sunrise* (1995), *The End of Summer* (2013), and *Rush* (2013) as well as recent ECM CD covers including Keith Jarrett’s *Back Hand* (1974) and *Rio* (2011) are reminiscent of the Color Field paintings of Barnett Newman and Ellsworth Kelly as well as Mark Rothko, as illustrated in Fig. 13.

Color Field painting also influenced textile and fashion design. The impact of this art movement continues with the fashion trend for color blocking that emerged in 2010 as well as the textile designs of Satu Maaranen (2011), Marimekko Sabluuna dress, Spring 2013, and the Gucci Spring/Summer 2013 collection, illustrated in Fig. 14.

MAJORENKO CONCEPT STORE UPPER EAST SIDE NEW YORK

marimekko

CONCEPT STORE

APRILS

MADE

INTERIORS

ACCESSORIES & GIFTS

SPRINGS 2012 LOOK BOOK

SALE

Shipping Bag 0 Items \$1,610

SPRINGS 2012 LOOK BOOK

SALE



Collection: Subluna by Pina Pina for Spring 2012

Print: Seta Marimekko 2011

Material: 100% woven cotton saten

Length: 65" (165cm) 2011 Nylon

Description: Above baby-doll style

Short sleeves with fold-over cuffs, Jewel neckline

Length: 34" (86cm) (measurement based on size 38/39)

Care: Machine wash cold, gentle cycle. Do not tumble dry

Click for the women's [sizing chart](#) also available in [metric sizes](#)

SHARE: [Facebook](#) [Twitter](#) [Pinterest](#) [Email](#)

Be the first of your friends to like this



Elie Soab



Bottega Veneta



Narciso Rodriguez



Sportmax



Christopher Kane

Color Field Painting, Fig. 14 Fashion and textiles inspired by Color Field painting

Color Field Painting: The Impact on Immersive Color Installations

The Color Field painters have influenced many subsequent artists, who create works that focus on color with the aim of connecting with the viewer and encouraging a deeper level of engagement often on a more intimate, immersive basis as per the works of Verner Panton, Peter Jones, Olafur Eliasson, Gabriel Dawe, Matthew Johnson, and James Turrell. A common theme of these artists is exploration of the interface between color and human response in ways that leave the viewer enriched with a greater understanding about color.

Immersive works and projections are not viewed at arm's distance but often envelope the viewer and elicit not only emotional response but trigger perceptual responses that form part of the work as a whole. While some of these works remain static, some works, such as those of Olafur Eliasson, involve a degree of radical subjectivity on the part of the viewer, thereby making their perceptive subjectivity a component of the work [23].

Cross-References

- [Art and Fashion Color Design](#)
- [Color Trends](#)
- [Impressionism](#)
- [Neon Lamp](#)
- [Palette](#)

References

1. Janson, H.W.: *The History of Art*, revised and expanded by A.F. Janson. Thames & Hudson, New York (1995)
2. Rosenberg, H.: The American action painters. *Art News* **51**(8), 22–23 (1952)
3. Greenberg, C.: *Art and Culture: Critical Essays*. Beacon, Boston, MA (1961)
4. Green, P. (ed.): *Monet and Japan*. National Gallery of Australia, Canberra (2001)
5. Kelsey, A.: The color of transcendence. *Anamesa* **3**(2), 3–14 (2005)
6. Bird, M.: *One Hundred Ideas that Changed Art*. Laurence King, London (2012)
7. Cabanne, P.: *Van Gogh*. Abrams, New York (1971)
8. Hughes, R.: *The Shock of the New: Art and Century of Change*. Thames & Hudson, London (1991)
9. Kandinsky, W.: *Über das Geistige in der Kunst*. (R. Piper, Munich (1912). English transl. *Concerning the Spiritual in Art*.) General Books, Memphis (2010)
10. Gage, J.: *Color in Art*. Thames & Hudson, London (2006)
11. MoMA: Mark Rothko http://www.moma.org/collect/object.php?object_id=78485 (2013). Retrieved 29 Mar 2013
12. Baal-Teshuva, J.: *Mark Rothko: Pictures as Drama*. Taschen, Bonn (2003)
13. Gershman, R.: Barnett Newman <http://www.theartstory.org/artist-newman-barnett.htm> (2012). Retrieved 2 Apr 2013
14. Adams, T.: He had to draw the line somewhere. *The Observer*, Sunday 22 September 2002, Features, p. 77. <http://www.guardian.co.uk/theobserver/2002/sep/22/features.review77>. Retrieved 2 Apr 2013
15. Glueck, G.: Helen Frankenthaler: Abstract painter who shaped a movement dies at 83. *The New York Times*, 28 December 2011. <http://www.nytimes.com/2011/12/28/arts/helen-frankenthaler-abstract-painter-dies-at-83.html>. Retrieved 10 Mar 2013
16. Benjamin, K.: Karl Benjamin – Biography <http://www.karlbenjamin.com/> (2013). Retrieved 5 Oct 2013
17. Richard Paul Lohse Foundation. Richard Paul Lohse: Biography http://www.lohse.ch/bio_e.html (2012). Retrieved 10 Mar 2012
18. Museum of Contemporary Art, Sydney: Almanac – The gift on Ann Lewis AO http://www.wagga.nsw.gov.au/_data/assets/pdf_file/0017/9116/Almanac_Education_Kit.pdf (2009). Retrieved 20 Mar 2013
19. Art Gallery NSW: Robert Owen – Cadence <http://www.artgallery.nsw.gov.au/collection/works/291.2004.a-e/> (2003). Retrieved 18 Nov 2010
20. Museum of Contemporary Art, Sydney: Rebecca Baumann - Automated Colour Field, 2012; Statement of significance by Anna Davis <http://www.mca.com.au/collection/work/201120/> (2012). Retrieved 28 Mar 2013
21. Institute of Modern Art: Rebecca Baumann, Automated Monochrome Field 2011. <http://www.ima.org.au/pages/exhibits/rebecca-baumann-at-imaksubi245.php?short=1> Retrieved 28 Mar 2013
22. Lawrence Wilson Art Gallery, University of Western Australia: Luminous Flux exhibition <http://www.lwgallery.uwa.edu.au/exhibitions/?a=2260278> (2013). Retrieved 10 Apr 2013
23. Grynstejn, M.E.: *Take Your Time: Olafur Eliasson*. Thames & Hudson, New York (2007)

Color from Motion

- [Motion and Color Cognition](#)

Color Harmonies for Fashion Design

- [Art and Fashion Color Design](#)

Color Harmony

Antal Nemcsics¹, Zena O'Connor² and
Renata Pompas³

¹Department of Architecture, Budapest
University of Technology and Economics,
Budapest, Hungary

²Architecture, Design and Planning, University of
Sydney, Sydney, Australia

³Milan, Italy

Synonyms

Balance; Coherence; Color accord

Definition

A sense of accord and balance among colors in a visual composition or design, resulting in a positive affective response and/or cognitive judgment about color combination.

Color Harmony Theories

Many theories about color harmony have evolved since antiquity across different fields including art, design, psychology, and physics. Each of these has a different focus, and hence there is little consensus in terms of defining and describing the construct “color harmony.” Most theories about color harmony are highly prescriptive, and this is their weakness – they assume that responses to color are deterministic, uniform, universal, and fixed irrespective of context rather than idiosyncratic, less deterministic, and open to the influence of factors relating to culture and context.

The problem of color harmony has long been regarded as an esoteric matter for the artist. But since the advent of interdisciplinary research on colors, color harmony has become an object of scientific research giving rise to several theories. These theories emphasize different relations between the role of color harmony in environment and man, his culture, and his message and

consider these as the exclusive laws of harmony. The scientifically established theories deal almost exclusively with the first level of the content of the concept of color harmony. They research the connections, which are mostly the same for all people. These relations express interactions between color perceptions, which can be described, in sum, in terms of relations between color perception parameters: hue, saturation, and lightness of harmonizing colors. These relations lend names to the harmony types such as complementary harmony, triadic harmony, scale of equal saturations, and others.

Color Harmony: A Twenty-First-Century Approach

Color harmony is a complex notion because human responses to color are both affective and cognitive, involving emotional response and judgment. Hence, our responses to color and the notion of color harmony are open to the influence of a range of different factors. These factors include individual differences (such as age, gender, personal preference, affective state, etc.) as well as cultural, subcultural, and socially based differences which give rise to conditioning and learned responses about color. In addition, context always has an influence on responses about color and notion of color harmony, and this concept is also influenced by temporal factors (such as changing trends) and perceptual factors (such as simultaneous contrast) which may impinge on human response to color. The following conceptual model illustrates this twenty-first-century approach to color harmony:

$$\text{Color harmony} = f(\text{color } 1, 2, 3 + n) \\ * (\text{ID} + \text{CE} + \text{CX} + \text{P} + \text{T})$$

wherein color harmony is a function (f) of the interaction between the color/s (col 1, 2, 3 + n) and the factors that influence positive aesthetic response to color which include the following: individual differences (ID) such as age, gender, personality, and affective state, cultural experiences (CE) and the influence of cultural and social conditioning, influence of the prevailing context

(*CX*) of color(s) which may include setting and ambient lighting, intervening perceptual effects (*P*) such as simultaneous contrast, and, finally, the effects of time (*T*) in terms of prevailing social trends [1].

Harmony as a Balance Between Psychophysical Forces

According to the first theory on color harmony, the rules for producing color harmony experiences are determined by the mechanics of color vision. The trigger of the theory has been the phenomenon of successive contrast: longer observation of green results in a red afterimage. Investigating this phenomenon, Rumford [2] has established that colors of successive color contrast mixed in an “appropriate proportion” in the Maxwell disk are resulting in a medium gray color. Goethe has built his theory of color harmony based on this observation. He thought the colors of successive contrast and simultaneous contrast are identical [3]. Therefore he stated that the “appropriate proportion” established by Rumford expresses the magnitude of color surfaces creating harmony experiences.

Chevreul, the French chemist, arrived at similar conclusions, stressing the role of complementarity in creating harmony, supporting his theory by successive contrast phenomena [4]. His message transmitted by Delacroix was translated by Seurat and Signac into the practice of painting [5, 6, 7]. Seurat himself experimented with harmonies, recapitulating his results by stating that harmony is a unity of opposites and similarities, a principle serving as the theoretical basis of pointillism.

Goethe’s theory has appeared with the mediation of Hölzel [8] in the works on the theory of art written by Kandinsky, Klee, Itten, Albers and Moholy-Nagy [9–13]. Itten, in his book written for the students of the Bauhaus school, has fixed this proportion in relation to main colors [14]. He established that the appropriate mutual proportion of the colors yellow, orange, red, violet, blue, and green is as follows: 9:8:6:4:6.

Krawkow has demonstrated by experiments that the colors of successive contrast are not identical to the colors of simultaneous contrast [15]. This experimental result has refuted the

conclusions of harmony theories originated from Rumford. Successive contrast is a physiological phenomenon, while the simultaneous color contrast is an aesthetic influence.

Harmony as Arrangement of Colors in Scales

It was observed by textile dyers and printers that mixing a color with white, gray, or black in different proportions led to very attractive, harmonic color complexes. It was one of the secrets of the trade of painters to make each of their colors dull by admixing a color in different proportions, helping colors in the picture to form harmonic complexes. Both approaches arose from the recognition that colors with uniformly varying saturation or lightness, i.e., those forming a scale, appeared harmonic. This observation was the starting point of the second generation of theories on harmony.

The first scale of color harmony was constructed by Newton and published in his *Optics* in 1704 [16]. Dividing the spectrum into seven colors, he paralleled it to the musical scale. His idea was further developed by Hoffmann in his book published in 1786 [17], explaining color harmonies with the aid of acoustic analogies [18]. In 1810, the painter Runge suggested and attempted to develop a unified system of musical and color harmonies. These ideas misdirected the research on color harmonies for a long time since even their followers were concerned only with hue scales.

Ostwald [19] and Plochère [20] were the first to define the arrangement in scales by writing relationships between saturation and lightness. Scale members were described by Ostwald as additive color mixing and by Plochère as subtractive color mixing components. In his theory of harmonies, Ostwald pointed out that in order to find every possible harmony, possible orders in the color solid have to be found [21]. The simpler this order, the more clear and self-evident is the harmony. There are essentially two of these orders: those in the equivalent color circle and in the isochrome triangle. This latter statement expresses the dependence of harmony on the uniform variation of saturation and lightness. Ostwald’s theory was progressive in that it

connected the laws of color harmony with the relations between exactly measurable components. For the achromatic scale, the laws of harmony are coincident with experience. In other respects, however, experience fails to support his findings. The essential deficiency of his laws of harmony resides in his color system. Color points of his color system represent color perceptions related only by mathematically definable quantitative variations of color mixing components. The interrelation between them comprises no perceptually equal or uniformly varying intervals.

Henri Pfeiffer, a painter starting from the Bauhaus school, later a graduate from the University of Cologne who became professor at the Ecole d'Architecture in Paris, considered the creation of a color harmony system as his *chef d'oeuvre*. He started from the acoustic meaning of "harmony" – namely, that the three main tones *do* (*C*), *mi* (*E*), and *sol* (*G*) of a vibrating chord are proportional in a way that the chord length for *mi* (*E*) is the harmonic mean between the chord lengths for *do* (*C*) and *sol* (*G*). The algebraic generalization of this rule was applied for determining color harmonies. Referring to tests by Rosenstiehl [22] and Fechner [23], Pfeiffer established that lightness intervals of the logarithmic scale are in a harmonic relation and called this scale a harmonic scale. Then he deduced correlations between the logarithmic scale and the scale obtained by the golden section. In his book he described in detail his tests using Plateau and Maxwell disks to create harmonic scales. He classified his scales into two groups: isochromic or equal hue scales and isophanic or equal lightness scales. First he defined each kind of harmony and then presented its mode of construction by means of a revolving disk, followed by the analysis of the psychic character of harmony. Characteristic of his work are his chromatologic tables, with the aid of which he defined harmonizing color groups.

Harmonic scales can be developed not only according to laws of additive or subtractive color mixing but also by taking perceptually uniform intervals between colors. Colors of the Munsell color system constitute such perceptually uniform hue, saturation, and lightness scales [24, 25]. In

his book, Munsell writes that a color scheme along a "path" is always harmonious.

Color Harmony as an Aesthetic Experience

According to the most recent theories of color harmony, the principles related to the establishment of color harmony experience could be recognized only by statistical surveys based on a multitude of experiences. Their experiments are related to determined population each time. Ultimately they intended to create a system which may predict whether a given color composition will be judged by the members of a defined population as harmonic or not.

Moon and Spencer developed a model to quantify color harmony by statistical methods based on surveys [26, 27]. The main elements of Moon and Spencer's color harmony model are color intervals, area ratio, and aesthetic measurement. The aesthetical measure developed by Moon and Spencer, despite of its controversial results, provides a basic quantitative approximation of the problem in a field possessing so far only qualitative contemplations lacking verification. According to Moon and Spencer, the colors can be harmonious if the color difference is determined between the individual components. The definite color difference is defined as an interval. The magnitude of intervals defined by Moon and Spencer differs by hues. According to Moon and Spencer, a harmonic balance between color samples can be only reached if the scalar momentums related to an adaptation point in a linear space are equal to or are products of each other. The adaptation point is a point of the linear color space which corresponds to the adaptation state of the eye, normally it is medium gray.

More people have dealt with the definition of the magnitude of those intervals between members of a scale being judged as harmonic one. Dimmick [28] and Boring [29] determined the smallest of such intervals. They found that for an interval less than a certain value no harmony can any longer develop. Moon and Spencer found that these intervals were different for different hues. To examine and confirm this observation, Mori et al. [30] made experiments, and their findings agreed with those of earlier tests on lightness intervals by Katz, Gelb, and Granit [31–33].

In fine arts, intervals between colors in a picture have always been important and characteristic of a given painter. Matisse [34] has always striven for uniform intervals in his color compositions. As Mondrian [35] wrote it in 1945, “For millennia, painting expressed proportionality in terms of color relations and form relations, achieving but recently the finding of proportionality itself.”

The results of a line of experiments registering the judgments of more than 50,000 observers have been used by Nemcsics [36, 37] to create the color harmony index number system. The system denotes every existing couple of colors with a number between 0 and 100. This number expresses the magnitude of harmony of a color couple felt by different groups of people, i.e., how much is the harmony content of the color couple. The conclusions of the color harmony theory of Nemcsics are the following:

1. Between the Coloroid saturation and luminance values of color compositions felt harmonic there is an identical number or logarithmically changing number of harmony intervals.
2. The members of compositions consisting of identical hues felt harmonic are located on straight lines of the actual axis section of the Coloroid color space [38].
3. The magnitude of harmony content of color compositions located on straight lines of the Coloroid color planes spaced at identical harmony intervals is felt different depending on the angle in relation to a straight line chosen as starting line being perpendicular to the gray axis.
4. The harmony content of hue pairs is higher than that of the others if they decline from each other in the Coloroid color space less than 10° , or decline between 30 and 40° , or decline between 130 and 140° , or their declination is near to 180° .
5. Both for men and women, it is valid that the intensity of the harmony experience of a composition is in synchronism with the preference of the colors in the composition bearing harmony.

Philosophical and Historical Overview

A great deal has been written about color harmony, from Chevreul [4], who was convinced that many different color hues and their harmony could be defined by means of the relationship between numbers and his color system as an instrument to find “harmony of analogy” and “harmony of contrasts” or from the numeric naming system of Munsell (1905) [24, 25], based on the criterion of perceptual uniformity and balance, passing through Bauhaus theories (1919–1933), until sophisticated contemporary three-dimensional color softwares.

However, in philosophical terms color harmony is a highly variable concept that is open to the influence of different including individual differences (such as age, gender, personality, affective state, and so on), cultural and subcultural differences, as well as contextual, perceptual, and temporal factors. In addition, from an ontological perspective, many early theorists championed an understanding of color harmony that was universal in nature and strictly deterministic, that is, color harmony was an effect that occurred for all people on a highly predictable basis irrespective of the context or situation. In the twenty-first century, theorists tend to take a more idiographic, stochastic view about color harmony. That is, the highly subjective nature of responses to color is acknowledged, and responses about color harmony are understood to occur on a more individual and less universal basis and are more likely to be probabilistic rather than deterministic and predictable [1, 39].

The art historian and theoretician Brandi (1906–1988) [40] writes that “the significance of words is not things, but the pre-conceptual scheme of things . . . , in other words, a summary of knowledge of things according to a given society”; the same is for the taste of color, connected with the geographical, historical, and social environment. This notion, which dates back to the ancient Greek philosophers, was also expressed by Hume (1711–1776): “Beauty in things exists merely in the mind which contemplates them” [41].

Color Harmony as Historical Model

Since all that relates to matters of taste and aesthetic pleasure belongs to the cultural sphere – and is therefore conditioned by, associated with, and subject to myriad factors – the definition of chromatic harmony periodically renews its model, creating new codes for new contents.

In olden times man sought perfection of his surrounds built on symmetry, as in the numeric unit of measurement of Greek temples, whose aesthetic quality comes about through internal, independent ordering of the whole. During the Renaissance, the architect Alberti (1404–1472) [42] defined beauty as “a harmony of all the parts . . . fitted together with such proportion and connection, that nothing could be added, diminished or altered, but for the worse”.

The different societies of the world have given rise to its own chromatic spectra, applying different criteria of harmony to the juxtaposition of colors, in a dynamic scheme, related to two essential properties: mimesis and ostentation.

Color Harmony as Mimesis

As in linguistics, there is onomatopoeic harmony, consisting in imitation of natural sounds by means of a phonic impression of the linguistic form, and in the animal kingdom, some creatures take on the chromatic features of the surrounding environment in order to camouflage themselves, so too in certain cultures, manufactured articles are modeled on the colors of the materials of the place in which they are introduced, thereby creating imitative harmony. Uniformity and similarity refer to something preexisting and therefore evoke aesthetic pleasure by reason of the fact that they are pleasing and reassuring.

Color Harmony as Ostentation

And yet, on the other hand, chromatic pleasantness is sought by enriching places with missing colors, so that full-bodied, vibrant, and rich colors stand out against the uniform backgrounds of landscapes, acting, contrary to imitation, through differentiation. In both aesthetic models, mimesis and differentiation, there is harmony when there is unity in multiplicity, ordering the various parts into a coherent whole, when colors in a multi-

chromatic background establish a dynamic and balanced relationship among dimension, distribution, saturation, whiteness, and blackness.

Color Harmony in the Society of Permanent Colors

Now that chemical colors deriving from processing petrochemical synthesis can be reproduced limitlessly, there has been a revolution in perception, which initially struck the eye of the viewer but which later accustomed the viewer's eyes and mind to their fixity and unchanging nature. The environment has again taken on stable yet inert colors.

The invention of the tin tube replacing an animal's bladder, together with chemical colors, has revolutionized painting, making it more convenient *plain air* (outdoors) and allowing the experimentation of the *pointillists*, in the same way as industrial enamel paints have made possible the *dripping style* painting of Pollock (1912–1956) [43], while quick-drying acrylic colors have given rise to a certain taste for Warhol's (1928–1987) [44] and pop art's wide-ranging, flat, uniform painting style.

It can be claimed that “abstract art was born with synthetic color, which became the most important and absolute principle of abstraction” [45].

While Max Bill (1908–1984) [46] sought chromatic harmony in laws of spatial values and in laws of color movement, distributed on the basis of proportions, Hartung (1904–1989) [47] was inspired by him in his appreciation of the energy of vinyl colors, which were scratched with different instruments and sprayed with compressors.

Later, after the invention of pearlescent and iridescent paints containing flakes of mica coated in titanium oxides, American minimalist painters covered surfaces with new changeable metallic colors, freed from fixatives. Today, *acrylic* resin and silicone made possible the work of Pesce (1939) [48], where the colors blend in three dimensions.

In this concept of harmony, colors are being considered all the more beautiful when they are able to create a perceptive illusion and hide the nature of the material under a uniform permanent patina.

Color Harmony in the Society of Permutable Colors

Nowadays, digital screens are shifting vision from substance to light, prompting an adjustment toward a chromatic panorama characterized by intangibility and exchangeability.

In art, colored lights alter the perception of the physical place and introduce another space-time dimension. While Flavin (1933–1996) [49] was among the first to compose environmental works using fluorescent neon tubes, in which floors, walls, and ceilings seemed canceled by the colored light, embracing viewers in a suspended dimensions, nowadays digital media makes limitless experimentation possible. Viola (1951) [50], a veritable maestro, treats his works with a chromatic pattern making them similar to a painting, with a soft painting-like light running through them, a light that immerses the viewer in an emotional, sensorial experience taking place in an intermediate, part-natural and part-artificial space, which is both real and unreal. Toderi (1963) [51], an up-and-coming Italian video artist, transforms her urban film shooting into luminous stellar spaces, linking heaven and earth in a pulsating flow.

Color harmony nowadays is becoming dematerialized, exchanging artificial light which cancels materials, space, and time.

Conclusion

Giulio Carlo Argan wrote that “recognizing beauty is an act of justice and a sentimental act, meaning an act which confers value” [52].

In this way, as all that relates to matters of taste and aesthetic pleasure belongs to the cultural sphere – and is therefore conditioned by, associated with, and subject to myriad factors – so also the definition of chromatic harmony is related to culturally accepted notions of taste and beauty. Given the changeable nature of notions such as beauty, taste, and aesthetic pleasure, any models to predict such are open to periodic renewal, creating new models or codes for new contexts.

Cross-References

- ▶ [Anchoring Theory of Lightness](#)
- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [Color Combination](#)
- ▶ [Color Contrast](#)
- ▶ [Color Preference](#)
- ▶ [Color Scheme](#)
- ▶ [Colorant](#)
- ▶ [Colorant, Natural](#)
- ▶ [Maxwell, James Clerk](#)
- ▶ [Munsell, Albert Henry](#)
- ▶ [Natural Color Distribution](#)
- ▶ [Neon Lamp](#)
- ▶ [Newton, \(Sir\) Isaac](#)
- ▶ [Ostwald, Friedrich Wilhelm](#)
- ▶ [Pearlescence](#)
- ▶ [Pigment, Inorganic](#)
- ▶ [Unique Hues](#)
- ▶ [Visual Illusions](#)

References

1. O'Connor, Z.: Color Harmony revisited. *Color Res. Appl.* **35**(4), 267–273 (2010)
2. Rumford, T.B.: *Recherche sur la Couleur*. Paul Migne, Paris (1804)
3. Goethe, J.W.: *Zur Farbenlehre*. Cotta, Tübingen (1810)
4. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colours*. Bell and Daldy, Cambridge (1879)
5. Noon, P., et al.: Delacroix. In: *Crossing the Channel: British and French Painting in the Age of Romanticism*, p. 58. Tate Publishing, London (2003)
6. “Seurat”: *Random House Webster’s Unabridged Dictionary*, New York (2007)
7. Ferretti-Bocquillon, M., et al.: *Signac, 1863–1935*. The Metropolitan Museum of Art, New York (2001)
8. Hölzel, A.: *Lehre von der harmonischen Äquivalenz*. Longen, München (1910)
9. Kandinsky, W.: *The Art of Spiritual Harmony*. Houghton Mifflin, London (1914)
10. Klee, P.: *Pedagogisches Skizzenbuch*. Longen, München (1925)
11. Itten, J.: *The art of color*. Van Rostand Reinhold, New York (1961)
12. Albers, J.: *Interaction of Color*. Yale University (1975)
13. Moholy-Nagy, L.: *Vision in Motion*. Theobald, Chicago (1961)
14. Droste, M.: *Bauhaus 1919–1933*. (Bauhaus Archiv) - Taschen/Vince K (2003)
15. Krawkow, G.V.: *Das Farbsehen*. Akad Verlag, Berlin (1955)

16. Newton, I.: *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press (1999)
17. Hoffmann, J.L.: *Versuch einer Geschichte der malerischen Harmonie überhaupt und der Farbenharmonie insbesondere*. Hendel, Halle (1786)
18. Runge, P.O.: *Die Farbenkugel*. Nossack, Hamburg (1810)
19. Ostwald, W.: *Die Farbenlehre*. Unesma, Leipzig (1923)
20. Plochere, G.: *Plochere Color System*. Vista, Los Angeles (1948)
21. Pfeiffer, H.E.: *L'harmonie des couleurs*. Dunod, Paris (1966)
22. Rosentsthiel, C.: *Traité de la couleur*. Dunod, Paris (1934)
23. Fechner, G.T.: *Element der Psychophysik*. Breitkopf und Härtel, Leipzig (1860)
24. Munsell, A.H.: *Munsell Book of Color*. Munsell Colour Co., Baltimore (1942)
25. Munsell, A.H.: *A Grammar of Color*. Van Nostrand Reinhold, New York (1969)
26. Moon, P., Spencer, D.E.: Area in Color Harmony. *J. Opt. Soc. Am.* **34**, 93 (1944)
27. Moon, M., Spencer, D.E.: Aesthetic measure applied to color harmony. *J. Opt. Soc. Am.* **34**, 4 (1944)
28. Dimmick, F.L.: The dependence of auditory experience upon wave amplitude. *Am. J. Psychol.* **45**, 463 (1933)
29. Boring, E.G.: *History of Experimental Psychology*. Apleton-Century-Croft, New York (1950)
30. Mori, N., Nayatani, V., Tsujimoto, A., Ikeda, J., Namba, S.: An appraisal of two-colour harmony by paired comparison method. *Acta Chromatica (Tokyo)* **1**, 22 (1967)
31. Katz, D.: *The World of Colour*. Kegan Paul, Trench, Trubner, London (1935)
32. Gelb, A. (1929) *Die Farbenkonstanz der dr hendinge*. In A. Bethe, P. Bergman (eds) *Handbuch de normalen und pathologischen Physiologie*. Springer, Berlin, p. 594
33. Granit, R.: *Sensory mechanisms of the retina*. Oxford University Press, London (1947)
34. Matiss, H.: *Les notices d, une peintre*. Dounod, Paris (1908)
35. "Piet Mondrian", Tate gallery, published in Ronald Alley, *Catalogue of the Tate Gallery's Collection of Modern Art other than Works by British Artists*, London (1981)
36. Nemcsics, A.: *Colour Dynamics*. Environmental Colour Design. Ellis Horwood, New York/London/Toronto/Sydney/Tokyo/Singapore (1993)
37. Nemcsics, A.: Experimental determination of laws of color harmony. Part 1–8. *Color Res. Appl.* **32**(6), 477–488 (2007); **33**(4), 262–270 (2008); **34**(1), 33–44 (2009); **34**(3), 210–224 (2009); **36**(2), 127–139 (2011); **37**(5), 343–358 (2012), online 2013
38. Nemcsics, A.: Coloroid colour system. *Color Res. Appl.* **5**, 113–120 (1980)
39. Hard, A., Sivik, L.: A theory of colors combination – a descriptive modes related to the NCS color order system. *Color Res. Appl.* **26**(1), 4–28 (2001)
40. Brandi, C.: *Art Italian*. Art, Milano (1985)
41. Hume, D.: *Natural History of Religion*. vB.Co., Edinburgh (1952)
42. Alberti, L.B.: *De pictura*, James Leon. Bázel (1540)
43. Polloch, P.J.: *Pollock-Krasner House & Study Center*, New York (2000)
44. Warhol, A.: *The Philosophy of Andy Warhol (From A to B & Back Again)*. Harcourt Brace Jovanovich, New York (1975)
45. Brusatin, M.: *Colore senza nome*. Marsilio, Venice (2006)
46. Bill, M.: *Funktion und Funktionalismus*. Schriften 1945–1988. Benteli, Bern (2008)
47. Hartung, H., Alley, R.: *Hans Hartung*. Oxford Art Online (2001)
48. Pesce: *Design Excellence Award of the Philadelphia Museum of Art* (2005)
49. Flavin, D.: *The Complete Lights, 1961–1996 by Michael Govan and Tiffany Bell*, p. 49. Yale University Press, New Haven (2004)
50. Bill Viola: *The Eye of the Heart*. Dir. Mark Kidal [DVD]. Film for the Humanities & Sciences (2005)
51. Grazia Toderi. *Charta*, Milà (2004)
52. Argan, G.C.: *Design e Mass Media*. Op. cit. Milano (1965)

Color Image Statistics

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

Color Mixture

► [Color Combination](#)

Color Models

► [Color Order Systems](#)

Color Order Systems

Antal Nemcsics¹ and José Luis Caivano²

¹Department of Architecture, Budapest University of Technology and Economics, Budapest, Hungary

²Secretaria de Investigaciones FADU-UBA, Universidad de Buenos Aires, and Conicet, Buenos Aires, Argentina

Synonyms

Color atlases; Color models; Color solids; Color spaces

Definitions

- (a) A system for categorizing colors. An arrangement of color perceptions, color stimuli, or material color samples according to certain rules
- (b) A subset of the world of color according to three attributes that constitute the coordinates of the color system
- (c) A rational plan for ordering and specifying all object colors by a set of material standards

Color System Review and Evaluation

The aim of making order in the vast set of colors that humans are able to distinguish has existed since the ancient times and appears along the whole human history. Among the people who have proposed some kind of color order systems, there are famous philosophers, architects, artists, scientists, and writers, for instance, Aristotle (c.350 BC), Leon B. Alberti (1435), Leonardo da Vinci (1516), Isaac Newton (1704), and Johann W. Goethe (1810). In the twentieth century, other scientists and theorists, such as Albert H. Munsell (1905, 1907, 1915), Wilhelm Ostwald (1916, 1917), Arthur Pope (1922, 1929, 1949), Cándido Villalobos and Julio Villalobos-Domínguez (1947), Antal Nemcsics (1975), Harald Küppers

(1978), and Frans Gerritsen (1987), just to mention a few, have been outstanding in formulating and building color order systems. This endeavor has also been pursued by organizations such as the Commission Internationale de l'Eclairage (1931, 1976), the Optical Society of America (1947–1977), the Swedish Standards Institution (1979), and others.

A color order system seeks to include all colors in a topological model, giving a specific position to each color and proposing some kind of logic that determines the whole organization. These models have adopted, according to the different authors and along times, the most diverse shapes: linear scales, chromatic circles, color triangles, and color solids. Within this last type, different solutions have been offered: cones, pyramids, double cones, double pyramids, spheres, and more or less irregular solids.

Old Linear Organizations and Two-Dimensional Diagrams

The most ancient color organizations are often simple lists of color names, linear scales – generally expressed in verbal form, without graphic representations – or two-dimensional diagrams in the form of color triangles or circles. Among the first group, two can be mentioned: the five colors of the old Chinese philosophy (blue, red, yellow, white, and black), related to the five elements (wood, fire, earth, metal, and water) and to the five metaphysical localizations (east, south, center, west, and north), and the color scale of Aristotle (born 384 BC – died 322 BC), with white and black at the ends and a series of intermediate chromatic colors linearly arranged between these two extremes of light and darkness. A verbal ordering of colors, which some authors reconstruct as a chromatic circle or a square, and even as a double cone, a double pyramid, or a sphere [1], is found in the book by Leon Battista Alberti (born 1404 – died 1472), *On Painting*.

A great deal of the *Treatise on painting* by Leonardo da Vinci (born 1452 – died 1519), compiled after his death from his notebooks, is dedicated to insights about color. For the arrangement of colors, he proposes a simple scheme based on

the oppositions black-white, blue-yellow, and green-red.

The Belgian Jesuit Franciscus Aguilonius (born 1567 – died 1617) published his *Opticorum libri sex* in 1613 in Antwerp, and the German Jesuit Athanasius Kircher (born 1602 – died 1680) published his *Ars magna lucis et umbrae* in 1646 in Rome. Both texts include diagrams with the arrangement and mixing of colors.

Aron Sigfrid Forsius (born 1550 – died 1624), a Finnish-born clergyman, cartographer, and astronomer who worked in Sweden, wrote a manuscript in 1611 where the first-known drawing of a ► **color circle** appears. Forsius' circle presents the following sequence: white, yellow, red, brown, black, green, blue, and gray, closing again in white. Forsius characterizes this circle as “ancient” and draws an alternative for the arrangement of colors. Various theorists have interpreted this second diagram as a sphere. If this can be demonstrated, it would be the first three-dimensional system in the history of color.

Isaac Newton (born 1642 – died 1727) discriminated seven colors in the spectrum produced by the dispersion of light through a prism and arranged them in the perimeter of a circle divided into seven portions, with the center occupied by the white light. While the chromatic circle of Newton responds to the additive mixture of lights, the circle devised by Johann Wolfgang von Goethe (born 1749 – died 1832), in his *Theory of Colors* published in 1810, responds to the subtractive mixture of pigments. Three ► **primary colors** – red, yellow, and blue – appear at the vertices of an equilateral triangle, opposed to another equilateral triangle with three secondary hues – orange, green, and purple – product of the subtractive mixture of the former.

Two Important Three-Dimensional Systems Before the Twentieth Century

By 1772, the first color solid expressly developed and drawn as such by its author, Johann Heinrich Lambert (born 1728 – died 1777), appears. The pyramid has a triangle with scarlet (equivalent to the red primary), amber (equivalent to the yellow primary), and blue at its base, from whose

mixtures Lambert is able to obtain black in the center of the triangle at the base. Hence, this is an arrangement of subtractive mixture of pigments. By placing white in the upper vertex, the system is completed.

The German painter Philip Otto Runge (born 1777 – died 1810) built a color sphere that is considered the ancestor of most of the twentieth-century color systems and could be termed “modern” for this. Runge's sphere, published in 1810, is the significant result of a sustained evolutionary process initiated in the Renaissance. The chromatic circle, at the equator of the sphere, is arranged on the basis of the red-yellow-blue triad of primaries, plus the green-orange-violet triad of secondaries, as in Goethe's circle.

Classification of Color Order Systems

Color systems can be categorized in two large groups: color stimulus systems and color perception systems. Color stimulus systems are established in the twentieth and twenty-first centuries. Their developers are almost without exception physicists and information specialists. They are systematizing radiation energy of different wavelengths creating the experience of different colors. These have been registered for the industry in international standards. Their detailed description may be found in chapters of the Encyclopedia dealing with the measurement of color and computerized displaying of color.

The roots of the establishment of color perception systems appear in the distant past. Their creators are representatives of the most different professions. There are, among others, painter artists, architects, writers, poets, philosophers, doctors, priests, bishops, chemists, botanists, and many other crafts being charmed by the diversified world of colors.

The established multitude of perception-based color systems can be categorized into eight groups in terms of aims, classification methods, and definition of color patterns by color notation. Irrespective of their categorization, they possess numerous identical attributes.

Every perception-based color system possesses a color atlas containing color samples. At the beginning, colors have been located onto planes, within a triangle, square, or circle region. Later, in the seventeenth century, it turned out that the multitude of colors can be arranged only within a three-dimensional space. With the exception of two systems (Munsell, Coloroid), colors were arranged on the surface of or within regular bodies. The most frequently applied bodies were the following: a triangle-based cone (Lambert system), a triangle-based prism (Mayer system), a circle-based double cone (Ostwald system [2]), a square-based double cone (Höfler system), a circle-based truncated cone (Baumann-Prase system), a cube (Hickethier system), a globe (Runge system), a semiglobe (Chevreul system), a body of rotation consisting of cones (Wilson system), a circle-based cylinder (Frieling system), and a body constructed of cubo-octahedrons (OSA system).

The irregular shapes of the color bodies of Munsell and Coloroid color systems were the result of the fact that, as opposed to the other color systems, they have taken into account the magnitude of saturation and luminance of colors of different hues distinguishable by humans.

All perception-based color systems are discontinuous, the single exception being the Coloroid system. They contain only a definite number of colors from the color space. In most of the systems, the letter or number notating the color indicates the location of the color in the color atlas. In most of the systems, the color notation gives the additive or subtractive components for color mixing. In some systems, it defines a particular color perception. In the case of some color systems (e.g., Munsell, NCS), the color coordinates of the CIEXYZ color system have been linked to a definite number of their colors. It allows the calculation of CIE coordinates – by interpolations with different degrees of inaccuracy – also for colors not presented in the color atlas. A direct transformation relation exists only between Coloroid and CIEXYZ which enables defining with equal accuracy in both systems all colors distinguishable by the human eye, i.e., more than a million colors.

Review List of Ancient and Modern Color Order Systems

Systems of Color Stimuli

Device-Independent Systems

Schrödinger 1920, Luther 1924, Nyberg 1928, Rösch 1928, Wright 1928, Guild 1931, CIEXYZ 1931, MacAdam 1942, Stiles 1946, Brown 1951, CIE 1960, Wyszecki 1963, CIE 1964, Judd 1963, CIELAB 1976, CIELUV 1976, CIECAM 1997, CIECAM 2002.

Device-Dependent Systems

RGB 1983, HLS 1983, HSV 1983, CMY 1983, CMYK 1983, Adobe RGB 1998, sRGB, HKS, HSV.

Systems of Color Sensations

Visual Systems Based on Experience

Grosseteste 1230, Alberti 1435, Leonardo da Vinci 1516, Forsius 1611, Aguilonius 1613, Fludd 1629, Boutet 1708, Harris 1766, Schiffermüller 1772, Baumgärtner 1803, Runge 1810, Hay 1828, Merimée 1839, Field 1846, Barnard 1855, Delacroix 1856, Jannicke 1978, Henry 1889, Kreevitzer 1894, Pope 1922, Becke 1924.

Systems Based on Music Parallelism

Newton 1704, Adams 1862, Seurat 1887.

Systems Based on Additive Color Mixing

Wünsch 1792, Chevreul 1839, Grassmann 1853, Maxwell 1855, Helmholtz 1860, Bezold 1874, Vanderpoel 1902, Ridgway 1913, Ostwald 1916, Baumann-Prase 1941, Jacobson 1948, Rabkin 1950.

Systems Based on Subtractive Color Mixing

Kircher 1646, Waller 1686, Mayer 1778, Sowerby 1809, Hayter 1826, Benson 1868, Hering 1878, Rood 1879, Lovibond 1887, Lacouture 1890, Höfler 1897, Plochère 1946, Colorizer 1947, Müller 1948.

Printed Screen Systems

Wilson 1942, Villalobos 1947, Hickethier 1963, Küppers 1978.

Educational Systems Based on Psychological Effects

Goethe 1810, Schopenhauer 1830, Wundt 1874, Ebbinghaus 1902, Klee 1922, Itten 1923, Boring 1929, Birren 1934, Frieling 1968, CMM Silvestrini 1986, Gerritsen 1987, Albert-Vanel 1990.

Perceptually Equidistant Color Systems

Munsell 1905, Johansson 1937, DIN-Richter 1953, Hesselgreen 1953, OSA-Wyszecki 1960, Manfred Adam 1966, Kobayashi 1966, Coloroid-Nemcsics 1975, Acoat Color Codification, NCS (Hård, Sivik, Tonnquist) 1979, Eusemann 1979, Chromaton 1981.

Practical Color Collections

Maerz 1930, Jeanneret 1937, ISCC-Kelly 1955, Gericke-Schöne 1969, Chroma Cosmos 1987, Master Atlas 1988, Cler 1992, RAL 1993, Pantone, EuroColor 1984.

Color Systems Used Today, Recommended for Art and Design

Munsell Color System

The foundations for the most up-to-date idea of color systems were first laid down by Munsell, who published his *Book of Color* [3] and the pertinent color sample collection in 1915 for the first time. This system has been further developed in 1943 and correlated to the CIE 1931 colorimetry system. In 1956, it was extended to include very dark colors. Later, the Munsell Color Company was founded for re-editing at regular intervals, in the original quality, the color collection of the system. Since 1979, it is also published in Japan under the title Chroma Cosmos 5000.

The Munsell system is still one of the most popular color systems. Its codes are up to this day the most common color identification numbers in the international literature. Colors are identified by three data: hue H, value V (lightness),

and chroma C. These data are the three coordinates of the Munsell color solid, characterized by cylindrical coordinates and corresponding to the three characteristics of visual perception.

The ► **color circle** of the color solid, the hue scale, is divided into 100 perceptually equal parts according to 10 shades each of the following 5 reference and 5 mixed colors: red (R), yellow-red (YR), yellow (Y), green-yellow (GY), green (G), blue-green (BG), blue (B), purple-blue (PB), purple (P), and red-purple (RP) (Fig. 1).

The axis of the Munsell color solid accommodates gray (neutral, N) colors. The achromatic scale is divided in a perceptually equidistant manner from 0 to 10. The Munsell value (lightness) of a perfectly absorbing surface (ideal black) is 0, while that of a perfectly and diffusely reflecting surface (ideal white) is 10. The Munsell chroma increases with the distance from the achromatic axis (Fig. 2). In the Munsell system, colors are marked as H V/C, for example, 10RP 5/6. The shape of the Munsell color solid is shown in Fig. 3. The relationship between the Munsell color system and CIE has been tabulated.

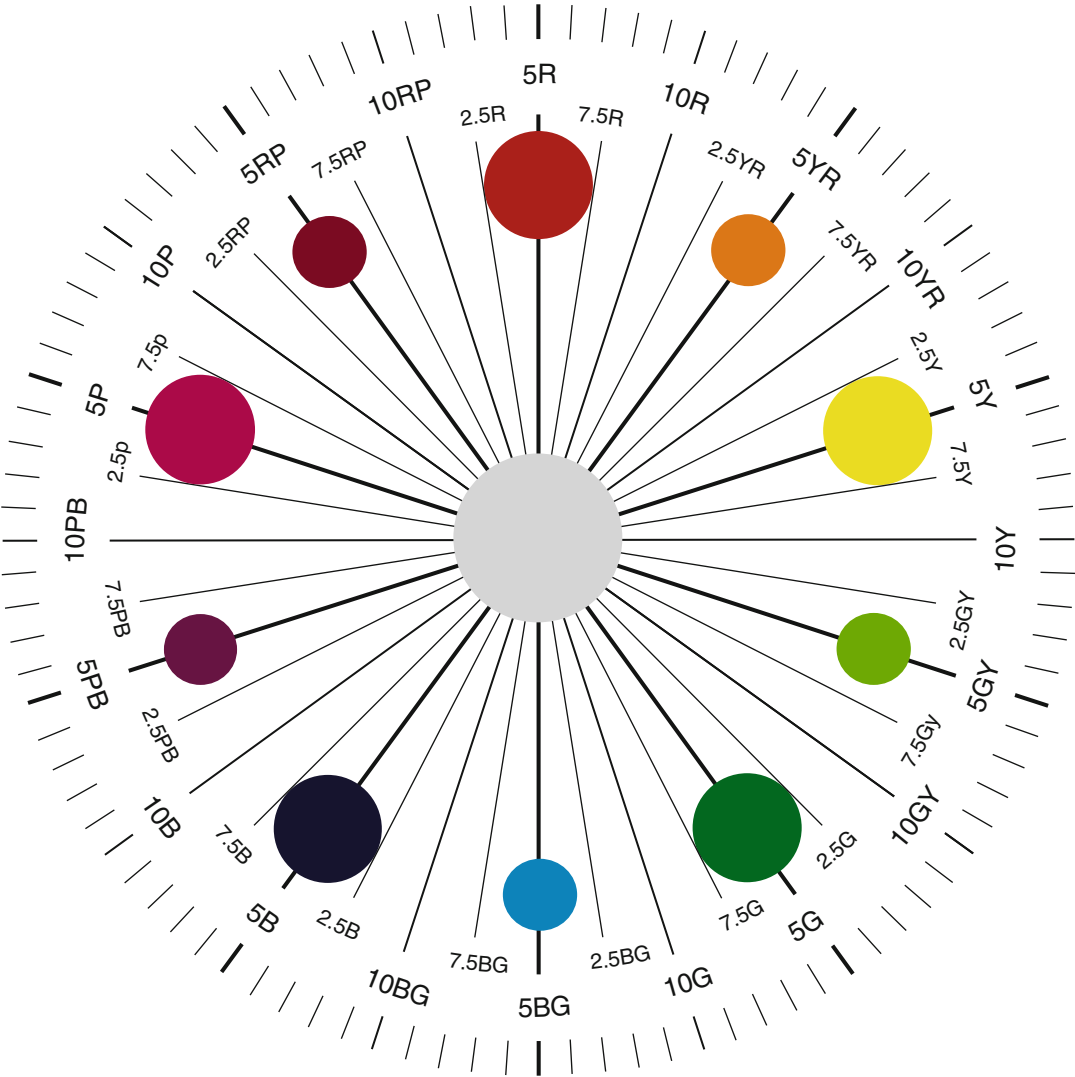
Natural Color System (NCS)

One of the latest ideas on color systematization relies on the Hering-Johansson theory, materialized as the color atlas by Hesselgren issued in 1953. Based on this atlas, in 1972 Hård, Sivik, and Tonnquist have developed the Natural Color System (NCS) adopted as a Swedish standard in 1979.

The authors of this system started from Hering's idea about six elementary color perceptions, i.e., white (W), black (S), yellow (Y), red (R), blue (B), and green (G) – all the other color perceptions being more or less related to them. In the NCS, every color is described by the degree of its similarity to the six elementary colors [4].

A color cannot be similar to more than two hues. Yellow and blue exclude each other, and so do red and green. The sum of the color variables defines the NCS chromaticness (c) of the perceived color and their ratio its hue (Φ).

In the NCS color system, colors are marked as $sc\text{-}\Phi$, for instance, 2040-G40Y, where s is blackness, c is chromaticness (a magnitude



Color Order Systems, Fig. 1 Color circle of the Munsell color system

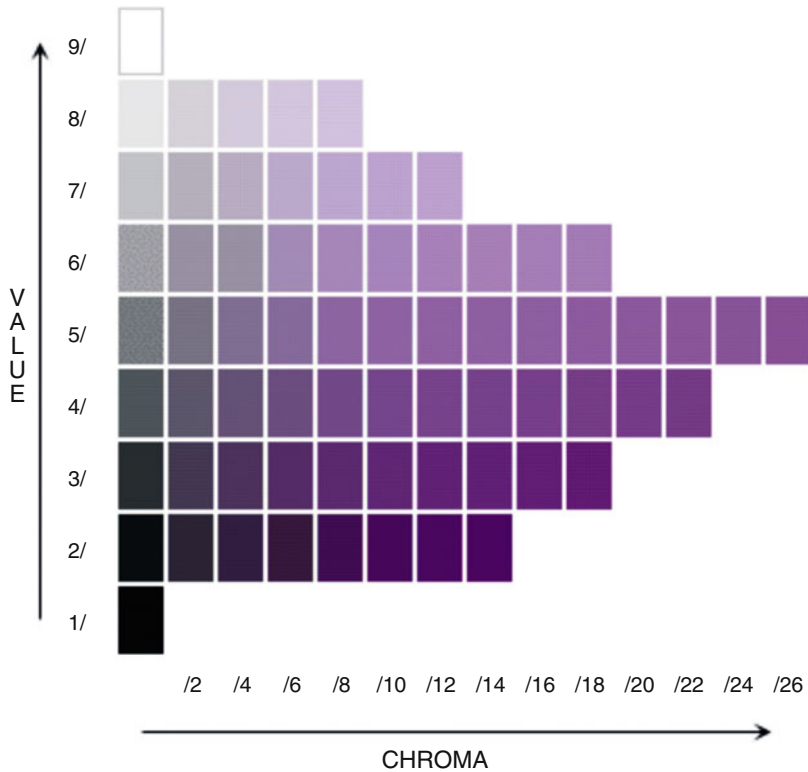
associated with saturation), and Φ is the hue of the color. In the color notation, the s and c values – both of two digits – are written without space and separated from Φ by a hyphen. The hue of the color in the example above is intermediary between green (G) and yellow (Y) in a ratio of 40 to 60.

The geometric form that the NCS assumes is a regular symmetric double cone with white and black at the vertices, while the other four preferential hues are on the circle of full colors, at corners of a square touching this circle (Figs. 4,

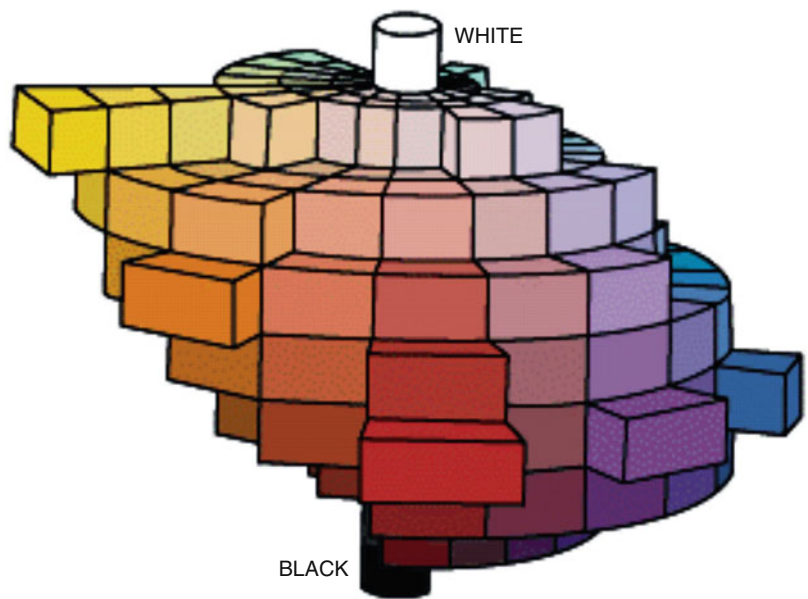
5, and 6). This color solid is of the same shape as that of the Ostwald system [2]. Also, the basic principle of some variables was adopted from the Ostwald system, but with a modified definition and scale. Similarities and differences between CIE and NCS color spaces have been tabulated, but no exact mathematical correlations were established.

The significance of this system is that it operates with a tetrachromatic color description as a possible way of visual color description.

Color Order Systems,
Fig. 2 Axial section of the
Munsell color solid, with
the Chroma and value
scales



Color Order Systems,
Fig. 3 Color solid of the
Munsell color system



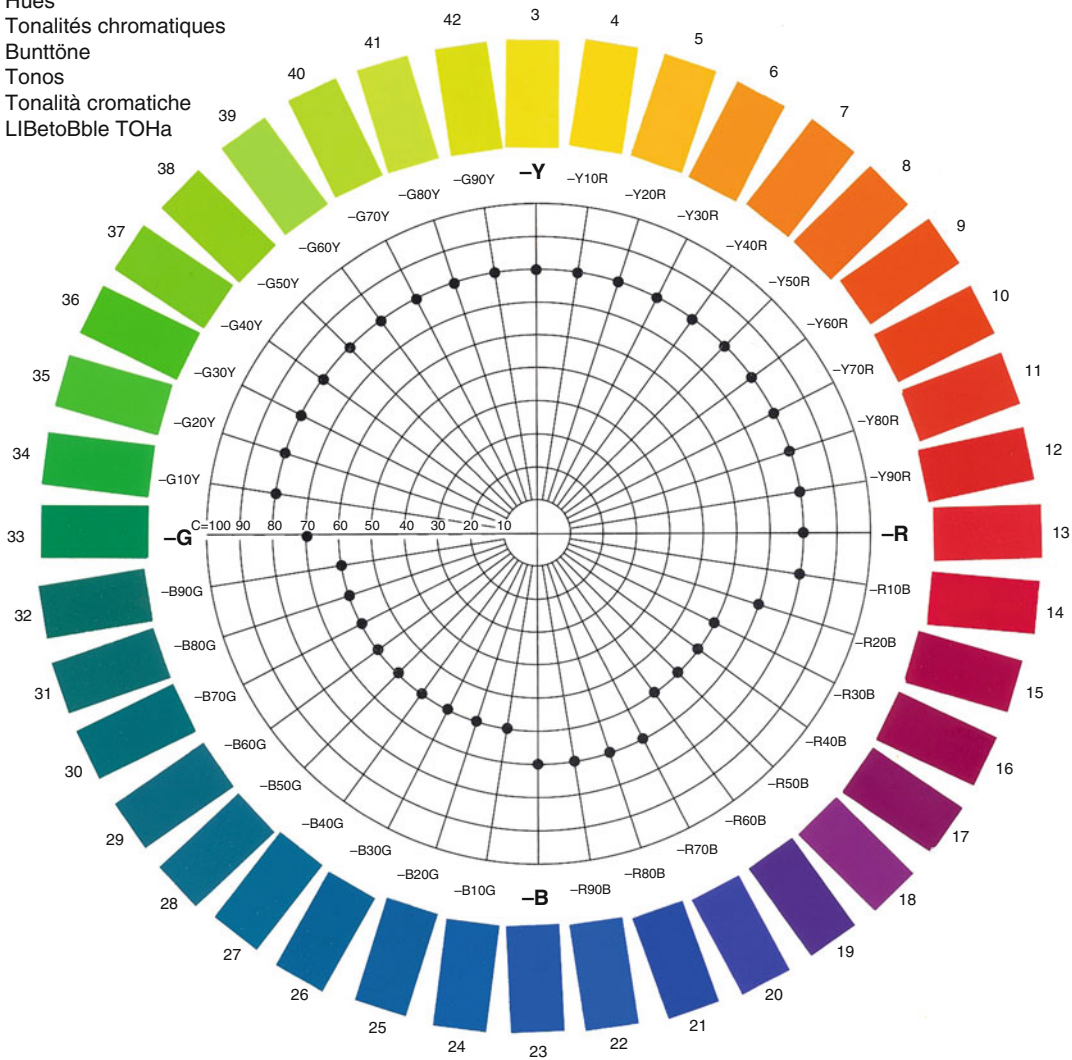
Coloroid System

Coloroid is a color system of surface colors enlightened by daylight and sensed by normal color vision observers, built on harmonic color

differences according to sensation, that well approximates the aesthetic uniformness. Coloroid is the only color system that has a direct transformation relation with the CIEXYZ system. It has



Hues
Tonalités chromatiques
Bunttöne
Tonos
Tonalità cromatiche
LIBetoBble TOHa



Color Order Systems, Fig. 4 Color circle of the NCS color system

been elaborated at the Budapest Polytechnical University by Nemcsics and was published in 1975 for the first time. From 1982 on, it has been a Hungarian technical directive, and from 2002 on, it is a Hungarian standard [5–7].

Coloroid coordinates are semipolar coordinates, representing the members of the population of colors placed inside a linear circular cylinder, to be used for the explicit definition of color points, namely, the angular coordinate representing numerically the hue of the color (A), the radial

coordinate representing numerically the saturation of the color (T), and the vertical axial coordinate representing the luminosity of the color (V).

Absolute white (W) is placed on the upper limit point of the axis of the color space. It is the color of a surface illuminated by CIE D65 beam distribution, with perfectly scattered reflection, having both Coloroid luminosity value and Yw color component value of 100.

Absolute black (S) is placed on the lower limit point of the axis of the color space. It is the color

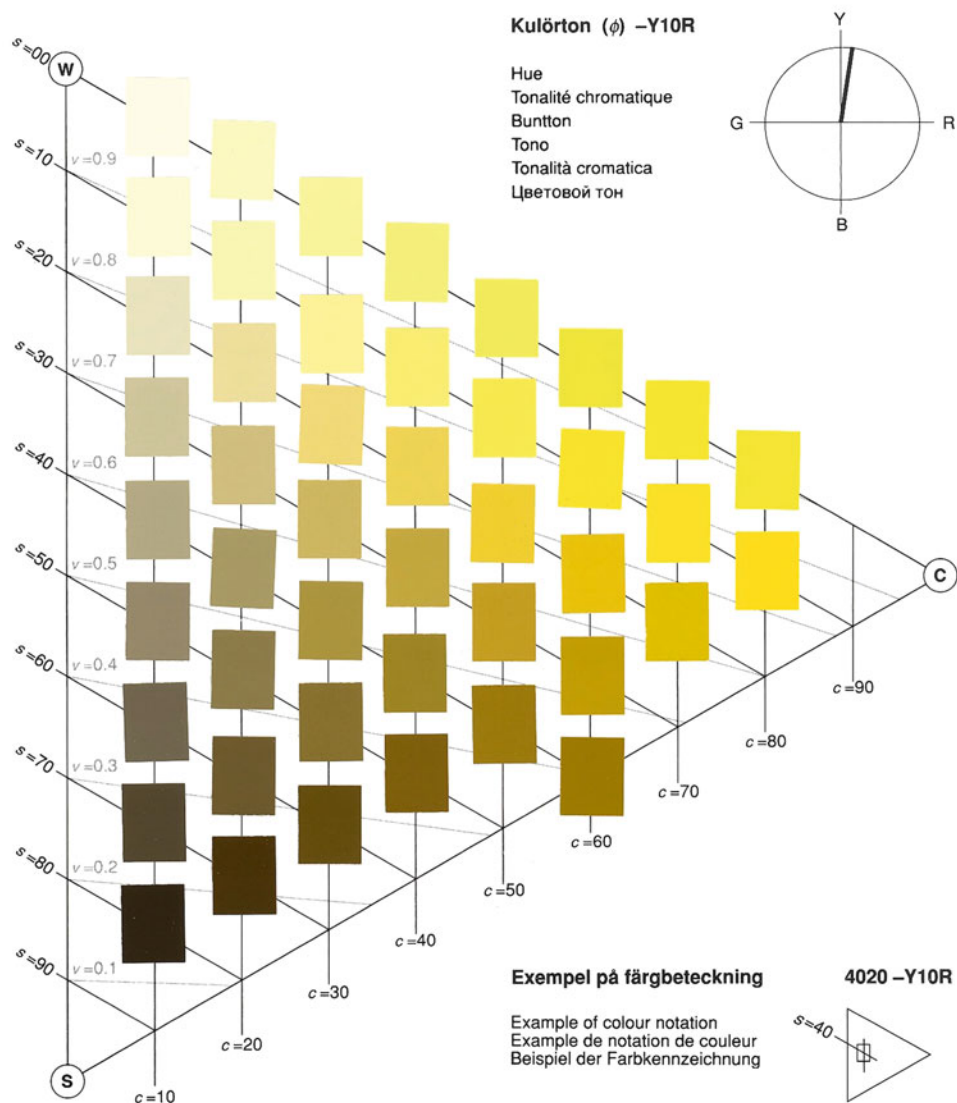
Färgatlas

Colour atlas, edition 2, sheet 4

SVENSK STANDARD SS 01 91 02

Utgåva 2

Blad 4



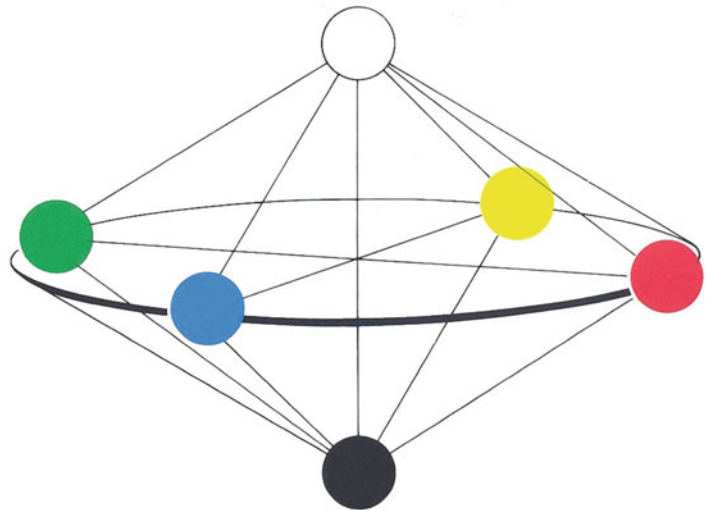
Color Order Systems, Fig. 5 Axial section of the NCS color solid

of a surface illuminated by CIE D65 beam distribution, perfectly light absorbing ($\beta = 0$ luminance factor), having both Coloroid luminosity value and Y_s color component value of 0.

The Coloroid limit colors are the most saturated colors that can be drawn onto the surface of the cylinder comprising the Coloroid space, located along a closed curve. In the CIE 1931

color diagram, colors are located along spectrum color lines between $\lambda = 450$ nm and $\lambda = 625$ nm; moreover, colors are located along the line connecting the points $\lambda = 450$ nm and $\lambda = 625$ nm.

The Coloroid basic colors are 48 different limit colors characterized with integer numbers, being located at approximately identical number of

Color Order Systems,**Fig. 6** Color solid of the NCS color system

harmony intervals to each other. The Coloroid basic colors are recorded in the CIE 1931 diagram by the jangle. The jangle is the angle of the half line originated from the D65 point of the CIE 1931 color diagram to the x axis.

The Coloroid color planes are the half planes delimited by the achromatic axis of the color space, having the same hue and dominant wavelength. In each color plane, colors are enclosed by the neutral axis and two curves, the so-called Coloroid delimiting curves. The shape of surfaces enclosed by delimiting curves is different for each hue and depends on the luminosity of the spectrum color or of the purple being located on one apex of the color plane. Along the vertical lines of the nets drawn on the color planes, Coloroid saturation values are identical; along their horizontal lines, Coloroid luminosity values are identical. Colors implemented with various means or colors created in the nature, belonging to individual color planes, are enclosed by internal delimiting curves.

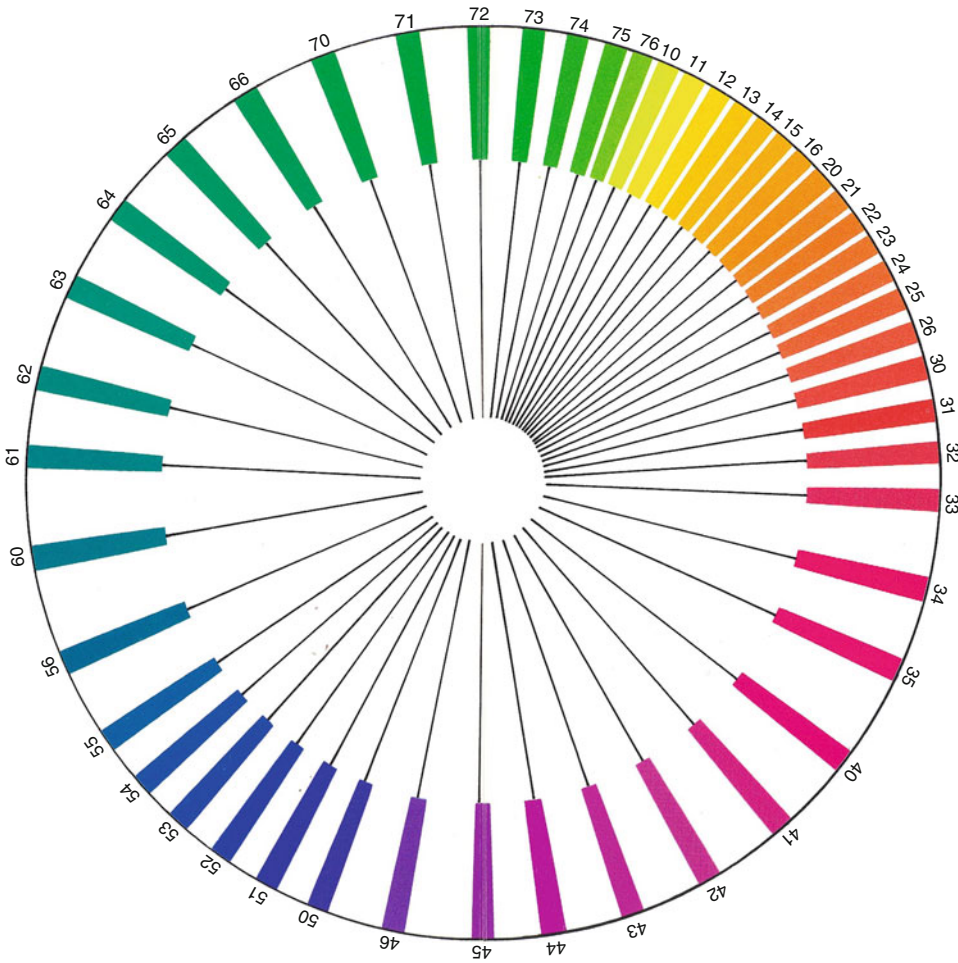
The Coloroid basic hues are the hues belonging to Coloroid basic colors. Similarly to basic colors, there are 48 Coloroid basic hues. In color planes A10, A11, A12, A13, A14, A15, A16 yellow, in color planes A20, A21, A22, A23, A24, A25, A26 orange, in color planes A30, A31, A32, A33, A34, A35 red, in color planes A40, A41, A42, A43, A44, A45, A46 purple and violet, in color planes A50, A51, A52, A53, A54, A55, A56 blue, in

color planes A60, A61, A62, A63, A64, A65, A66 cold green, in color planes A70, A71, A72, A73, A74, A75, A76 warm green hues colors exist.

Coloroid saturation is a characteristic feature of the surface color quantifying its saturation, i.e., its distance from the color of the same Coloroid achromatic luminosity measured on a scale that is aesthetically near to uniform. Its denotation is T . The saturation of the limit colors is equal to 100. The saturation of absolute white, absolute black, and gray is equal to 0. In the Coloroid space, colors of identical saturation are located equidistant to the achromatic axis, on a coaxial cylinder.

Coloroid lightness is a characteristic feature of the surface color denoting the distance measured from absolute black on an aesthetically near-uniformly graduated scale. Its denotation is V . The lightness of absolute black is equal to 0. The lightness of absolute white is 100. In the Coloroid space, colors of identical lightness are located in planes perpendicular to the achromatic axis. Numerical values of the Coloroid lightness of a surface color are determined by the expression $V = Y^{1/2}$.

Coloroid hue is a characteristic feature of the surface color denoting its hue on a scale distributed into 48 sections on an aesthetically near-uniformly graduated scale. Its denotation is A . The Coloroid hue of the surface color is a



Color Order Systems, Fig. 7 Color circle of the Coloroid color system

function of the dominant wavelength. In a Coloroid space, surface colors having identical hues lie in the Coloroid color planes.

The color notation in the Coloroid system is expressed by hue (A)–saturation (T)–lightness (V), for instance, 13-22-56 (Figs. 7, 8, and 9).

Küppers' Atlas and Rhombohedric Color System

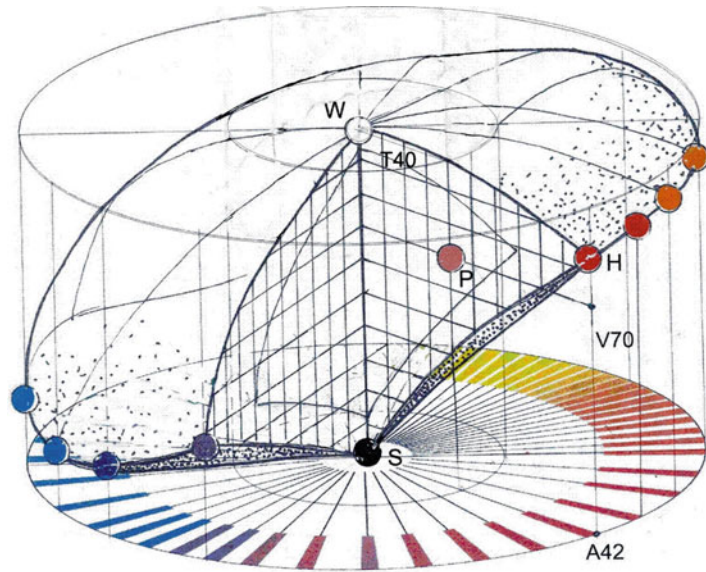
The German engineer Harald Küppers has published an atlas specifically useful for the graphic arts and the printing industry, containing more than 5,500 nuances [8]. The color samples of this atlas have been produced by the technique of four-color printing, mixing the four transparent

dyes, yellow, magenta, cyan, and black, with the concurrence of the white background of the paper.

The gradations of the mixtures are expressed in percentages directly equivalent to the proportion of surface covered by each dye, so that this notation not only is useful to designate the different nuances but also as a formula to produce the colors. The variation is made by differences of 10 % between an individual sample and the next one and between a chart of colors and the next one.

The published color charts are divided into five series; three of them are said to be of achromatic mixture because of the intervention of black, and the other two are considered of chromatic mixture

Color Order Systems,
Fig. 9 Color solid of the
 Coloroid color system



due to the exclusive intervention of yellow, magenta, and cyan. In the first three series, black is added to a mixture of two chromatic dyes – made in 10 % steps of variation, and held constant for the whole series – in 10 % steps of variation from one chart to another. In the fourth series, yellow is added in successive charts to a fixed mixture of two chromatic dyes (magenta and cyan). Series 1–4 have 11 charts, with 0 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, and 99 % of the dye being added to the fixed binary mixture. The fifth series has two additional charts: one with yellow-cyan mixtures and 99 % of magenta and the other with yellow-magenta mixtures and 99 % of cyan.

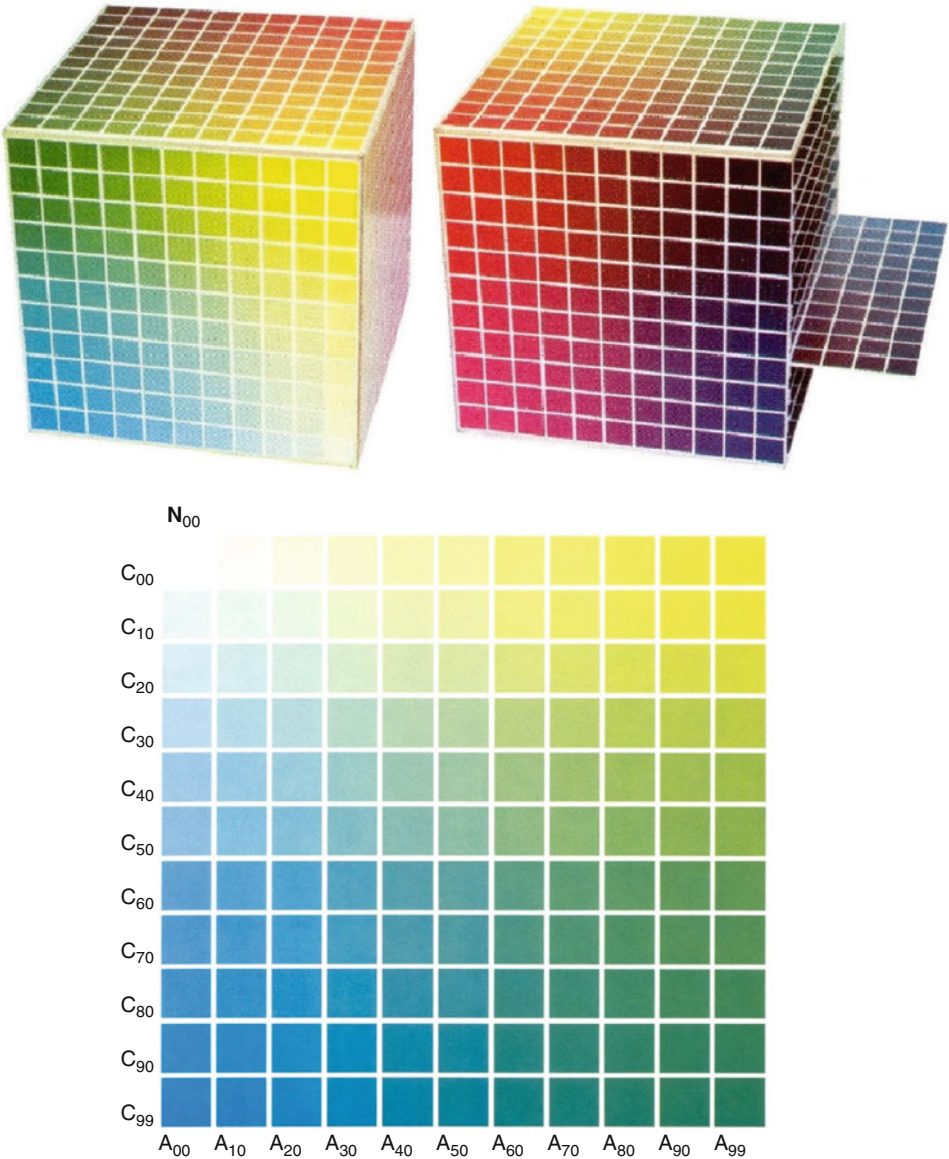
Küppers also represents his system as a three-dimensional space, either as a cube (Fig. 10) or with a rhombohedral shape, constituted by an upper tetrahedron, a central octahedron, and a lower tetrahedron (Fig. 11). In the rhombohedral model, white is placed in the upper vertex of the upper tetrahedron, in whose base the three subtractive primaries are placed: yellow, magenta, and cyan. This base is also one of the faces of the central octahedron. The lower triangular face of the octahedron holds the three colors resulting from the mixture of the subtractive primaries in pairs: green (yellow with cyan), red (yellow with magenta), and blue (cyan with magenta). This

face is, in turn, shared with the lower tetrahedron, in whose lower vertex black is placed.

Additional Studies, Classifications, and Evaluations of Color Order Systems

The bibliography about historical, comparative, or classificatory analysis of color order systems is extensive. In addition to the already stated references, it is possible to mention articles by Spillmann [9], Robertson [10], Sivik [11], and Tonnquist [12]. The International Color Association devoted a whole congress to the theme of color order systems in 1983 [13] and had a study group on this subject from 1978 to 1990, chaired successively by Günter Wyszecki, Fred Billmeyer, James Bartleson, and Nick Hale (see [14]).

Specially interesting is the analysis made by Billmeyer [15], who deals with the history and the principles of various systems, establishes comparisons and differences, and offers data about the attempts of conversion among the notations of various systems. William Hale [16] discusses the uses of color order systems, including those that have a relevant theoretical or scientific basis as well as those more pragmatic and with a specific commercial purpose. Tonnquist [17] offers another criterion to classify color order systems.



Color Order Systems, Fig. 10 Harald Küppers, the model of his atlas in a cubic shape, and a page from it

He divides them into (a) *physical*, those that have actual samples as reference and that only exist in terms of an atlas or a color chart; (b) *psychophysical*, those whose definition is given by means of valences for a set of points in a color space, and whose definition is only valid to a combination of observer, illuminant, and measuring instrument, as, for instance, the CIE system; and (c) *perceptual*, those defined in terms of

elementary color percepts, mental references of certain basic colors which serve to describe all the remaining colors, as, for instance, the NCS system.

Some countries have adopted certain color systems as national standards. However, no color system is favored with the acceptance as an international standard. During the 6th Congress of the International Color Association, a specific



Color Order Systems, Fig. 11 Küppers' rhombohedral model

discussion about the question if any particular system is better than the others took place ([18], vol. I, pp. 163–172). The conclusion was that there is no particular system that is *the* best for all the applications, which may cover fields so different as color education, artistic practice in various disciplines, professional practice in diverse branches of design (architectural, graphic, industrial, textile, landscape design), specification of color in materials so dissimilar as paper, clothes, leather, plastics, metals or in complex industries such as the automotive, the food industry, color reproduction in TV, video and computer displays; some systems are more useful than others for certain specific problems. In the 7th Congress of the AIC, a round table on color order systems was also organized ([19], vol. A, pp. 173–174) in which there was a consensus on some points, among them (a) that physical color samples, although they are useful, are not essential to a color system and (b) that broadly, the systems can be divided into two types: systems of color appearance (a typical example would be the NCS) and systems of color stimuli (an example would be the CIE system).

The interested reader can look for additional information on description, analysis, and comparison of color order systems in Caivano [20], Stromer and Baumann [21], Silvestrini, Fischer and Stromer [22], Stromer [23], Kuehni [24], and Spillmann [25], among other sources. There is also an excellent website devoted to this subject: www.colorsystem.com.

Cross-References

- Chevreul, Michel-Eugène
- CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments
- Color Circle
- Color Contrast
- Color Harmony
- Complementary Colors
- Itten, Johannes
- Munsell, Albert Henry
- Ostwald, Friedrich Wilhelm
- Philosophy of Color
- Primary Colors
- Richter, Manfred
- Runge, Philipp Otto
- Unique Hues

References

1. Parkhurst, C., Feller, R.L.: Who invented the color wheel? *Color. Res. Appl.* **7**(3), 217–230 (1982)
2. Ostwald, W.: *Die Farbenlehre*. Unesma, Leipzig (1923)
3. Munsell, A.H.: *Munsell Book of Color*. Munsell Color Co., Baltimore (1942)
4. Hård, A., Sivik, L.: A theory of colors in combination: a descriptive model related to the NCS color order system. *Color. Res. Appl.* **26**(1), 4–28 (2001)
5. Nemcsics, A.: Coloroid colour system. *Color. Res. Appl.* **5**(2), 113–120 (1980)
6. Nemcsics, A.: Color space of the Coloroid color system. *Color. Res. Appl.* **12**(3), 135–146 (1987)
7. Nemcsics, A.: *Colour Dynamics*. Environmental Colour Design. Ellis Horwood, New York (1993)
8. Küppers, H.: *Das Grundgesetz der Farbenlehre*. DuMont, Cologne (1978) (English transl. *The Basic Law of Color Theory*. Barron's Educational Series, Woodbury, New York) (1982)
9. Spillmann, W.: Color order systems and architectural color design. *Color. Res. Appl.* **10**(1), 5–11 (1985)
10. Robertson, A.R.: Principles of colour order systems. In: *AIC Colour 93, Proceedings of the 7th Congress*, vol. A, pp. 149–153. Hungarian National Colour Committee, Budapest (1993)
11. Sivik, L.: Systems for descriptive colour notations – implications of definitions and methodology. In: *AIC Colour 93, Proceedings of the 7th Congress*, vol. A, pp. 89–94. Hungarian National Colour Committee, Budapest (1993)
12. Tonnquist, G.: 25 years of colour with the AIC -and 25000 without. *Color. Res. Appl.* **18**(5), 353–365 (1993)
13. AIC (Association Internationale de la Couleur): *The Forsius Symposium on Colour Order Systems*, 2 vols. Scandinavian Colour Institute, Stockholm, Colour Reports F26 and F28 (1983). Also available in http://www.aic-color.org/congr_archivos/aic1983procF26.pdf and http://www.aic-color.org/congr_archivos/aic1983procF28.pdf
14. Billmeyer Jr., F.W.: *AIC Annotated Bibliography on Color Order Systems*. Mimeoform Services, Beltsville (1987)
15. Billmeyer Jr., F.W.: Survey of color order systems. *Color. Res. Appl.* **12**(4), 173–186 (1987)
16. Hale, W.N.: Color order systems and color notations. In: *AIC Color 89, Proceedings of the 6th Congress*, vol. 1, pp. 43–51. Grupo Argentino del Color, Buenos Aires (1989)
17. Tonnquist, G.: Colour order systems and colour atlases. In: *AIC Color 89, Proceedings of the 6th Congress*, vol. 2, pp. 162–165. Grupo Argentino del Color, Buenos Aires (1989)
18. AIC (Association Internationale de la Couleur): *AIC Color 89, Proceedings of the 6th Congress*, 2 vols. Grupo Argentino del Color, Buenos Aires (1989)
19. AIC (Association Internationale de la Couleur): *AIC Colour 93, Proceedings of the 7th Congress*, 3 vols. Hungarian National Colour Committee, Budapest (1993)
20. Caivano, J.L.: *Sistemas de orden del color*. Facultad de Arquitectura, Diseño y Urbanismo, UBA, Buenos Aires (1995). Also available in <http://www.fadu.uba.ar/sitios/sicyt/color/1995scol.pdf>
21. Stromer, K., Baumann, U.: *Color Systems in Art and Science*. Regenbogen, Konstanz (1996)
22. Silvestrini, N., Fischer, E.P., Stromer, K.: *Farbsysteme in Kunst und Wissenschaft*. DuMont, Cologne (1998)
23. Stromer, K.: *Farbsysteme*. DuMont, Cologne (2002)
24. Kuehni, R.G.: *Color Space and its Divisions: Color Order from Antiquity to the Present*. Wiley, New York (2003)
25. Spillmann, W. (ed.): *Farb-Systeme 1611–2007: Farb-Dokumente in der Sammlung Werner Spillmann*. Schwabe, Basel (2009)

Color Palette

► Color Scheme

Color Perception and Environmentally Based Impairments

Galina V. Paramei

Department of Psychology, Liverpool Hope University, Liverpool, UK

Synonyms

[Acquired color vision impairment](#); [Acquired color vision loss](#); [Dyschromatopsia](#)

Definition

Decreased discrimination of colors caused by adverse environment, such as long-term occupational exposure to or consumption of drugs, substances, and food containing neurotoxic chemicals.

Color vision early manifests adverse effects of exposure to an environment that contains

neurotoxic substances [1, 2]. The acquired color vision impairments, or dyschromatopsias, can be very subtle (subclinical) but also may vary considerably in severity, increasing or decreasing as long as the responsible agent persists, and can become irreversible under long-term exposure and/or agent dose.

There are several scenarios of exposure to hazardous chemical agents in the environment:

- i. Long-term occupational exposure to certain substances (e.g., neurotoxic metals, organic solvents, carbon disulfide, etc.)
- ii. Self-administered chronic consumption of substances containing neurotoxic chemicals (e.g., alcohol, tobacco)
- iii. Side effects from pharmacological treatment of medical conditions (e.g., cardiovascular, antiepileptic, or antituberculosis drugs)
- iv. Consumption of food contaminated by neurotoxic elements through the food chain (e.g., mercury)

General Characteristics of Neurotoxin-Induced Color Vision Impairments

Acquired color vision defects, unlike congenital ones, are noticeable to the observer: recently affected subjects name the stimuli as they see them – in contrast to subjects with congenital color vision defects where there is compensatory adaption of their color naming to that of normal trichromats. Acquired dyschromatopsia may not be identical in the two eyes, which requires testing the two eyes separately.

Neurotoxic substances can affect one or more loci in the color vision system. At the pre-receptor level, hazardous chemicals can accelerate yellowing of the crystalline lens which results in an increase in absorption of blue light and hence decreased discrimination of blue colors. In the retina, the main mechanism of color vision loss is selective damage to specific photoreceptor classes, short-wavelength (S-), middle-wavelength (M-), or long-wavelength (L-) cones [► [Cone](#)

[Fundamentals](#)]. Most vulnerable among these are the S-cones, damage of which is manifested by blue color vision defects. Post-receptoral processing can also be disrupted – at the level of ganglion cells, optic nerve, optic radiation, or visual cortex – causing color vision impairment. Often the damage is nonselective; i.e., patterns of color discrimination loss are not always specific to one of the color subsystems and differ from those in congenital abnormalities.

Classification of Acquired Dyschromatopsias

The wide variation in acquired color vision defects, according to Verriest [3], can be classified in four major types, I, II, and III and a nonspecific defect. The first two types are associated with impaired color discrimination along the red-green axis in perceptual color space [Cross-Ref. Bimler], much like the patterns found in congenital red-green deficiency, i.e., both involve mild to severe confusion of reds and greens [Cross-Ref. Bonnardel]. Type I is protan-like and reveals little or no loss of blue-yellow discrimination; type II is deutan-like and is manifested by concomitant mild loss of discrimination between blues and yellows. Type III, tritan-like, is manifested by mild to moderate blue-green and yellow-violet confusions, with a lesser or absent loss of red-green discrimination.

According to Köllner's rule [1], impairment of blue-yellow discrimination – the range most frequently affected by exposure to hazardous chemicals – suggests toxic retinopathy, i.e., a more external retinal dysfunction; by contrast, a preponderance of red-green loss is associated with pathology in the optic nerve; finally, complex color vision loss, blue-yellow and red-green, suggests a more advanced stage with a damage to both the retina and the neuro-optic pathway. However, numerous exceptions to Köllner's rule instruct one to be cautious about making a clear-cut attribution of blue-yellow loss to damage at a

retinal level and of red-green loss to damage at a neural level. Both color systems appear to be selectively susceptible to damage by various types of neurotoxins.

Tests of Color Vision for Assessing Acquired Dyschromatopsias

In epidemiological studies, color arrangement tests are predominantly used. These can be rapidly administered and easily interpreted, and in addition, they allow color vision ability to be quantified graphically [4]. In an arrangement test, the observer is presented with a set of color caps and requested to arrange them in (“rainbow”) sequence. The number of erroneous cap transpositions provides a measure of overall color discrimination; the pattern of the transpositions indicates whether the defect is closer to the blue-yellow or red-green axes or with no discernible pattern [4, 5] (Fig. 1).

Three tests, whose caps sample a color circle at even intervals, are traditionally used for the purpose: the Farnsworth-Munsell 100-hue test, the Farnsworth Dichotomous panel D-15 test, and the Lanthony Desaturated Panel D-15d test. The FM 100-hue test consists of 85 caps and takes 20–30 min to perform; it is designed so that error scores will be concentrated in the region of the poorest discrimination. The D-15 test contains a sample of the latter, including a fixed cap and 15 movable ones; it requires ca. 5 min to complete and is designed to diagnose moderate to severe color defects. The Lanthony D-15d test is similar in design and identical to the D-15 test in administration but consists of color samples that are lighter and paler [5]. The D-15d test was designed specifically to capture mild or subclinical color defects in observers who pass the standard D-15 test. The two tests are often used in conjunction, though the more sensitive D-15d is widely employed for early detection of mild neurotoxin-induced dyschromatopsias. Outcomes of both tests are reported via a Color Confusion Index (CCI), where 1.0 corresponds to color perfect

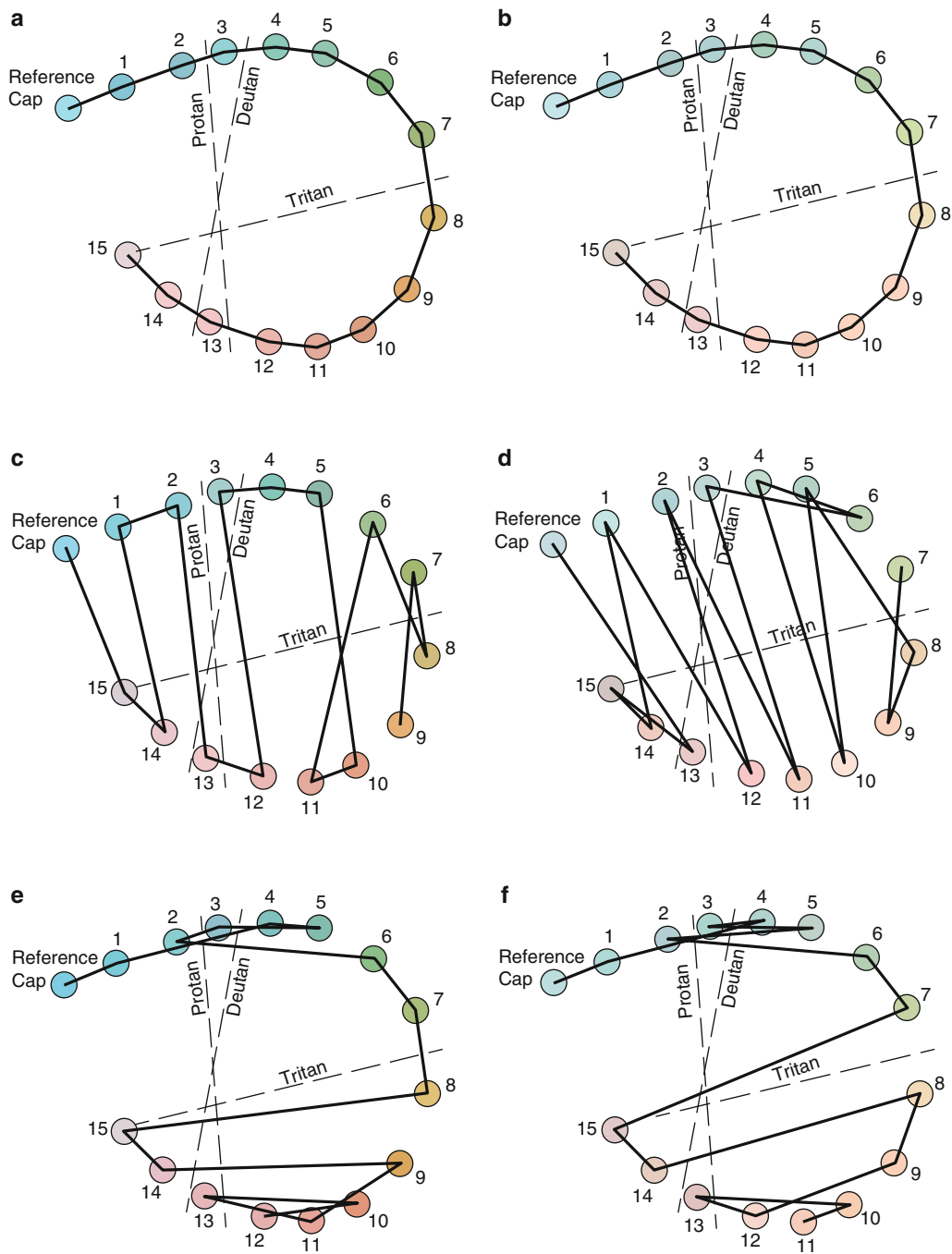
arrangement; CCI values greater than 1.0 indicate progressive impairment of color discrimination [4, 5].

Occupational Exposure to Neurotoxic Substances

A number of occupations involve exposure to volatile neurotoxic substances (e.g., printers, aircraft maintenance workers, dry cleaners in automotive and metalworking industries, viscose rayon workers, microelectronics assembly workers, gold miners, dentists, etc.). Such substances include organic solvents (toluene, styrene, benzene, perchlorethylene, *n*-hexane, carbon disulfide), solvent mixtures, and metals like mercury (in its elemental or methyl forms). Even when neurotoxic substances are applied within the occupational limits, long-term exposure has been shown to result in mild impairment of color discrimination [6–8].

Using the D-15d test, the degree and pattern of color vision loss was intensively investigated with regard to exposure to organic solvents, in particular, toluene and styrene [2, 6–9], and to mercury [10]. The main finding across these studies is significant increase of the CCI in the occupationally exposed observers (compared to age-matched controls). For example, in a meta-analysis of 15 sample studies of the effects of toluene, styrene, and solvent mixtures [8], the grand mean CCI for the exposed workers was 1.22 ± 0.08 , significantly greater than 1.13 ± 0.06 for the controls ($p = 0.003$). Similarly in [10], for workers of fluorescent lamp production exposed to mercury vapor $CCI = 1.14 \pm 0.14$ was significantly greater than 1.04 ± 0.06 for controls ($p = 0.002$). The impairment of color discrimination was shown to be subject to cumulative exposure, i.e., product of duration and current level of exposure [2, 6–9], and may become irreversible even when the hazardous agent is withdrawn [10].

Neurotoxin-induced dyschromatopsias are predominantly of type III, i.e., tritan-like pattern. Less common are types I and II, red-green



Color Perception and Environmentally Based Impairments, Fig. 1 Scoring sheet for the D-15 (a) and the D-15d (b); for illustrative purposes, the numbers are accompanied by colors simulating those of the original test caps. At perfect color arrangement of a normal trichromat, lines connect the “reference cap” through 1–15; CCI = 1.0. (c–d) Lines are drawn between consecutive

caps as placed by a protanope, observer with a congenital red-green deficiency; (e) CCI = 2.48; (f) CCI = 3.11. (e, f) Mild acquired tritan type of color (blue-yellow) discrimination impairment; (e) CCI = 2.06; (f) CCI = 2.19. Note that the low-saturated stimuli of the D-15d result in a more prominent color confusion (d, f)

dyschromatopsias. In comparison, the nonspecific type of dyschromatopsia, which implies difficulty in discriminating colors along both the red-green and blue-yellow axes of color space, is also highly prevalent.

Self-Administered Consumption of Substances Containing Neurotoxic Chemicals

Chronic excessive consumption of alcohol (ethanol) affects color discrimination capacity [5, 6]. When assessed by the D-15d test, the prevalence of dyschromatopsia was shown to increase with alcohol intake. Further, heavy drinkers (with an intake larger than 750 g/week) manifested primarily loss of blue-yellow discrimination, whether or not they were undergoing treatment in a detoxification center. However, 25 % of persons undergoing detoxification revealed dyschromatopsia of the nonspecific type, including red-green loss [11].

Tobacco smoke contains a range of compounds including nicotine, cyanide, and carbon monoxide and, when consumed excessively, can affect color vision [1]. When tested by means of the color arrangement tests, chronic smokers (>20 cigarettes/day, for at least a year) revealed a subtle but statistically significant reduction in sensitivity to red-green differences compared to nonsmokers [12] or showed a diffuse character of color vision disturbance, without a particular dyschromatopsia axis [13].

Side Effects from Pharmacological Treatments

A number of medications (e.g., cardiovascular, antiepileptic, antituberculosis, and antirheumatism drugs, oral contraceptives, etc.) are known to produce measurable color vision disturbances, some of which affect color vision even at therapeutic levels, most commonly as type III, tritan-like blue-yellow dyschromatopsia [1, 5, 6, 14]. For instance, patients taking antiepileptic drugs develop mild blue-yellow deficiency which may show signs of progression with lasting intake

of the drugs. Intake of Viagra was shown to cause transient adverse visual events described as a blue color tinge to vision, accompanied by mild blue-yellow deficiency in about 11–14 % of those taking the medication, the disturbance being reversible after the medication has been discontinued. Treatment by the tuberculostatic ethambutol shows mild blue-yellow deficiency as the earliest sign of the drug's neurotoxicity, but this can also develop as type II, deutan-like red-green, or a nonspecific dyschromatopsia; these defects are transient and reversible after stopping the therapy.

Consumption of Food Contaminated by Neurotoxic Elements

Certain industrial activities, like gold mining or mercury mining, are associated with pollution of mercury which bioaccumulates mainly through the aquatic food chain (seafood and fish). Even at low levels of dietary exposure, using the FM 100-hue test, and the Cambridge Colour Test (► [Paramei & Bimler, Deuteranopia](#)), mercury was shown to chronically reduce color discrimination, with the error pattern indicating that both blue-yellow and red-green systems are affected [15].

Cross-References

- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Cone Fundamentals](#)
- [Deuteranopia](#)
- [Psychological Color Space and Color Terms](#)

References

1. Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L. G. (eds.): *Congenital and Acquired Color Vision Defects*. Grune and Stratton, New York (1979)
2. Gobba, F.: Color vision: a sensitive indicator of exposure to neurotoxins. *Neurotoxicology* **21**, 857–852 (2000)
3. François, J., Verriest, G.: On acquired deficiency of colour vision, with special reference to its detection and classification by means of the tests of Farnsworth. *Vision Res.* **1**, 201–219 (1961)

4. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
5. Geller, A., Hudnell, H.K.: Critical issues in the use and analysis of the Lanthony desaturate color vision test. *Neurotoxicol. Teratol.* **19**, 455–465 (1997)
6. Iregren, A., Andersson, M., Nylén, P.: Color vision and occupational chemical exposures: I. An overview of tests and effects. *Neurotoxicology* **23**, 719–733 (2002)
7. Gobba, F., Cavalleri, A.: Color vision impairment in workers exposed to neurotoxic chemicals. *Neurotoxicology* **24**, 693–702 (2003)
8. Paramei, G.V., Meyer-Baron, M., Seeber, A.: Impairments of colour vision induced by organic solvents: a meta-analysis study. *Neurotoxicology* **25**, 803–816 (2004)
9. Benignus, V.A., Geller, A.M., Boyes, W.K., Bushnell, P.J.: Human neurobehavioral effects of long-term exposure to styrene: a meta-analysis. *Environ. Health Perspect.* **113**, 532–538 (2005)
10. Feitosa-Santana, C., Costa, M.F., Lago, M., Ventura, D.F.: Long-term loss of color vision after exposure to mercury vapor. *Braz. J. Med. Biol. Res.* **40**, 409–414 (2007)
11. Mergler, D., Blain, L., Lemaire, J., Lalande, F.: Colour vision impairment and alcohol consumption. *Neurotoxicol. Teratol.* **10**, 255–260 (1988)
12. Bimler, D., Kirkland, J.: Multidimensional scaling of D15 caps: color-vision defects among tobacco smokers? *Vis. Neurosci.* **21**, 445–448 (2004)
13. Erb, C., Nicaeus, T., Adler, M., Isensee, J., Zrenner, E., Thiel, H.-J.: Colour vision disturbances in chronic smokers. *Graefes Arch. Clin. Exp. Ophthalmol.* **237**, 377–380 (1999)
14. Zrenner, E., Hart, W.: Drug-induced and toxic disorders in neuro-ophthalmology. In: Schiefer, U., Wilhelm, H., Hart, W. (eds.) *Clinical Neuro-Ophthalmology: A Practical Guide*, pp. 223–229. Springer, Berlin (2007)
15. Silveira, L.C.L., Damin, E.T.B., Pinheiro, M.C.C., Rodrigues, A.R., Moura, A.L.A., Cortes, I.M.I.T., Mello, G.A.: Visual dysfunction following mercury exposure by breathing mercury vapour or by eating mercury-contaminated food. In: Mollon, J.D., Pokorny, J., Knoblauch, K. (eds.) *Normal & Defective Colour Vision*, pp. 409–417. Oxford University Press, Oxford (2003)

Color Phenomenology

Don Dedrick

Department of Philosophy and Department of Psychology, University of Guelph, Guelph, ON, Canada

Synonyms

Color experience; Color qualia

Definition

The word “phenomenology” finds its original application in philosophy, and it has two distinct meanings in that discipline. In the first, most substantive meaning it refers to a philosophical tradition originating in the work of G. W. F. Hegel and developed in the work of Edmund Husserl, Martin Heidegger, Jean-Paul Sartre, and others, with the psychologist Franz Brentano as a major influence. In this primordial sense, it refers to a nonpsychological description of the fundamental constituents of experience. It may sound peculiar to call an account of experience “nonpsychological” since experience might be thought of as necessarily psychological. The rationale for the usage is this: one may possibly describe the fundamental constituents of human experience – concepts, ideas, propositions, temporality, mental images, etc. – in a way that captures their generic character and hence their “universality” rather than their specific contents. Such a view articulates, so classical phenomenologists say, the logical or conceptual structure of experience. The second application of the term, the origin of which is equally philosophical and psychological, refers more directly to the *mere* appearance of things. There is no universality attached to such description. It is, rather, a reference to the way things *seem* to perceivers: red looks this way (and perhaps just to an individual), pain feels this way (ditto), dogs bark in a way that sounds as it sounds (again, perhaps just to the individual). This sense of the word “phenomenology” (the word “qualia” is sometimes used) describes the way that many Anglo-American philosophers have deployed the term throughout much of the twentieth century, into the twenty-first. This second sense has clear links to the way experimental psychologists use and have talked about phenomenology, and it is this sense that is of interest here.

Color Phenomenology and Ontology

If the phenomenology of X concerns the way that some X *seems*, is a contrast to the way things *are* implied? The answer is yes. Consider the following example (discussion will now concern examples

and issues specific to color phenomenology). As one moves a color stimulus from one illumination source to another, one will notice that it appears to change color, though likely within the bounds of color constancy. This may lead to a question as to its “real color,” and that question may lead in a number of directions. Perhaps its real (object) color is identical with a physical property that does not change across illuminations: surface spectral reflectance, for example. Perhaps the idea of real color should be abandoned and replaced with a conception of color relativized to viewing conditions and observers. Perhaps one could speak of a “normal observer” (as the CIE does) so as to wring some objectivity out of the phenomenology. These positions take up different views as to the relation between the way things seem and the way things are for color, and there is a spirited, scientifically informed philosophical literature that covers the many permutations of these views (For a representative sampling, see Ref. [1]).

Epistemological Issues

Questions about real colors are questions about ontology. Does the catalogue of the real include colors? What is a real color, if there is such? As important as these questions are, much of the interest when it comes to color phenomenology is not ontological but epistemological: what can be known, and what are the limitations to what can be known, about color experience? The most famous query along this line comes from John Locke. In the *Enquiry Concerning Human Understanding* (2), Locke proposes that a “spectral inversion” would not be detectable. Subject A and subject B, indistinguishable in terms of their behavior on discrimination tasks, nonetheless have different experiences. A’s experiential color space is “inverted” relative to B’s. For Locke, this meant that A and B use color words the same way, describe their color experiences the same way, discriminate color stimuli in the same way, and yet have distinct color experiences, e.g., A’s green is B’s red, and vice versa, and the same goes for blue and yellow. Thus, the inversion is behaviorally undetectable. Many consequences

have been thought to flow from Locke’s proposal, but vision science provides good reason to believe that even if such inversion was possible, it would be detectable. The inverted spectrum proposal depends on a color space that is symmetrical so that one can map discriminable differences one-to-one from, say, the “greens” to the “reds.” This condition is not satisfied for a standard human trichromatic color space, as specified, for instance, in the asymmetrical CIE $L^*a^*b^*$ space. The upshot, in terms of behavior, is that A and B would behave differently – confusing or discriminating different color stimuli. The inversion, with the cleverness of psychophysics, would be detected (see Ref. [2]).

Despite the fact that Locke’s proposal fails, its implications are not easily dispatched. Even if spectral inversion can be detected, the question remains as to what color experience is like. Consider the following thought experiment. What is to be learned from the psychophysics of color? This is a broad question, but broadly the answer would be that one learns correlations between different types of stimuli and different types of behavior. From these correlations, serving as constraints, ideas as to what the properties of the neural substrate of color experience need to be like may be formulated, models constructed, convergence with other areas of physiology and psychology sought. Psychophysics (visual being the concern here) is a mature subdiscipline of psychology, but does it deal with the way things – colored things – look? This sounds an odd question for a science that is based on subjects looking at visual stimuli and responding to them. How could it not deal with this? Yet vision scientists are uneasy about the claim that discrimination data say something reliable about the content of experience. If one asks a subject what red looks like, the subject will revert to demonstrative claims – it looks like *that* – or to relational claims locating a red color presentation in relation to that of other colors: more like orange and yellow, less like blue and green. While these descriptors are often robust for subjects, they do not, or so it is often claimed, get at the subjective nature of color experience. In an influential article, “What is it like to be a bat?”[3] The philosopher Thomas Nagel argued exactly that. Bats, Nagel proposed, have

experience – there is something that it is like to be a bat (as opposed to, say, a stone). While one can use the techniques of animal ethology, biology, and bat psychophysics to determine bat discrimination space, the content of the bat's experience is beyond the grasp of those third-person methods – beyond an objective 3rd person science and its “view from nowhere.” Nagel's argument really has little to do with bats. They are a useful exemplar since it is easy to imagine (a) that bats have experience and (b) that their experience is distant from human experience. In this sense, the bat is a useful foil to arguments from analogy: one can easily believe that bats have experience, but there is no analogy to human experience to guide us because echolocation is a different sensory modality than either sight or hearing. That having been said, the central thrust of Nagel's argument concerns subjectivity. There is a subjective view of the world that science, as humans know it, cannot access. Subjectivity, in the end, is just as mysterious from the 3rd person point of view as is the bat's. This is why many psychophysicists are likely to be in agreement with philosophers skeptical as to the knowledge of color phenomenology – why it is that psychophysicists are uneasy about the inference from discrimination to experience – inference to the way colors “look.”

Not all philosophers, and certainly not all scientists, vision or otherwise, accept the view that subjective experience – including color experience – is mysterious from a third-person perspective. But disagreement over this issue is profound. One might argue that a good model of an individual's discrimination space for color is as good as the human behavioral sciences get. If, for example, a subject fails to discriminate images in some set of pseudo-isochromatic color plates, then one can make predictions about their future discriminatory behavior and also explain that behavior. Such tests do more than identify types of “color blindness”; they identify the axes on which colors are confused and may correlate such confusion with genetic differences in opsin expression at the retinal level. What more could one want? It seems that there are two things: (1) an account of *what* the subjective point of view is like, as opposed to a third-person take on subjective experience – this is Nagel's concern, and (2) an account of *how* subjective

experience is generated (how it fits in with the ontological “catalogue of the real”) – this is a concern most closely identified with the Australian philosopher David Chalmers [4].

Chalmers writes mainly about consciousness, but his views on that subject have clear implications for color phenomenology. Unlike Nagel, who makes a case for the subjectivity of experience and is concerned with how that experience might be understood objectively, Chalmers argues that the real problem with experience is that science has no idea as to how subjective experience is generated by a physical system (and, more radically, *why* there should be experience in the first place). Chalmers is not denying there is subjective experience. He is claiming that its causal story is incomplete. Suppose one could understand color perception “all the way up”: from stimulus presentation, to photon-absorption at the retinal level, to retinal and LGN opponent processing, to cortical processing in the visual areas of the brain, to integrative processing in the executive areas, to the output of discrimination-based behavior which is a function of this whole process. While vision science understands some elements of this causal story quite well, and others not as well, even *perfect* understanding of it might leave science in the dark as to how color experience is generated. At what point do the biological properties of brains cause or constitute experiences of color, and how? Chalmers argues that the science of consciousness, such as it is, has no idea how to even address, let alone answer this so-called hard question – “hard” not in virtue of the difficulty of the science (as with the molecular biology of vision, say) but “hard” in the sense that science offers *no advice* on how to bridge the gap between its cognitive-neurobiological accounts of the brain and the brain's generation of experience. On the basis of this and related concerns, Chalmers has drawn a number of unusual conclusions: that some form of mind-brain dualism is true and that consciousness is both an emergent property of brains *and* a fundamental property of the universe.

These are very controversial claims. Critics of Chalmers have typically adopted one of two approaches (see the commentaries reprinted in

Ref. [4]): (1) argue that Chalmers assumes the limitations of current brain science are permanent (this involves the positive claim that problems that look hard at a given point in time may become easier with new developments in theory and practice) and (2) argue that early progress has, even now, been made on a complete theory of conscious experience.

With respect to the first strategy, one can quite agree that, in the future, science is likely to succeed in areas currently unimagined. Yet (1) does not address the request for a way forward on the problem of brains generating experience but merely points out that science will almost certainly find new ways of addressing (or disposing of) the problem. This may be true, but the argument is not substantive, given the claim that the nature of experience and its origin is a different sort of problem. If one, further, (1) assumes that the development of science will be sufficient to explain conscious experience at some point in time, then one is simply denying Chalmers' claim and that, arguably, assumes what it should demonstrate. As for (2), the view that progress has been made on the explanation of consciousness, Chalmers has pointed out that contemporary empirical theories of conscious experience (a) shift the problem of experience and how it is caused or constituted to accounts wholly within cognitive neuroscience and biology – an account of attention, say, or an account of neural opponent processing for the case of color. As a consequence (b), such accounts will provide, at best, accounts of the “neural correlates” of conscious experience rather than an account of how conscious experience is generated by neurobiological function. This, however, may not be such a bad thing nor is it quite the limitation that, at first glance, it appears to be.

Color Phenomenology and the Structure of Color Experience

Claims that science does not understand how phenomenological experience arises or is constituted from neurobiological function or that a complete understanding of human phenomenology surpasses what can be known from a third-person scientific perspective sound dire: as if experience

is not understood at all. And yet it is remarkable how much about the phenomenology of color can be known “from the outside” – and from a position of ignorance as to the ultimate causes of that phenomenology. As noted in paragraph 3, psychophysics would be able to identify the spectral inversion that has troubled philosophers. Such detection really is no different, in principle, than the detection of different forms of dichromacy: protanopia, deuteranopia, tritanopia – all involving failure to discriminate among stimuli that a normal trichromat would discriminate. Each of these deficiencies (relative to trichromacy) is a consequence of the lack of one or another of the three typical photoreceptors. They are physiologically based effects that can be identified through behavioral tests, and, moreover, vision science can explain the relevant deficiencies and their fine-grained differences at physiological and genetic levels. Color science has, in other words, a good grasp of the *structure* of human (primate) color experience, as well as its causes, even if the precise, personal, subjective nature of that experience remains epistemically problematic.

References

1. Cohen, J., Matthen, M. (eds.): *Color Ontology and Color Science*. MIT Press, Cambridge, Massachusetts (2000)
2. Byrne, A.: Inverted qualia. In: *Stanford Encyclopedia of Philosophy* (2004). Online (substantially revised 2010). <http://plato.stanford.edu/archives/spr2010/entries/qualia-inverted/>.
3. Nagel, T.: What is it like to be a bat? *Philos. Rev.* **83**, 435–450 (1974)
4. Chalmers, David.: *Explaining Consciousness: The Hard Problem*. MIT Press, Cambridge, Massachusetts (2000)

Color Pollution

Malvina Arrarte-Grau
Architecture, Landscape and Color Design,
Arquitectura Paisajismo Color, Lima, Peru

Synonyms

Inappropriate use of color; Visual contamination;
Visual pollution

Definition

Color pollution consists of an inappropriate color arrangement which causes or increases disorder in the perception of the Visual Field within an urban or natural environment. It is an important aspect of visual pollution.

Overview

Color Pollution as a Component of Visual Pollution

In the context of environmental design visual pollution refers to all non-architectural elements which spoil, in an invasive and simultaneous way, the perception of outdoor spaces. These features range from plastic bags trapped between tree branches to publicity panels, street signs, posts and wires [1]. Besides the unaesthetic consequences the unnecessary exposure of these elements may bring, visual pollution leads to the overstimulation of the senses, increasing the load of information to be processed by viewers, drivers and pedestrians.

Light and color are important aspects of visual pollution. The first one generates luminous pollution or ► [Light Pollution](#) due to artificial lighting and the second one, color pollution by the introduction of cultured elements. Color pollution may originate in visual pollution, when it is the consequence of an inconveniently positioned feature which reinforces its presence by color. Otherwise it may be a color that is incoherent within the composition and causes disorder. The disharmonious result may be produced by one or more color dimensions (hue, saturation and lightness) [2]. Achromatics are necessary for creating transition spaces between color information, though black, white and gray may also produce an effect of color pollution by lack of tone and contrast in luminosity.

Negative Impact of Color in the Environment

Color affects the perception of objects and compositions within a setting. It serves to codify elements and establish visual hierarchies. If used randomly, disregarding its power, color may

easily draw attention to misleading data, distracting the viewer from relevant information for the interpretation of a scene [3]. When inappropriate colors are used in minor elements such as urban furniture, advertising panels or building elements, color, far from being a useful signal, ends up invading the environment with an ambiguous and arbitrary presence.

Color as Intended by Nature and Adaptation to Color Coding

Colors in nature convey physical and chemical characteristics. They also give information on material processes and ethereal substances. In the animal kingdom color has a biological role, as an aid to survival [4]. In natural conditions human beings would respond to colors by instinct, using them as intended by nature: for alerting and announcing danger, for recognizing food and bodily functions and for perceiving space and distance.

The man-made environment is full of visual information and signals, many of which are based on color. Universal coding by color distinguishes hot from cold, gas from water, forbidden from allowed and so forth. Color is present in domestic and work activities. When shopping, colors in packaging and logos inform about names and brands, colored tags serve to mark sales from regular prices, while bright and fluorescent colors indicate special deals.

When color is used inadequately, the information it conveys becomes irrelevant and negative for the user, causing confusion and contributing to visual pollution. Man possesses a biological memory that keeps him aware of color signals, but using this faculty amidst color excess and randomness may prove a waste of energy, causing fatigue and stress.

Color Pollution in Cities

Color pollution exists in the urban environment, predominantly in cities and villages. Commercial and mixed-use areas are good examples of competition for capturing attention. In isolation, color communication is effective. The use of written signs and subtle colors may work too, but among competing signs, these no longer accomplish their

intended function, promoting other formulas are to be put into practice: brighter colors, contrasting backgrounds, larger letter types, images, lighting and many others [5]. The arrangement of the signs may easily become visual pollution, while the excessive and inappropriate use of color contributes to a chaotic environment.

The effects of color pollution are negative. Bright and saturated colors tend to spread in the scene, as others are induced to use the same strategy. In cities color tends to go out of control in commercial panels. Traffic and safety signs [2] diminish their effectiveness when there are objects with similar colors around. Associations between certain colors and objects, which have not been planned carefully, tend to become ingrained in a cityscape, i.e. urban furniture, bridges and signals. From that base, the rest is decided or added on.

Architectural elements may cause color and visual pollution within a façade by producing imbalance in the composition. When building exteriors invade the visual field with bright and saturated colors, far from enhancing architecture, these threaten the aesthetic aspect of the environment.

Color Pollution in Natural Settings

Natural settings are highly susceptible to visual and color pollution. This occurs when man-made elements are introduced in a landscape, with no regard to material quality and color, resulting in alien and obtrusive. It is common to find posts, cables, advertising and constructions invading the Visual Field, interrupting the continuity of mountain ranges, seascapes, agricultural fields and woods. The position, size and frequency of these elements are relevant. Additionally their impact may be emphasized or diminished by materials and colors. For instance, a black plastic water tank may be very disruptive when placed against a mountain backdrop. Its visual impact would be attenuated if it were beige, echoing the color of the background.

Consequences of Color Pollution

As occurs with noise and visual contamination, color pollution may be aggressive, causing fatigue

and stress. Slow reactions and traffic accidents may be caused by excess of information and distracting colored elements. As color pollution affects the aesthetic aspect of a place, users may find it hard to enjoy and develop a sense of belonging in visually disrupted settings.

Though the perception of color in animal species differs from that of humans, pollution by color may diminish the chances of wildlife establishment near human conglomerates. In prey birds and day birds, which have an acute and sophisticated sense of vision, color plays specific functions in mating and feeding [2, 6]. The presence of artificially colored elements in their habitat may affect their behavior by alerting and confusing them.

Causes of Color Pollution

If color does not correspond to function nor to survival or aesthetics, who is to blame for its randomness? Lack of regulation and guidelines for the use of color allow the invasion of contaminating colored elements in the environment. This is a major deficiency of cities and natural settings. The causes of color pollution also rely on the consumer society, the commercial offer of paint and cladding materials, and visual marketing strategies.

Man, with his ability to synthesize and create from the materials which are granted to him, uses color to express, symbolize and communicate ideas. The evolution of coloring mediums, tools and technologies has resulted in an overwhelming variety. By contrast, the rules for the application of these advancements are scarce.

Municipalities rent strategic spaces to corporations for publicity. Owners will rent the roofs of their houses to advertising companies, if they are not prevented to do so. To the detriment of the setting a gigantic beer can or a credit card could become a focal point. The colors used in these advertisements are centered in the product, not in the environment, and very often result in color pollution.

It appears as common practice in many parts of the world that, at the time of political elections, propaganda invades city and countryside. Usually the combination of two primary colors of medium

to high saturation and black is used over a white background, producing a contrasting figure-ground relationship which is legible from a distance [5]. Recurring examples of visual and color pollution of this sort remain on walls and along roadsides long after the campaigns are over.

It is for sure that the cultivated and sensitive individual values unspoiled scenery more than authorities and inhabitants of rural towns, where the beauty of natural sights seems to be taken for granted.

Solutions for Color Pollution

Authorities have in their hands the possibility to control the environmental impact caused by political propaganda and commercial advertising through regulations and sanctions.

Decisions for preserving and improving environmental colors are common ground of authority and designer. It is necessary to assess the environment and develop regulations according to its unique qualities, visual advantages and important buildings. These should be taken as standpoints for design and planning. The design of features to go in a landscape or cityscape should be thoughtful and consider the particular characteristics of the setting, including color. Color is a powerful tool but is just one of many in a composition. Parameters regarding order, geometry, repetition, size, shape and material should be part of the regulation criteria.

Reducing the variety and amount of elements in the visual field is crucial. Color may be protecting or decorating some element, which is obtrusive per se. The addition of features, such as signals and urban equipment, requires planning and restrictions in the color aspect too. In particular cases, the introduction of greenery may help to cover up the obtrusive elements, to create structure and order in the visual field or to balance a composition.

The adjustment of color dimensions (hue, saturation and lightness) may be effective for resolving certain visual conflicts. Through the use of adequate colors, annoying features could be neutralized or kept inconspicuous. In this way color would serve the purpose of contributing to visual order.

Cross-References

► [Light Pollution](#)

References

1. Couto, M.: Contaminación visual del paisaje. <http://www.monografias.com/trabajos-pdf2/contaminacion-visual-paisaje/contaminacion-visual-paisaje.pdf> (2007). Accessed 10 Oct 2014
2. De Grandis, L.: Teoría y uso del color. Ediciones Cátedra, Madrid (1985)
3. Burga, J.: Del espacio a la forma. Facultad de Arquitectura, UNI, Lima (1987)
4. González, G. (ed.): El gran libro del color. Blume, Barcelona (1982)
5. Folis, J.: Architectural signing and graphics. Whitney Library of Design, New York (1979)
6. Wikipedia: Bird vision. http://en.wikipedia.org/wiki/Bird_vision. Accessed 10 Oct 2014

Color Pop-out

► [Color and Visual Search](#), [Color Singletons](#)

Color Preference

Stephen E. Palmer¹ and Karen B. Schloss²

¹Department of Psychology, University of California, Berkeley, Berkeley, CA, USA

²Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

Synonyms

[Color aesthetics](#); [Color harmony](#); [Favorite colors](#)

Definition

How much people like different colors.

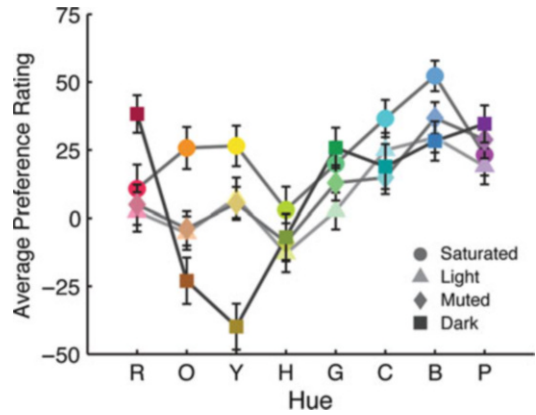
Overview

One of the most fascinating aspects of the perception of colors is that people have relatively strong preferences, liking certain colors and color combinations much more than others. This entry discusses what is known about human color preferences, not only in terms of *which* colors and color combinations people like but also *why* they like them.

Preference for Single Colors

Average relative color preferences for a given sample of colors are typically measured behaviorally by asking a group of people to perform one of three tasks. First, the observers can be asked to indicate which of two simultaneously presented colors they prefer for each possible pair of colors in the sample. The probability, averaged over observers, of choosing each color versus all other sample colors is then taken as a measure of its average relative preference within that sample. Second, observers can be shown all of the colors in the sample simultaneously and be asked to rank order them from most preferred to least preferred. The average rank of each color across observers provides a measure of average relative preference within the sample. Third, observers can be shown a single color on each trial and be asked to rate their preference for it on a discrete (e.g., 1–7) or continuous (e.g., marking along a line segment) rating scale. Average ratings across observers can then be taken as a measure of average relative preference for colors within the sample. Correlations among these different measures tend to be quite high when the same observers (or large samples of different observers from the same population) judge the same colors.

Although early researchers often claimed that color preferences were simply too idiosyncratic to be worth an empirical study, modern measurements of well-calibrated, computer-generated displays of standardized colors using improved data analysis techniques have now established that there are indeed reliable and repeatable patterns in group data [1]. These average color preferences



Color Preference, Fig. 1 Color preference ratings in the USA as measured by Palmer and Schloss [2]. Preferences for the saturated, light, muted, and dark colors are plotted as a function of hue: red (R), orange (O), yellow (Y), chartreuse (H), green (G), cyan (C), blue (B), and purple (P)

are most easily understood in terms of the three primary dimensions of human color experience (see entry for ► [Psychological color space and color terms](#)): hue (its “basic color”), saturation (how vivid or pure the color is), and lightness or brightness (how light versus dark the color is). Figure 1 plots average adult preference ratings in the USA for a wide gamut of 32 chromatic colors, consisting of eight hues – red, orange, yellow, chartreuse (yellow green), green, cyan (blue green), blue, and purple – in shades that are either highly saturated, light, muted (desaturated, mid-level lightness), or dark [2].

As Fig. 1 shows, average color preferences show a clear maximum around saturated blue and a clear minimum around dark yellow (greenish brown or olive). The majority of this variation is due to differences along the blue-to-yellow dimension of hue, with bluer colors being generally preferred to yellower colors. There is much less variation in the red-to-green dimension. In addition, people generally prefer more saturated colors over less saturated ones, with little difference between the light (pastel) and muted tones, at least in the USA (see the entry for ► [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)). The most interesting finding theoretically is the rather striking difference between the shape of the hue preference curve

for the dark colors versus those for the Munsell chroma matched light and muted colors. In particular, there are dramatic decreases in preference for dark orange (brown) and particularly for dark yellow (greenish brown, or olive) relative to the light and muted oranges and yellows, but there are also modest increases in preference for dark red and dark green relative to the light and muted reds and greens [2]. Although gender differences among American adults are relatively slight, men like saturated colors more than women do, whereas women tend to like muted colors more than men do [3]. The overall pattern of preferences for single colors is thus complex but clear and replicable. For a more extensive review of modern studies of single color preferences, the reader is referred to Whitfield and Wiltshire [1].

Infant Color Preferences

Infant color preferences are studied by examining the looking behavior of babies when they are shown pairs of colors side by side. It is generally assumed that the color at which the baby looks longer and/or first is preferred to the alternative. Color preferences are therefore measured by determining the average looking times and/or the probabilities of first looks [4]. Infants younger than about 3–4 months tend not to be studied because the short-wavelength sensitive cones do not mature until that age, making them functionally color deficient relative to adults (see entry on ► [Color Perception and Environmentally Based Impairments](#)).

When infant looking preferences are measured for highly saturated colors, the hue preference function tends to have roughly the same shape as the corresponding adult hue preference function, with a peak around blue and a trough around yellow to yellow green [5]. Great care has to be taken to ensure that hue-based color preferences actually reflect differences in hue by controlling for luminance, brightness, discriminability, colorimetric purity, and saturation. More recent studies that directly compared infant preferences with those of adults for less saturated, but better matched, colors have found important differences, however. In particular, infant preferences for these color samples vary primarily on the red-to-green

dimension, with redder colors being more preferred, and do not vary much on the blue-to-yellow dimension [6]. Because this pattern for infants is opposite that for adults, color preferences must either be subject to learning as a result of experiences with differently colored objects or there must be a substantial maturational process that influences color preferences.

Color Preferences in Different Contexts

A question of considerable applied interest is how adult color preferences for patches of “context-free” colors, as described above, generalize to preferences for colored objects. Clearly, they do not generalize for natural objects that have prototypical colors, because yellow bananas and red strawberries are strongly preferred to blue ones, but better generalization is evident for artifacts that could, in principle, be produced in virtually any color, such as shirts, walls, sofas, and cars.

Schloss, Strauss, and Palmer [7] studied preferences for the same 32 colors graphed in Fig. 1 when they were judged as the colors of walls, trim, couches, throw pillows, dress shirts/blouses, ties/scarves, and T-shirts. They found that the shape of the hue preference function for context-free colored squares (i.e., as in Fig. 1, but averaged over different lightness and saturation values) was largely the same as that of hue preference functions for all the different object contexts they studied. The only clear exceptions were that large, red objects (e.g., walls, trim, and couches) were liked less than smaller red objects. In contrast, there were marked differences in preferred lightness and saturation levels across different objects, often depending on practical considerations, such as walls being preferred in lighter tones that make rooms appear more spacious and couches being preferred in darker shades that hide dirt. And although saturated colors are generally the *most* preferred colors for context-free squares of color (see above), they are actually the *least* preferred colors for all of the objects tested. Other researchers have found similar results: although context-free color preferences are dominated primarily by hue, object-specific color preferences were more strongly affected by lightness and saturation levels (e.g., in car colors, with darker tinted/shaded colors being more preferred than

lighter, grayish colors) [8]. Even within the basic-level object category of cars, however, Schloss et al. found striking differences: the most preferred colors for a luxury sedan were achromatic (black, gray, or white), consistent with their conventional formality as serious, sophisticated cars, whereas color preferences for a VW “Bug” tended toward brighter, warmer, more saturated colors (e.g., yellow), consistent with their conventional informality as fun, sporty cars [7]. Such results can be interpreted as weaker cases of the prototypical banana and strawberry examples mentioned above but reflecting sociocultural conventions rather than natural prototypes. Emotional reactions to colors can also be important in preferences for colors chosen for different residential rooms. People prefer room colors to correspond with their desired feeling when inhabiting the room: e.g., light blue is preferred for the living room because it feels calm, whereas “near white,” green, blue, and yellow are preferred for the bathroom because they feel hygienic and/or pure [9].

Theories of Color Preference

Thus far this entry has focused on describing *which* colors people like and dislike. But *why* do they like the ones they do? Indeed, why do people have color preferences at all? Although color in the natural world sometimes carries significant information (e.g., about ripe versus unripe fruit), it is relatively inconsequential in most modern artifacts (see above). Several theoretical explanations of color preference have been proposed and tested, including ones grounded in physiology, psychophysics, emotion, and ecological objects. Because all of these models have been tested against the data shown in Fig. 1, those data will be used as a benchmark.

The most physiologically oriented theory [10] suggests that people like colors to an extent that depends on a weighted average of cone contrasts relative to the background believed to be computed very early in visual processing: $L-M$ and $S-(L+M)$, where S , M , and L represent the outputs of cones maximally sensitive to short, medium, and long wavelengths of light. Hurlbert and Ling’s model fits their own data very well (accounting for 70 % of the variance) but fits the

data in Fig. 1 only about half as well (37 %), no doubt because their sample of colors did not include the highly saturated and nameable colors in Palmer and Schloss’s [2, 3] color sample.

A related but purely psychophysical hypothesis is that color preferences are based on conscious color appearances. Palmer and Schloss [2, 3] tested this possibility using a weighted average of observer-rated redness-greenness, blueness-yellowness, saturation, and lightness of each color, roughly analogous to their coordinates in the Natural Colour System (see entry on ► [Color Order Systems](#)). This model did a much better job in accounting for the data in Fig. 1 (60 % of the variance), suggesting that a later, conscious representation of color provides a better basis for color preference than an early, nonconscious, one based on retinal cone contrasts.

A third type of explanation can be constructed in terms of the emotional associations of colors. The basic hypothesis is that people may like colors to the extent that they like the emotions that are evoked by or associated with those colors. Ou et al. measured *color emotions* through subjective judgments of many emotion-related terms and related those ratings to color preferences [11]. Their results showed that three factor-analytic dimensions underlay color emotions: *active-passive*, *light-heavy*, and *cool-warm*, explaining 67 % of the variance in their preference data. Palmer and Schloss fit observers’ subjective ratings of these dimensions to the data in Fig. 1 and found that it accounted for 55 % of the variance, with people liking *active*, *light*, and *cool* colors more than *passive*, *heavy*, and *warm* ones [2, 3].

The ecological valence theory (EVT) of color preference was formulated by Palmer and Schloss to test the hypothesis that people like colors to the degree that they like the environmental objects that are characteristically those colors [2, 3]. For example, people tend to like blues and cyans because they like clear sky and clean water, and they tend to dislike browns and olive colors because they dislike feces and rotting food. The theoretical rationale of the EVT is that it will be adaptive for organisms to approach objects and situations associated with the colors they like and to avoid objects and situations associated with the

colors they dislike to the extent that their color preferences are correlated with objects and situations that they like versus those that they do not like. They reported strong support for the EVT through empirical measurements of the weighted affective valence estimates (WAVEs) for the 32 chromatic colors in Fig. 1. The WAVE for each color measures the extent to which people like the objects that are associated with that color. It was computed from observers' average valence (liking/disliking) ratings for *all* things named as associates for each color studied, with each valence rating being weighted by the similarity of the given color to the color of that associate. Although blue is strongly associated with objects that are almost all liked (e.g., clear sky, clean water, swimming pools, and sapphires), most colors have associates with both positive and negative valences. Nevertheless, the average WAVE, computed as specified above, explained 80 % of the variance in the data shown in Fig. 1 with no estimated parameters. This does not mean that color preferences are irrelevant to object preferences: clearly they matter for functionally identical artifacts that are available in many colors (e.g., clothes, furniture, cars, and appliances). However, the EVT suggests that those preferences arose initially from associations with characteristically colored objects and were then positively or negatively reinforced to the extent that people have positive or negative experiences with them.

Preference for Color Combinations

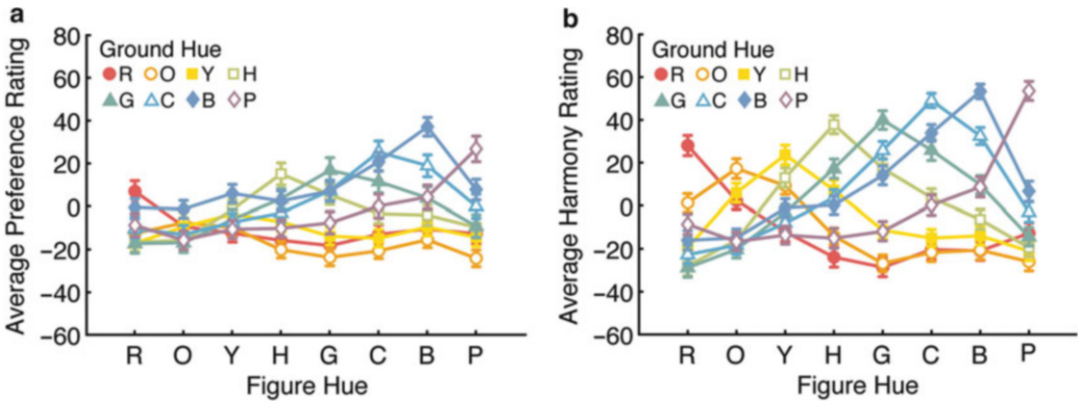
Chevreul formulated the most influential art-based theory of color harmony (or color preference, because he used the terms interchangeably), which claimed that there are two distinct types: *harmony of analogous colors* and *harmony of contrast* [12]. In brief, harmony of analogous colors includes *harmony of scale* (colors of the same hue that are similar in lightness) and *harmony of hues* (colors that are similar in hue and the same in lightness). Harmony of contrast includes *harmony of contrast of scale* (colors of the same hue that differ in lightness), *harmony of contrast of hues* (colors of similar hue that differ in

lightness), and *harmony of contrast of colors* (colors that are different in both hue and lightness). Other theories include Itten's claim that combinations of colors are harmonious provided that the colors produce neutral gray when mixed as paints and Munsell's and Ostwald's theories that colors are harmonious when they have certain relations in color space (e.g., they are constant in hue and saturation but vary in lightness) [13]. None of these theories was formulated on the basis of aesthetic measurements, although some have since been tested empirically.

Empirical Research on Color Pair Preference/Harmony in Combinations

Schloss and Palmer attempted to clarify the confusion surrounding preferences for color pairs by explicitly distinguishing among three different concepts: *pair preference*, *pair harmony*, and *figural preference* for a color against a colored background [14]. They defined *pair preference* as how much an observer *likes* the combination of the two colors as a whole. They defined *pair harmony* as how well the two colors *go together*, regardless of whether the observer likes the combination or not (analogous to the distinction between harmony and preference in music, wherein nearly everyone agrees that Mozart's music is more harmonious than Stravinsky's, but some prefer Mozart and others prefer Stravinsky). They defined *figural preference* as how much the observer likes the single color of the figure when viewed against a different color in the background. Although figural preference involves a judgment about the single color of the figure, it is relevant to preferences for color combinations because the same color can look strikingly different on different background colors (see entry on ► [Simultaneous Color Contrast](#)).

Figure 2a shows average preference ratings for color pairs as a function of the hue of the figure (x-axis) and that of the ground (the different curves). The primary pattern in the data is that, for every background hue, people prefer combinations in which the figure has the same or a very similar hue. Clearly, people tend, on average, to like color combinations that are the same or similar in hue but differ in lightness, which Chevreul



Color Preference, Fig. 2 Rated preference (a) and harmony (b) for pairs of colors as a function of the figural hue (x-axis) and ground hue (different curves)

called harmonies of analogous colors. There is no evidence for Chevreul's harmonies of contrast, however, because there are no reliable increases in the functions at opposite hues (e.g., red and green). A secondary fact is that people tend to like color combinations to a degree that reflects their preferences for figure and ground colors, with combinations on blue backgrounds being generally most preferred and those on orange backgrounds being least preferred. This tendency is relatively minor, however, accounting for only 22 % of the variance in preference ratings.

Harmony ratings of the same color pairs are plotted in Fig. 2b. Clearly, they are very similar to the preference ratings ($r = +.79$), except that the peaks at same-hue combinations are even more pronounced. This high correlation largely explains why preference and harmony have so often been equated: fully 62 % of the variance in preference ratings is explained by harmony ratings. An additional 14 % of the variance can be explained by including preferences for the individual figure and ground colors, explaining a total of 76 % of the variance in people's color preferences for figure-ground combinations.

Ratings of figural preference for colors against colored backgrounds are measurably different from ratings of both pair preference and pair harmony, showing clear effects due to the hue contrast and lightness contrast of the figure against the background: warm figures (e.g., red, orange, and

yellow) were preferred against cool backgrounds (e.g., green, cyan, and blue) and vice versa [14]. It appears that what Chevreul termed harmonies of contrast actually apply to judgments about preferences for figural colors when viewed against different colored backgrounds. Given the general preference for saturated colors described previously, it is not surprising that observers prefer figural colors against highly contrastive background hues, because these would produce the strongest simultaneous color contrast effects, thus increasing the perceived saturation of the figural region.

Theories of Preferences for Color Combinations

The foregoing describes which color combinations people prefer and which ones they find harmonious, but why do these variations in preference and harmony arise? To the extent that pair preferences are influenced by preferences for the component colors, ecological associations of colors with objects are one important factor. Pair preferences are also influenced by people's positive/negative associations with objects and/or institutions that are associated with those colors in combination. For example, Schloss, Poggesi, and Palmer [15] investigated preferences for school colors among Berkeley and Stanford students: blue and gold for Berkeley and red and white for Stanford. They found that Berkeley students liked Berkeley color combinations better

than Stanford students did and Stanford students liked Stanford colors better than Berkeley students did, with these effects increasing with increasing amounts of self-rated school spirit. Such results clearly imply that ecological effects are present in preference for color combinations as well as for individual colors.

Perhaps the most important finding about preferences for color pairs is that people generally like harmonious combinations of the same (or similar) hue that differ in lightness. Although it is not immediately obvious why this might be the case from an ecological viewpoint, Schloss and Palmer suggested that color harmony might stem from ecological color statistics in natural images corresponding to different areas of the same ecological object [14]. A red sweatshirt, for example, would be darker red where it was in shadow and lighter red where it was brightly illuminated. Accordingly, pairs that are judged to be most harmonious (i.e., that “go together” best) may, in fact, be those that are most likely to co-occur within the same object in natural images.

Acknowledgments This material is based upon work supported by the National Science Foundation under Grant Nos. 0745820 and 1059088. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Cross-References

- [Color Perception and Environmentally Based Impairments](#)
- [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)
- [Psychological Color Space and Color Terms](#)
- [Simultaneous Color Contrast](#)

References

1. Whitfield, T.W.A., Wiltshire, T.J.: Color psychology: a critical review. *Genet. Soc. Gen. Psych.* **116**, 385–411 (1990)
2. Palmer, S.E., Schloss, K.B.: An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8877–8882 (2010)

3. Palmer, S.E., Schloss, K.B.: Ecological valence and human color preference. In: Biggam, C.P., Hough, C. A., Kay, C.J., Simmons, D.R. (eds.) *New Directions in Color Studies*, pp. 361–376. John Benjamins, Amsterdam (2011)
4. Teller, D.Y.: The forced-choice preferential looking procedure: a psychophysical technique for use with human infants. *Infant Behav. Dev.* **2**, 135–153 (1979)
5. Teller, D.Y., Civan, A., Bronson-Castain, K.: Infants’ spontaneous color preferences are not due to adult-like brightness variations. *Vis. Neurosci.* **21**, 397–401 (2004)
6. Franklin, A., Bevis, L., Ling, Y., Hurlbert, A.: Biological components of colour preference in infancy. *Dev. Sci.* **13**, 346–354 (2010)
7. Schloss, K.B., Strauss, E.D., Palmer, S.E.: Object color preferences. *Color Res. Appl.* **38**, 393–411 (2013)
8. Saito, T.: Latent spaces of color preference with and without a context: using the shape of an automobile as the context. *Color Res. Appl.* **8**, 101–113 (1983)
9. Manav, B.: Color-emotion associations and color preferences: a case study for residences. *Color Res. Appl.* **32**, 144–151 (2007)
10. Hurlbert, A., Ling, Y.: Biological components of sex differences in color preference. *Curr. Biol.* **17**, 623–625 (2007)
11. Ou, L., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. Part III: colour preference modeling. *Color Res. Appl.* **29**, 381–389 (2004)
12. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colors and Their Applications in the Arts*. Van Nostrand Reinhold, New York (1939)
13. Westland, S., Laycock, K., Cheung, V., Henry, P., Mahyar, F.: Colour harmony. *Colour: Des. Creativity.* **1**, 1–15 (2007)
14. Schloss, K.B., Palmer, S.E.: Aesthetic response to color combinations: preference, harmony, and similarity. *Atten. Percept. Psychophys.* **73**, 551–571 (2011)
15. Schloss, K.B., Poggesi, R.M., Palmer, S.E.: Effects of university affiliation and “school spirit” on color preferences: berkeley versus stanford. *Psychon. Bull. Rev.* **18**, 498–504 (2011)

Color Processing, Cortical

Daniel C. Kiper

Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

Synonyms

[Chromatic processing; Cortical](#)

Definition

The transformation of color signals and chromatic properties of receptive fields within the visual cortex of primates.

Processing of Chromatic Signals in the Early Visual Pathways

The processing of chromatic signals in the retina and lateral geniculate nucleus (LGN) has been the focus of numerous studies and is well understood. Less is known about the fate of color signals in the cortex. This entry first reviews central aspects of color processing in the primary visual cortex (V1) and discusses how it differs from subcortical processes. It then discusses the processing of color signals in extrastriate visual areas.

Color in the Striate Cortex (V1)

Chromatic Properties of Individual Neurons

In the primate primary visual cortex, it had been estimated that about 50 % of the cell population is selective for color [1]. Estimates of the proportion of color-selective cells, however, are complicated by the use of different criteria for the classification across studies. Interestingly, color signals in primary visual cortex have long been thought to be relatively weak relative to black and white stimuli. Studies based on imaging techniques of the active human brain, however, have now clearly demonstrated that primary visual cortex contains a large proportion of color-responsive and color-selective cells [2]. The majority of color-selective cells in V1, like those of its subcortical input, the lateral geniculate nucleus (LGN), simply add or subtract their chromatic inputs. Indeed, most V1 [3] cells' responses to chromatic modulations are well accounted for by a model postulating a linear combination of the signals derived from the three cone classes. Although there are some V1 cells that are more selective for color than predicted [4], the model adequately fits the responses of the majority of V1 cells.

The preferred colors of V1 neurons [2], however, do not cluster around the three cardinal directions found in the LGN [5]. Instead, cells often prefer colors that lie intermediate to these "cardinal" directions. For cortical cells, the classification of chromatic cells into red-green or blue-yellow opponent cells is therefore not valid anymore.

Moreover, most color-selective cells in V1 [6] respond also vigorously to luminance variations, a property that seems ubiquitous in the visual cortex. It also illustrates the fact that color selectivity does not imply color opponency. A response to stimuli containing chromatic but no luminance information (isoluminant stimuli) does not imply that the neuron receives opponent inputs from two or three cone classes. True color opponency can be deduced in neurons that give stronger responses to isoluminant than to luminance stimuli, provided that the stimuli are equated for cone contrasts [7]. It thus appears that in the cortex, there is a whole continuum of cells, ranging from strict color opponency to strict luminance [8].

Recent studies have also shown that, unlike earlier levels, primary visual cortex contains so-called double-opponent cells [6]. Double-opponent cells are cells whose receptive field combines color and spatial opponency. The defining functional property of these cells is that they respond strongly to color patterns but weakly or not at all to uniform (full-field) color modulations. The existence of pure double-opponent cells, cells with non-oriented receptive fields that respond only to color patterns [9], has been hotly debated in the last years. While early reports have been contested, it now appears that double opponency does exist in primary visual cortex but linked with other functional properties such as orientation selectivity and a high sensitivity for luminance stimuli as well [6].

Clustering and Specialization of Color-Selective Cells in V1

In the last decades, the localization and functional specialization of color-selective cells within area V1 have been the matter of debate. A number of studies reported that color-selective cells in V1 are not orientation selective (a canonical property of

virtually all primate V1 neurons) and that these cells are clustered within the cytochrome oxidase (CO)-rich blobs that have been described previously [10]. Both claims, however, have been contested by a number of anatomical and physiological studies [8]. Some of the controversy can be attributed to the different techniques and stimuli that have been used to localize and characterize color-selective cells. A number of extensive and careful studies using single-cell recordings combined with histological processing have clearly shown that color-selective cells are not restricted to the CO blobs.

The same studies have shown that color selectivity and orientation selectivity are not, as originally proposed, segregated within the V1 cell population. The most convincing, but not unique, evidence for conjoint selectivity to color and orientation comes also from single-cell recordings [11].

The Case of Blue-Yellow Signals

For a number of mostly technical reasons, many color-vision studies have focused primarily on red-green opponency and less on the evolutionary older blue-yellow system. The discovery of the koniocellular layers in the LGN and their predominant role in the processing of blue cone signals (necessary for blue-yellow opponency) [12] raised the possibility of a specialized treatment of blue cone signals within the primary visual cortex. To what extent blue cone signals are clustered within V1 and how much they contribute to cortical color processing is not clear yet. Recent studies seem to indicate that blue cone signals are distributed uniformly within V1 (no clustering) and that these signals might be combined with achromatic signals in the double-opponent cells described above.

Color in the Extrastriate Cortex

Although several studies showed that individual neurons in the dorsal visual pathway, in particular in area MT of the macaque monkey, can

significantly respond to chromatic variations [13], these responses are typically smaller than those obtained with luminance stimuli and do not account for the animal's behavioral performance. The following sections are thus restricted to areas of the ventral pathway that are known to play a critical role in color processing.

Proportion of Color-Selective Cells

In extrastriate areas of the ventral pathway, the number of neurons whose responses are affected by color variations remains surprisingly constant despite the variability of the criteria used to classify neurons. This proportion reaches 50 % in area V2 and 54 % in V3 [1]. Estimates in later areas of the ventral pathways (reviewed in 8) are more variable. In area V4, often but mistakenly considered a "color" area of the primate brain, original estimates ranged from less than 20–100 %. A more recent estimate of 66 % has been reported. In the IT cortex, it has been estimated around 48 %.

Chromatic Properties of Individual Neurons

As already described in V1 neurons, the chromatic properties of extrastriate neurons differ from those in the retina and LGN in two important ways. First, a significant proportion of neurons in each area possess a high degree of color selectivity. These neurons show a narrow tuning in color space. Narrowly tuned neurons have been reported, in different proportions, within areas V2, V3, and V4. The second major difference already described in V1 concerns individual neurons' preferred colors. As in V1, the distribution of preferred colors of extrastriate neurons is not clustered in color space but is uniformly distributed [8].

Color Versus Other Visual Attributes

The question of conjoint selectivity for color and other visual attributes has been posed and answered for V1 neurons (see above). Similarly, the question whether color and other visual attributes such as orientation, motion, and size are segregated in the extrastriate cortex has been

studied. Numerous studies (see [8]) in the last decade have shown that cells within V2, V3, and V4 can be concomitantly tuned for several dimensions of the visual stimulus. It thus appears that color is not processed independently but by the same neuronal populations that also code orientation or size. Note however that a recent study reported the existence of a significant subpopulation of V4 cells that respond to chromatic but not luminance variations leading the authors to suggest that color and luminance might be treated by different channels within area V4 [14].

Clustering of Color-Selective Cells

The debate concerning the organization of color-selective cells into clusters within a given area has extended to several visual areas beyond V1. In V2, the thin bands defined by CO staining have been reported to represent clusters of color-selective cells and to be the source of the color signals sent to area V4 and areas of the inferotemporal cortex. Moreover, optical imaging studies of area V2 concluded that color is represented in an orderly fashion within the thin stripes in the form of well-defined color maps that resemble those based on human color perception. As in V1, however, the clustering of color selectivity within the thin CO stripes has been challenged.

In areas more posterior within the temporal pathway, color-selective cells have been reported to cluster into subregions. This clustering is in fact often offered as an explanation for the large variance in estimates of the proportion of color-selective cells in the temporal cortex. Microelectrode recordings in dorsal V4 are thought to encounter many color clusters that seem to be much less prevalent in ventral parts of V4.

In the most anterior parts of the ventral pathways, the distinction between cortical areas is much less clear, and the relationship between areas reported in the macaque brain with those of the human is still controversial. It has nonetheless been suggested that color might be treated within a specialized pathway that extends across several of the ventral visual areas including V2,

V4, and the dorsal portion of the posterior inferotemporal cortex (PITC) [15]. Within PITC, color-selective cells would be clustered into islands themselves containing orderly, columnar color maps, reminiscent of the organization previously reported in V2.

Is There a Color Center in the Primate Brain?

Several of the issues discussed above are intimately related to the question whether the primate brain contains one color center, a cortical area whose main function would be to support most or all aspects of color perception. The notion of a color center is a natural consequence for the proponents of a strictly modular view of cortical organization. Most researchers agree that there are areas in the ventral cortex that are highly activated by color stimuli, but none of these areas have been established as predominant over the others.

Moreover, experimental lesions in macaque V4, the most popular candidate for the role of color center, do not result in a complete loss of color vision. These results cast doubts on the notion of a unique color center and support the idea that color, like many other visual attributes, is treated within a network of neuronal populations distributed within the ventral occipitotemporal pathway.

Unresolved Issues

Important perceptual phenomena such as color constancy, unique hues, or color categorization remain largely unexplained. Further studies reconciling imaging or lesion results in the human brain with those obtained by methods revealing single activity in the macaque brain are thus necessary to fill these gaps. Moreover, the relationship between color signals and those associated with other visual attributes such as object shape or motion needs further scrutiny, particularly in extrastriate areas. Today, it is safe to say that a full understanding at the neuronal level of perceptual phenomena associated with color is still eluding the vision science community.

Cross-References

- [Color Vision, Opponent Theory](#)
- [Unique Hues](#)

References

1. Zeki, S.: The distribution of wavelength and orientation selective cells in different areas of the monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* **217**, 449–470 (1983)
2. Engel, S., Zhang, X., Wandell, B.: Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature* **388**(6637), 68–71 (1997)
3. Lennie, P., Krauskopf, J., Sclar, G.: Chromatic mechanisms in striate cortex of macaque. *J. Neurosci.* **10**(2), 649–669 (1990)
4. Cottaris, N.P., De Valois, R.L.: Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature* **395**(6705), 896–900 (1998)
5. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic properties of neurons in macaque LGN. *J. Physiol.* **357**, 241–265 (1984)
6. Johnson, E.N., Hawken, M.J., Shapley, R.: The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nat. Neurosci.* **4**, 409–416 (2001)
7. Schluppeck, D., Engel, S.A.: Color opponent neurons in V1: a review and model reconciling results from imaging and single-unit recording. *J. Vis.* **2**(6), 480–492 (2002)
8. Gegenfurtner, K.R., Kiper, D.C.: Color vision. *Annu. Rev. Neurosci.* **26**, 181–206 (2003)
9. Shapley, R.M., Hawken, M.J.: Color in the cortex: single- and double-opponent cells. *Vision Res.* **51**, 701–717 (2011)
10. Conway, B.R., Chatterjee, S., Field, G.D., Horwitz, G.D., Johnson, E.N., Koida, K., Mancuso, K.: Advances in color science: from retina to behavior. *J. Neurosci.* **30**(45), 14955–14963 (2010)
11. Leventhal, A.G., Thompson, K.G., Liu, D., Zhou, Y., Ault, S.J.: Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *J. Neurosci.* **15**(3), 1808–1818 (1995)
12. Henry, S.H., Reid, R.C.: The koniocellular pathway in primate vision. *Annu. Rev. Neurosci.* **23**, 127–153 (2000)
13. Gegenfurtner, K.R., Kiper, D.C., Beusmans, J.M., Carandini, M., Zaidi, Q., Movshon, J.A.: Chromatic properties of neurons in macaque MT. *Vis. Neurosci.* **11**(3), 455–466 (1994)
14. Bushnell, B.N., Harding, P.J., Kosai, Y., Bair, W., Pasupathy, A.: Equiluminance cells in visual cortical area v4. *J. Neurosci.* **31**(35), 12398–12412 (2011)
15. Conway, B.R., Tsao, D.Y.: Color architecture in alert macaque cortex revealed by fMRI. *Cereb. Cortex* **16**(11), 1604–1613 (2006)

Color Psychology

Zena O'Connor

Architecture, Design and Planning, University of Sydney, Sydney, Australia

Synonyms

[Chromotherapy](#); [Psychological impact of color](#)

Definition

Color psychology refers to a branch of study which postulates that color has a range of psychological or behavioral responses.

Overview

Color psychology and color therapy exist on the periphery of alternative medicine and are generally not accepted under the auspices of mainstream medical science and psychology. Despite this, a plethora of articles can be found in mainstream and digital media that discuss a link between color and a range of psychological, cognitive, biological, and behavioral effects. While it is often promoted that such a link exists on a universal, causal basis by some sectors of the media, there is minimal evidence to support this hypothesis. Furthermore, there are a number of factors that influence color and human response, and these include individual differences (such as age, gender, affective state, belief systems, and environmental stimuli screening ability), social and cultural differences, as well as contextual and temporal differences. Despite this, claims such as *red is stimulating and arousing* and *blue is calming, relaxing, and healing* are often quoted without substantiation or evidence of any nature in popular media. The source of such claims can be traced to a number of key authors including Faber Birren, Kurt Goldstein, Robert Gerard, and Max Lüscher, whose works have been either superseded or debunked for lack of scientific

rigor. However, some proponents have a vested interest in promoting a link between color and human response due to the popularity of such claims and the opportunity to capitalize on this popularity through the sale of books, workshops, courses, and consultancy services. While there is a place for such products and services, there are also compelling reasons to apply the principle of caveat emptor to color psychology and color therapy claims found in nonacademic sources.

Defining Color Psychology

The term “color psychology” suggests that the interface between color and human response is underpinned by a causal link wherein color has the capacity to influence a range of human responses including affective, cognitive, and related behavioral responses. In mainstream media and popular culture, “the psychological effects of color” is assumed to be a causal link that exists on a universal basis irrespective of individual differences or the impact of cultural, contextual, or temporal factors [1, p. 9]. This understanding about color psychology remains fairly consistent; however, some sources extend the definition to include a larger gamut of responses from ► [color preference](#) to precognitive, psychophysical, and “biology-based responses,” defined as responses on human metabolism, circulation, and respiratory systems [2, p. 92].

The word “therapy” as in “color therapy” generally refers to a remediation of a health or psychological issue intended to address a diagnosed problem or stop a medical or psychological condition from occurring. Hence, color therapy implies that color can be used as a therapeutic device in the remediation of a health or psychological issue. Some writers take this concept further and suggest that the “biological consequences of color responses can be a valuable tool in health management” for the treatment of various ailments, and this appears to be a fairly common understanding of the term [2, p. 93], while other writers suggest that color therapy, or chromotherapy, can be used prescriptively as a “truly holistic, non-invasive and powerful therapy which dates back thousands of years” [3, p. 1].

The Interface Between Color and Human Response: Recent Research

Academic literature includes a broad range of studies that discuss the effects of light as well as the influence of colored light waves in respect to human response. While human vision is a complex and not yet fully understood process, the receptor system for detecting light has been found to be different from that associated with the circadian cycle [4]. In reference to human circadian cycles, light tends to regulate our sleeping patterns and changes in light–dark exposure can desynchronize the circadian cycle affecting the ability to sleep and wake, as well as impacting on physiological and metabolic processes. Disruptions to the circadian rhythm may result in changes in mood and behavior as evidenced by studies that focus on seasonal affective disorder, known by the acronym SAD (seasonal affective disorder) [5–7]. Light has also been found to have an effect on the human neuroendocrine system and may also suppress melatonin and elevate cortisol production, which can have negative impacts [7, 8]. Furthermore, a number of recent studies have indicated that certain wavelengths of light may have specific impacts. For example, blue light may improve cognitive performance; different colored lenses may assist with reading difficulties such as dyslexia; and the human circadian system may be particularly sensitive to short-wavelength light [9–11]. Despite the many advances in recent research, the precise roles of the rods and cones of the retina as well as melanopsin in the control of circadian cycles remain to be determined [12].

A plethora of studies exist which suggest that color may influence a range of psychological, physiological, and behavioral responses; however, the range and diversity of research findings was highlighted by an analysis of 30 studies [13]. For example, it has been suggested that red has a greater capacity for arousal than blue [14–16]. However, findings from a more recent study suggest that there is no statistically significant difference between these two colors in terms of physiological arousal and that it may be hue rather than saturation (intensity) of color that has an impact [17]. In addition, recent studies have

found that responses to color may vary depending on age, gender, culture, and preference [18, 19]. It is important to note that while many of the recent studies that focus on psychological, physiological, and behavioral responses are scientifically rigorous, the findings are often based on an extremely limited range of color samples or a small sample group. In terms of color therapy, while associations may exist between various colors and a range of different human responses, it does not necessarily follow that such colors can be effectively used in therapy as some sources in popular media suggest.

Color Psychology in Popular Media

Information about color psychology in popular media is abundant and accessible via myriad magazines, online magazines, and Internet websites. Internet search engine Google currently provides access to 365,000 websites for information relating to color psychology. For example, www.about.com and www.colortheraPyhealing.com provide numerous pages on color psychology and color healing, respectively, with the latter providing fairly detailed information about the use of color as a therapeutic tool and a range of color therapy workshops. Similarly, www.colour-affects.co.uk offers detailed comment about the so-called psychological properties of color and provides a summary of four personality types supposedly linked to specific colors. In addition, mass media magazines and online magazines such as *WellBeing* (www.wellbeing.com.au) feature various articles on color psychology and color therapy.

The information available from popular, mainstream media can vary from broad, generalized commentary to detailed, pseudoscientific discussions peppered with relatively harmless generalizations, platitudes, and motherhood statements, at best, and unsubstantiated and highly dubious claims, at worst, such as “colors are the mother tongue of the subconscious” and “color heals.”

Some sources of information provide claims about perceptual color effects and color associations intermingled with unsubstantiated claims about color psychology such as:

Although red, yellow and orange are in general considered high-arousal colors and blue, green and most violets are low-arousal hues, the brilliance, darkness and lightness of a color can alter the psychological message. While a light blue-green appears to be tranquil, wet and cool, a brilliant turquoise, often associated with a lush tropical ocean setting, will be more exciting to the eye. The psychological association of a color is often more meaningful than the visual experience.

Colors act upon the body as well as the mind. Red has been shown to stimulate the senses and raise the blood pressure, while blue has the opposite effect and calms the mind. People will actually gamble more and make riskier bets when seated under a red light as opposed to a blue light. That's why Las Vegas is the city of red neon. [20]

Other sources of information suggest that color has therapeutic effects such as the claims provided in the online magazine *Conscious Living*.

Colour can repair and heal the body, when the frequency of the colour aligns with the emotion needed to activate the micro-particulars so healing can take place. . . The use of colour in visualisation is most effective, and easiest for the novice to utilise, as colour has a very strong radiating effect on the whole body. Every other form of colour therapy is fundamentally symbolic. [21]

In a similar vein, Wright suggests that

In practice, colour psychology works on two levels: the first level is the fundamental psychological properties of the eleven basic colours, which are universal, regardless of which particular shade, tone or tint of it you are using. Each of them has potentially positive or negative psychological effects and which of these effects is created depends on personality types and—crucially—the relationships within colour combinations, the second level of colour psychology. [22]

Wright offers further clarification from the colour affects system, and it should be noted that the colour affects site offers color consultancy services plus a range of industry and personal courses and workshops.

Similarly, Rewell provides a detailed discussion about specific responses to color in the online magazine *WellBeing* as follows:

Babies cry more in yellow rooms. Tension increases in people in yellow rooms and people who drive yellow cars are more prone to become aggravating in heavy traffic. . . Spend time exposed to a lot of yellow and you'll feel like time has sped up. . . A rejection of yellow indicates a fear of change. . .

Red stimulates the physical and adrenalin. It raises blood pressure, the heart rate and respiration. [23]

Other sources of information about color psychology include architectural and interior design books (e.g., see Kopacz and Mahnke) as well as technical reports [2, 24, 25]. In addition, various courses and workshops discuss issues relating to color psychology including the Colour Therapy Healing organization in the UK, the International School of Colour and Design in Sydney, and the Nature Care College in Sydney [26–28].

Information from sources such as those discussed above often quotes a range of color associations co-mingled with various psychological, cognitive, or behavioral responses to color. In addition, such sources rarely if ever provide evidence to substantiate any claims regarding the impact of color on human response. An example of the intermingling of a range of different types of response is provided by Cherry as follows:

Red is a bright, warm color that evokes strong emotions. Red is also considered an intense, or even angry, color that creates feelings of excitement or intensity.

Blue calls to mind feelings of calmness or serenity. It is often described as peaceful, tranquil, secure, and orderly. Blue can also create feelings of sadness or aloofness. Blue is often used to decorate offices because research has shown that people are more productive in blue rooms.

Green is restful, soothing, cheerful and health-giving. Green is thought to relieve stress and help heal. Those who have a green work environment experience fewer stomach aches. Green has long been a symbol of fertility. [1]

Some sources go so far as to state symbolic color associations and then extend these further to include different situations or contexts such as the following:

Red is the colour for courage, strength and pioneering spirit... It is the colour of anger, violence and brutality.

Blue is calming, relaxing and healing (but) not as sedentary as indigo.

Green is the colour of balance and harmony and can, therefore, be helpful in times of stress. If one has experienced trauma, a green silk wrapped around the shoulders can have a very therapeutic effect. [3]

It is not uncommon for claims such as those above to be presented in an authoritative manner, exhorting the reader to believe and accept; however, the lack of evidence indicates that such claims are often personal opinion masquerading as scientific fact or pseudoscience.

An Irrefutable, Causal, and Universal Link Between Color and Human Response?

Various sources of information in mainstream media and related sources tend to suggest that an irrefutable, causal relationship exists between color and human response, as illustrated by Hill's statement: "Based on numerous studies by Drs Morton Walker, Gerard and Faber Birren, the link between color and physiological responses has been well documented" [24, p. 7].

In addition, many sources imply that the link between color and human response is universal irrespective of individual or cultural differences. For example, Logan-Clarke asserts: "Red... is stimulating and energizing therefore it is helpful for tiredness and lethargy, to stimulate low blood pressure, to boost sluggish circulation... Red is energizing and excites the emotions, and can stimulate the appetite" [3, p. 10]. Similarly, Rewell contends "Red stimulates the physical and adrenalin. It raises blood pressure, the heart rate and respiration" [23], while Kopacz suggests "Red is believed to sensitise the taste buds and sense of smell, increasing the appetite... all this occurs because the heart rate instinctively quickens, which causes a release of adrenalin into the bloodstream raising blood pressure and stimulating the nerves" and "the sight of the color blue causes the body to release tranquilizing hormones when it is surveyed, particularly a strong blue sky" and "many believe (blue) can lower blood pressure, slow the pulse rate and decrease body temperature" [2, pp. 76, 79].

In regard to such claims, Kopacz notes the lack of evidential support but offers the work of a number of authors and designers including Birren and Mahnke as well as Morton Walker (author of *Bald No More, Foods for Fabulous Sex, Your Guide to Foot Health*, as well as *The Power of Color*) and Wright (color consultant of www.colour-affects.co.uk) to support his color

psychology claims. Similarly, Mahnke notes the inconclusive nature of findings from research that focuses on the interface between color and human response, but nevertheless refers to the work of Birren and Goldstein among others as support for various psychophysiological effects of color.

In regard to color therapy, it is suggested that color can be used as a treatment tool in conjunction with the seven chakras of the body [2, 3, 25]. The concept of chakras, considered to be energy centers within the human body, belongs to a belief system originating in the Hindu scriptures known as the Upanishads, dating from the first millennium BCE. Under the color chakra theory, a color is linked to each of the seven chakra and these colors are associated with bodily functions and dysfunctions within each chakra area. For example:

Red: Activates the circulation system and benefits the five senses; used to treat colds, paralysis, anaemia, ailments of the bloodstream and ailments of the lung;

Blue: Raises metabolism; is used to stabilize the heart, muscles and bloodstream; used to treat burns, skin diseases, glaucoma, measles and chicken pox, and throat problems;

Green: Strengthens bones and muscles, disinfects bacteria and virus, and relieves tension; used to treat . . . malaria, back problems, cancer, nervous disorders, and ulcers, and to manage heart problems and blood pressure. [2, p. 93]

The allocation of colors to each of the chakras is reminiscent of the doctrine of the four color-linked humors of the body from ancient Greek medicine: black bile, yellow bile, green phlegm, and red blood. The linking of color with the humors, the four elements (earth, fire, water, and air) as well as the seasons, represented the color correspondence beliefs that emerged in antiquity and continued through to the Renaissance as evidenced by the color correspondences depicted in the engraving by Nicoletto Rosex [29].

Aside from ancient belief systems, much of the information currently available in popular media about color psychology echoes, if not directly quotes, the work of a number of key theorists, including Faber Birren, Robert Gerard, Kurt Goldstein, and Max Lüscher (e.g., see [1–3, 22, 25, 30–32]).

Birren, Gerard, Goldstein, and Lüscher

Extensively quoted by recent authors, Birren published more than 40 books and over 250 articles on color psychology, color therapy, as well as color application. Tunney advises that Birren was a leading authority on color in the mid-twentieth century who was retained as a color consultant by DuPont, General Electric, Sears, Roebuck and Company, and the United States Navy [33]. Birren championed an unambiguous, irrefutable, and universal causal link between color and human response, and he cites Goldstein's assertion that "It is probably not a false statement if we say that a specific color stimulation is accompanied by a specific response pattern of the entire (human) organism" [34, p. 144]. Given his role as color consultant, it was in his interests to champion such a link as he was able to then provide advice addressing this link in a range of commercial contexts.

Goldstein, who published *The Organism* in 1939, was considered a significant authority on the psychological and physiological impacts of color by Birren, who quoted Goldstein in his *Color Psychology and Color Therapy* publication: "under the influence of red light, time is likely to be overestimated. Conversely, under the influence of green or blue light, time is likely to be underestimated" and "under red light, weights will be judged as heavier; under green light they will be judged lighter" [35, p. 211]. Birren as well Mahnke cites Goldstein's 1942 study, which discussed the perceived stimulating effects of red and the opposite effects of green [16].

Gerard, whose 1958 doctoral thesis ("Differential effects of colored lights on psychophysical functions") was subsequently presented as a conference paper ("Color and emotional arousal"), is also frequently cited by Birren and Mahnke [15, 36]. Gerard's key findings from his study on the arousal properties of red, blue, and white illumination, as reported by Wise, Wise, and Beach, include: "statistical differences between red-blue (illumination) conditions for all physiological measures except heart rate . . . responses to the white light varied but most often were similar to those of the red condition" [37, p. 5]. Wise et al. also note that Gerard, whose study involved

a small sample of only 24 male university students, advised caution in terms of the generalizability of his findings. It is unfortunate that Gerard's advice was not reported by subsequent authors.

Lüscher, who developed and published the Lüscher Color Test, included four basic colors (which he referred to as psychological primaries) in the color test: orange-red, bright yellow, blue-green, and dark-blue, as well as four auxiliary colors – violet, brown, black, and neutral gray [38]. Citing anecdotal evidence, Lüscher assigned specific associations and affective characteristics to each color. For example, (red) has a “stimulating effect on the nervous system – blood pressure increases, respiration rate and heartbeat both speed up”; while blue “has the reverse effect – blood pressure falls, heartbeat and breathing both slow down” [38, p. 12]. The Lüscher Color Test is essentially a ► [color preference](#) test and its use in personality testing and assessment has been soundly cautioned [39].

While the works of Birren, Gerard, Goldstein, and Lüscher have their place in the literature on color, it is important to note that their theories have, to a large extent, been debunked due to methodological shortcomings. In addition, more recent and methodologically more rigorous studies have superseded these earlier authors, and this, plus a number of other reasons discussed below, provides a compelling case for applying the principle of caveat emptor to information about color psychology and color therapy found in mainstream media.

Color Psychology and Color Therapy: Caveat Emptor

Latin for “buyer beware,” the term caveat emptor suggests that in the absence of a warranty, the buyer is at risk, and therefore onus for carefully assessing goods and services prior to purchase remains with the buyer. There are a number of reasons why the principle of caveat emptor should prevail in regard to information about color psychology and color therapy.

Firstly, the existence of an irrefutable and universal causal link between color and an unlimited range of psychological, biological, and behavioral

responses remains a largely unsupported hypothesis. While numerous studies exist that focus on the interface between color and human response, a systematic and critical review of over 200 studies was conducted for the National Aeronautics and Space Administration (NASA) in 1988. This study concluded that “there are no ‘hard-wired’ linkages between environmental colors and particular judgmental or emotional states” [37, p. 46].

In addition, while a number of studies exist which focus on the interface between color and human response, the findings of many of these studies are limited by the setting and context of the study, the size and composition of the sample group, as well as the limited range of colors used in the stimuli. Extrapolating beyond the data in such studies is methodologically unsound, especially when small sample sizes are used [40]. Very small sample groups were used by many early studies including the often-quoted studies of Goldstein and Gerard:

- Sample size of 3–5 [16]
- Sample size of 24 [15]
- Sample size of 48 [41]
- Sample size of 14 [42]
- Sample size of 25 [43]

In addition, many early studies focused on the color attribute of hue alone without any consideration of the attributes of tonal value or saturation, and many research studies used small colored squares of cardboard as stimuli and the findings from such studies were extrapolated to a whole range of other contexts or situations. The findings of such studies have extremely limited relevance due to the methodological weakness of studying a complex and highly subjective phenomenon as color in isolation with limited stimuli.

Ancient wisdom is often cited as a reason to believe claims about color psychology and color therapy, with the underlying logic implying that ancient wisdom embedded within such claims represents evidential proof. For example, Kopacz cites the link between color and chakras of ancient Hindu scriptures as evidence, while Van Wagner suggests that chromotherapy was practiced by ancient cultures, including the Egyptians and

Chinese [1, 2]. Such authors imply that not only is the wisdom of the ancients above question but it provides evidential proof of the veracity of various color-related claims. Clearly, ancient belief systems are not always a guarantee of veracity as evidenced by the superseded beliefs that the world is flat, and the sun and moon orbit the earth. Without disparaging the wisdom of the ancients, it is a fair comment that a proportion of ancient wisdom has been superseded by later scientific discovery. The existence of a link with ancient wisdom should not of itself be used as evidential proof of any claim regarding color psychology and color therapy.

Factoids are not facts, although many authors who write about color psychology and color therapy would have us believe otherwise. Coined by author Norman Mailer, factoids are suppositions or inventions presented as fact, and a comedic illustration is the line in the film *Anchorman*: “It’s anchorman not anchor-lady, and that’s a scientific fact!” [44]. Color psychology and color therapy information available in popular culture often appears as factoids rather than facts, especially when nil evidence is cited to support claims and assertions. Examples of unsubstantiated factoids include: “We are hard-wired to yellow as a stimulus. . . . If you’re environment is boring and time passes slowly, surround yourself with small amounts of yellow” and “People who dislike yellow often favour blue to calm themselves and feel secure. If you drink coffee for a pick-me-up, try drinking it from a yellow cup” [23, p. 32]; similarly, “Yellow can also create feelings of frustration and anger. While it is considered a cheerful colour, people are more likely to lose their tempers in yellow rooms and babies tend to cry more in yellow rooms” [1].

The use of control groups, measures to control internal and external validity, and adequate levels of statistical significance are absent from some studies that focus on the interface between color and human response, and these impact negatively on the scientific rigor of such studies. It is empirically and methodologically unsound to draw generalized conclusions from such studies and transfer them to different settings and contexts [40, 45–47].

The fallacy of the single cause provides another reason to be cautious about claims regarding color psychology and color therapy. This fallacy suggests that one single cause for an outcome represents causal oversimplification, and under the post-positivist paradigm, recent theorists such as Hård and Sivik consider the interface between color and human response to be highly complex and open to the influence of a wide range of factors and mediating variables [48]. While correlation may indicate the existence of an association between one variable (such as color) and human response, correlation does not imply causation and suggesting a causal link without allowing for mediating variables is considered empirically and methodologically unsound [40, 45, 47].

Inherent bias, subjective validation, and data cherry picking in research studies always call into question the findings of such studies [40]. Specifically, “Scientists have a vested interest in promoting their work,” and this is often the case for color theorists and researchers whose publications support their role as consultants [40, p. 337]. In addition, subjective validation is evident in many studies that focus on the interface between color and human response. This occurs when two unrelated events are judged to be related because of an expectation of such or because an existing hypothesis demands such a relationship. For example, to apply the notion that “green relieves stress” to the following claims: “green silk wrapped around the shoulders can have a very therapeutic effect” and “Those who have a green work environment experience fewer stomach aches” indicates subjective validation and renders such claims invalid.

It is unwise to ignore any mediating variables that may impact a research study and this is particularly relevant to research that focuses on the interface between color and human response. In this context, mediating factors include an individual’s personality, personal bias and feelings, as well as cultural experiences and affective state. Mehrabian found that individual differences exist in terms of stimulus screening ability, and high screeners are able to automatically screen out less important components of environmental

stimuli such as color and sound as opposed to low screeners [49]. Stimulus screening ability is rarely if ever mentioned in research studies that focus on color and human response nor is an individual's own bias in terms of whether or not they consider color to have an impact on their responses to color.

To conclude, color psychology claims information found in mainstream media suggests that color prompts a range of different human responses: psychological, biological, and behavioral. Many of these claims lack substantiation in terms of empirical support, exhibit fundamental flaws (such as causal oversimplification and subjective validation), and are generally personal opinion and factoids masquerading as fact. Color psychology is a discipline that is not accepted within the medical profession and exists on the periphery of alternative medicine and therapies. It is therefore recommended that the principle of caveat emptor is applied when evaluating information about color psychology plus check with a medical practitioner or psychologist before acting upon such information.

Cross-References

- [Color Preference](#)
- [Primary Colors](#)

References

1. Cherry, K.: Color psychology: how colors impact moods, feelings and behaviours. About.com: Psychology. <http://psychology.about.com/od/sensationandperception/a/colorpsych.htm> (2014). Accessed 4 Jan 2014
2. Kopacz, J.: Color in three-dimensional design. McGraw-Hill, New York (2003)
3. Logan-Clarke, V.: What is colour therapy? <http://www.colourtherapyhealing.com/> (2014). Accessed 4 Jan 2014
4. Brainard, G.C., Hanifin, J.P., Greenson, J.M., Byrne, B., Glickman, G., Gerner, E., et al.: Action spectrum for melatonin regulation in humans. Evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**(16), 6405–6412 (2001)
5. Harmatz, M.G., Well, A.D., Overtree, C.E., Kawamura, K.Y., Rosal, M., Ockene, I.S.: Seasonal variation of depression and other moods: a longitudinal approach. *J. Biol. Rhythms* **15**(4), 344–350 (2000)
6. Kasper, S., Wehr, T.A., Bartko, J.J., Gaist, P.A., Rosenthal, N.E.: Epidemiological findings of seasonal changes in mood and behavior. *Arch. Gen. Psychiatry* **46**, 823–833 (1989)
7. Stevens, R.G., Blask, D.E., Brainard, G.C., Hansen, J., Lockley, S.W., Provencio, I., et al.: Meeting report: the role of environmental lighting and circadian disruption in cancer and other diseases. *Environ. Health Perspect.* **115**(9), 1357–1362 (2007)
8. Skene, D.J., Lockley, S.W., Thapan, K., Arendt, J.: Effects of light on human circadian rhythms. *Reprod. Nutr. Dev.* **39**(3), 295–304 (1999)
9. Irlen, H.: Reading problems and Irlen coloured lenses. *Dyslexia Rev.* **8**(3), 4–7 (1997)
10. Lehl, S., Gerstmeier, K., Jacob, J.H., Frieling, H., Henkel, A.W., Meyer, R., et al.: Blue light improves cognitive performance. *J. Neural Transm.* **114**, 457–460 (2007)
11. Warman, V.L., Dijk, D.J., Warman, G.R., Arendt, J., Skene, D.J.: Phase advancing human circadian rhythms with short wavelength light. *Neurosci. Lett.* **342**, 37–40 (2003)
12. Webb, A.: Considerations for lighting in the built environment: non-visual effects of light. *Energy Build.* **38**, 721–727 (2006)
13. Mikellides, B.: Emotional and behavioural reactions to colour. In: Sivik, L. (ed.) *Colour and psychology: from AIC interim meeting 96 in Gothenburg. Colour report F50*. Scandinavian Colour Institute, Stockholm (1997)
14. Ali, M.R.: Pattern of EEG recovery under photic stimulation by light of different colors. *Electroencephalogr. Clin. Neurophysiol.* **33**, 332–335 (1972)
15. Gerard, R.M.: Differential effects of colored lights on psychophysiological functions. University of California, Los Angeles, Unpublished Ph.D. thesis (1958)
16. Goldstein, K.: Some experimental observations concerning the influence of colors on the function of the organism. *Occup. Ther. Rehabil.* **21**(3), 147–151 (1942)
17. Mikellides, B.: Color and psychological arousal. *J. Archit. Plann. Res.* **7**(1), 13–20 (1990)
18. Manav, B.: Color-emotion associations and color preferences: a case study for residences. *Color Res. Appl.* **32**(2), 144–150 (2007)
19. Ou, L.C., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. Part III: colour preference modelling. *Color Res. Appl.* **29**(5), 381–389 (2004)
20. Pantone: color psychology. <http://www.pantone.com/pages/pantone/Pantone.aspx?pg=19382&ca=29> (2014). Accessed 10 Jan 2014
21. Campbell, L.: Mind body healing – how it works. <http://www.consciousliving.net.au/magazine> (2009). Accessed 11 Apr 2009
22. Wright, A.: Psychological properties of colours. <http://www.colour-affects.co.uk/how-it-works> (2014). Accessed 11 Jan 2014
23. Rewell, C.: What do colours mean? Part II. <http://www.wellbeing.com.au/article/features/wisdom/What-do-colours-mean-Part-II> (2009). Accessed 11 Jan 2014

24. Hill, T.R.: Using color to create healing environments (report prepared for Dupont Corian). Little Fish Think Tank/Dupont, Atlanta (2008)
25. Mahnke, F.: Color, environment and human response. Wiley, New York (1996)
26. CTH: Colour therapy healing workshop. <http://www.colourtherapyhealing.com> (2009). Accessed 10 Apr 2009
27. ISCD: International school of colour and design. <http://www.iscd.edu.au/default2.asp> (2009). Accessed 10 Apr 2009
28. NCC: Nature care college. <http://naturecare.com.au> (2009). Accessed 10 Apr 2009
29. Gage, J.: Colour and culture. Thames & Hudson, London (1995)
30. Feisner, E.A.: Colour: how to use colour in art and design. Laurence King, London (2000)
31. Holtzschue, L.: Understanding color. Wiley, New York (2006)
32. Stone, T.L., Adams, S., Morioka, N.: Color design workbook. Rockport, Beverly (2006)
33. Tunney, S.: Guide to the Faber Birren papers. <http://mssa.library.yale.edu/findaids/stream.php?xmlfile=mssa.ms.1567.xml#biogFull> (2006). Accessed 11 Apr 2009
34. Birren, F.: Color psychology and color therapy. Citadel, Secaucus (1961)
35. Goldstein, K.: The organism: a holistic approach to biology derived from pathological data in man. Zone, New York (1995 [1939])
36. Gerard, R.M.: Color and emotional arousal (Abstract from the program of the 66th annual convention of the American Psychological Association). *Am. Psychol.* **13**(7), 340 (1958)
37. Wise, B.K., Wise, J.A., Beach, L.R.: The human factors of color in environmental design: a critical review. NASA Grant No. NCC 2-404. NASA Ames Research Centre, Moffett Field (1988)
38. Lüscher, M.: The Lüscher color test. Random House, New York (1969)
39. Walters, J., Apter, M.J., Svebak, S.: Color preference, arousal, and the theory of psychological reversals. *Motiv. Emot.* **6**(3), 193–215 (1982)
40. Sutherland, W.J., Spiegelhalter, D., Burgman, M.A.: Twenty tips for interpreting scientific claims. *Nature* **503**, 335–337 (2013)
41. Nakshian, J.S.: The effect of red and green surrounding on behavior. *J. Gen. Psychol.* **70**(1), 143–161 (1964)
42. Nourse, E.W., Welch, R.B.: Emotional attributes of color: a comparison of violet and green. *Percept. Mot. Skills* **32**(2), 403–406 (1971)
43. Goodfellow, R.A., Smith, P.C.: Effects of environmental color on two psychomotor tasks. *Percept. Mot. Skills* **37**(1), 296–298 (1973)
44. Ferrell, W., McKay, A., Apatow, J., McKay, A.: Anchorman: the legend of Ron Burgundy. DreamWorks Pictures, Los Angeles (2004)
45. Argyrous, G.: Statistics for social and health research. Sage, London (2001)
46. Campbell, D.T., Stanley, J.C.: Experimental and quasi-experimental designs for research. Houghton Mifflin, Boston (1963)
47. Coolican, H.: Research methods and statistics in psychology, 4th edn. Hodder & Stoughton, London (2004)
48. Hård, A., Sivik, L.: A theory of colors in combination – a descriptive model related to the NCS color-order system. *Color Res. Appl.* **26**(1), 4–28 (2001)
49. Mehrabian, A.: Individual differences in stimulus screening and arousability. *J. Pers.* **45**, 237–250 (1977)

Color Qualia

► Color Phenomenology

Color Range

► Palette

Color Roles

► Functionality of Color

Color Scene Statistics, Chromatic Scene Statistics

Yoko Mizokami

Graduate School of Advanced Integration
Science, Chiba University, Chiba, Japan

Synonyms

Chromatic image statistics; Color environment;
Color image statistics; Color statistics; Natural
color distribution; Natural scene statistics

Definition

Color scene statistics or chromatic scene statistics are statistical characteristics of scene color. There

are a number of ways to analyze color scenes (natural scenes) statistically in relation to vision, such as the average color, color distribution, Bayesian model, principal component analysis, and probabilistic models. Color scene statistics are closely related to the evolution and development of the color vision mechanism at every level of the visual system.

Overview

Color vision has evolved with the natural environment. Thus, the color statistics of a natural scene must have an enormous impact on the evolution and development of the color vision mechanism. There have been many attempts to analyze natural scene statistics and find the connection to the color vision mechanism as well as other aspects of the visual mechanism. Traditionally, the average color and color distribution of a scene are associated with color adaptation and color constancy. Based on those statistical data, physiological, empirical, and statistical models of color constancy were introduced. Other aspects of visual perception, such as color discrimination or color appearance, would be formed by the environment to obtain the best performance or representation under a given color environment. Measurement techniques, such as two- and three-dimensional measurements of scenes and multispectral measurement, provide additional information for deeper analysis or analysis from different perspectives. Powerful computational analyses introduce new statistical approaches for investigating the mechanism of color vision. Here, the topics of color scene statistics related to color vision, visual perception, and their evolution are discussed.

Measurement of Objects, Illumination, and Scenes

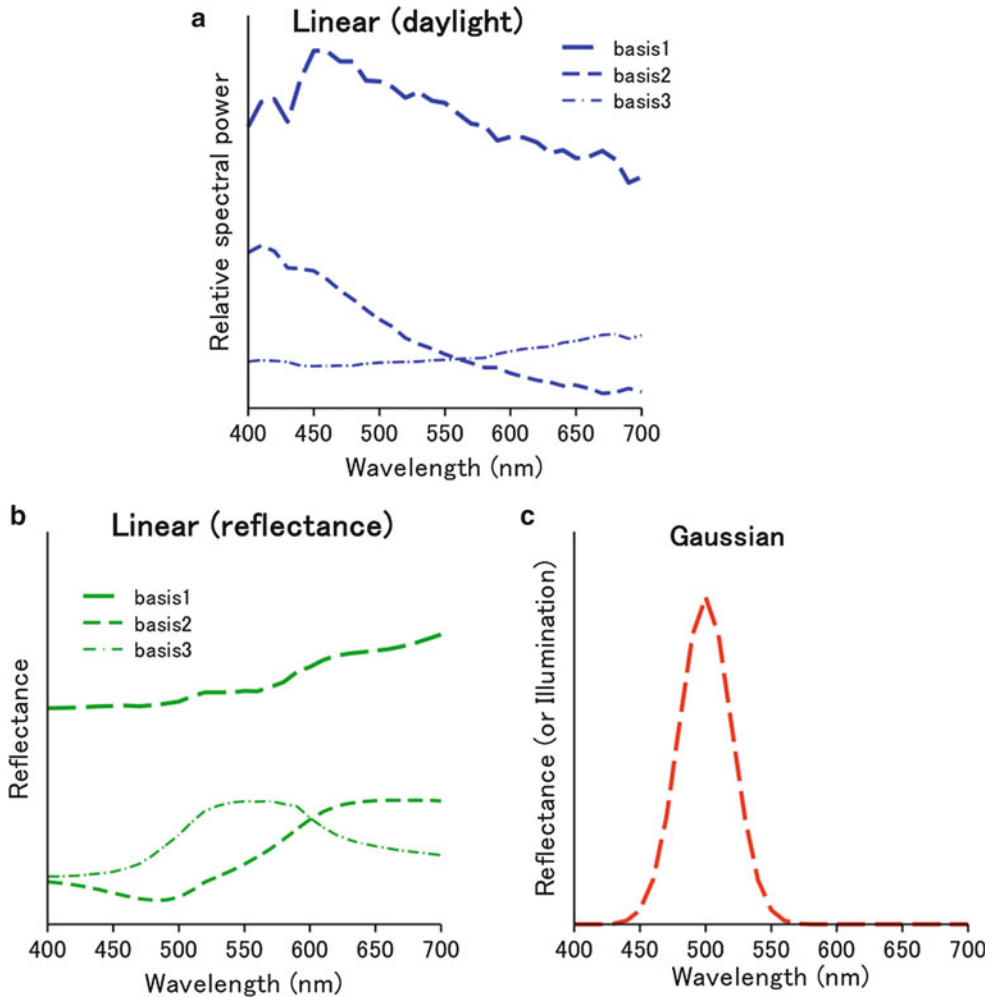
Obtaining accurate color information of scenes is important to analyze color scene statistics. Camera calibration techniques and calibrated images in which the RGB values of images are

transferable to information on chromaticity coordinates have been developed.

Another important aspect for analyzing colors in natural scenes is the spectral information. RGB information has only three dimensions, and consequently, some information is lost. For instance, RGB information cannot differentiate metameric color, which results when a pair of stimuli exhibit the same apparent color but with different spectral power distributions. Many groups have measured the daylight spectra and spectral reflectance of natural and printing surfaces. Published databases (e.g., SOCS [1]) and personal databases on websites are available.

For obtaining the spectral color information of an entire scene, multispectral or hyperspectral image capture techniques using a multispectral camera are used. Those data are used for studies on color vision mechanisms. It is difficult, however, to obtain a hyperspectral image manually because it takes a long time and the scene must be still (no moving objects or wind) to capture an image with multiple wavelengths by changing bandpass filters. Therefore, collections of multispectral images of outdoor scenes are limited. Recent development of fast and/or accurate multispectral cameras will allow further statistical analysis based on a large number of multispectral data sets.

Although the multispectral information is important and useful, it includes a massive amount of data that is inconvenient to handle. Thus, there have been many attempts to represent multidimensional data with a smaller number of dimensions by compressing the amount of information using principal component analysis (PCA). The natural spectra can be represented well by three basis functions. Those for the illuminant obtained by Judd [2] are shown in Fig. 1. The basis functions give the relative spectral radiant power distribution of the CIE daylight illuminant at different color temperatures. Cohen [3] applied this analysis to the reflectance of Munsell color paper and showed three basis functions that were different from those obtained by Judd. Later studies suggested that Cohen's data were applicable to natural surfaces (see Fig. 1) and that three basis functions were necessary and probably



Color Scene Statistics, Chromatic Scene Statistics, Fig. 1 Examples of three basis functions for daylight (left) and reflectance (center), respectively. A Gaussian

fitting function with three parameters: peak, amplitude, and bandwidth (for both illumination and reflectance) is also shown (right)

sufficient for representing the spectral reflectance functions of natural objects. (Note that some studies suggested that a larger number of basis functions are needed depending on the context.) These results led to the hypothesis that the human visual mechanism has a similar representation in the visual process [4]. However, it was recently suggested that Gaussian fitting with three variables to natural spectra (see Fig. 1) is as good as the fitting by three basis functions [5]. More investigations would be needed to determine how human visual system represents the spectral world.

Analysis of Natural Color Statistics

There are a number of ways to analyze the color properties of a scene statistically. The most basic method would be by average color and color distribution. Analysis taking into account spatial frequency is also important because the spatial contrast sensitivities of luminance, and the L-M (reddish–greenish) and S-LM (bluish–yellowish) opponent-color channels are different. Those of chromatic channels tend to have a low-pass shape, suggesting the low-frequency component contributes more to color perception. Recent

developments in statistical modeling, along with powerful computational tools, have enabled researchers to study more sophisticated statistical models of visual images and to evaluate these models empirically against large data sets [6].

Based on the different information obtained from color scene statistics, the mechanisms of vision and color perception have been investigated and revealed from aspects of the peripheral to central visual system to perspectives on evolution, development, and short- to long-term adaptation.

Effect on Evolution of Color Vision

The sensitivity of the photoreceptors and how the cone signals are combined in postreceptoral channels have been explained by assuming that they optimize the efficient coding of natural color signals and that they are tuned to specific signals.

For example, it has been suggested that trichromacy in Old World primates and the reflectance functions of tropical fruits coevolved and that primate color vision has been shaped by the need to find reddish (ripe) fruit or young (edible) leaves among green foliage, as shown by the example in Fig. 2. For instance, the tuning of the L and M cones and the later opponent-color system (L-M) based on the signal representing their difference are optimized to detect the color signals provided by ripening fruit or edible foliage [7]. This hypothesis is a good example of how color statistics of the visual environment affected the early-stage evolution of the color vision system.

It has also been shown that seasonal variations in the color statistics of natural images alter both the average color and color distribution in scenes, as shown in Fig. 3. On opponent-color space, arid periods are marked by a mean shift toward the + L pole of the L-M chromatic axis. A rotation in the color distributions away from the S-LM chromatic axis and toward an axis of bluish–yellowish variation, both primarily due to changes in vegetation, implies that these changes contribute to the construction of the mechanism of both visual sensitivity and color appearance [8].

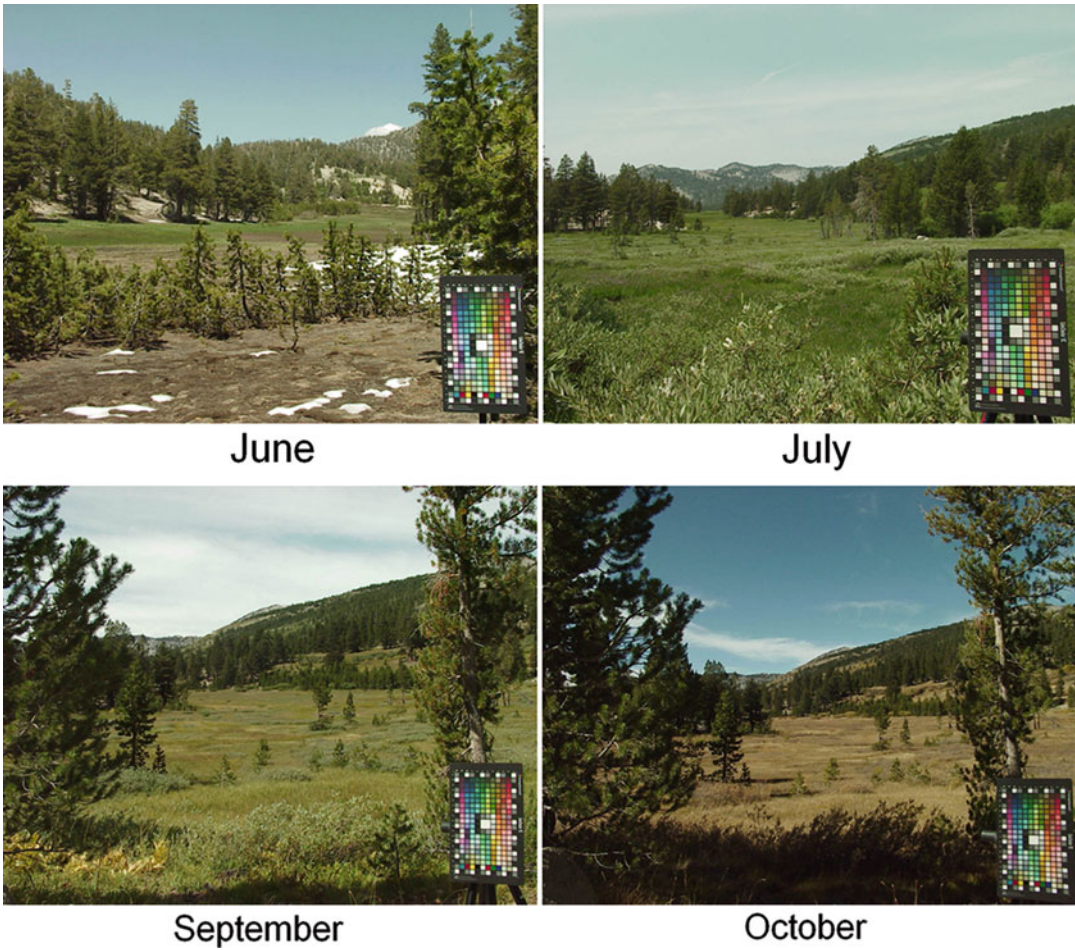


Color Scene Statistics, Chromatic Scene Statistics, Fig. 2 Examples of the advantage of trichromacy (*top*), capable of *red–green* discrimination, compared to dichromacy (*bottom*) on finding *red* fruits among *green* leaves

There are other hypotheses, such as skin tone discrimination and predator detection affecting evolution. There are a number of variations or changes in natural environments other than the examples shown above, which may contribute to formation of the color vision mechanism. It is interesting to pursue how the color vision mechanism was optimized during evolution and development [9].

Cortical Mechanism and Probabilistic Approaches

Cortical color-coding is also thought to have evolved to represent important characteristics of the structure of color in the environment. Probabilistic modeling allows us to test experimentally



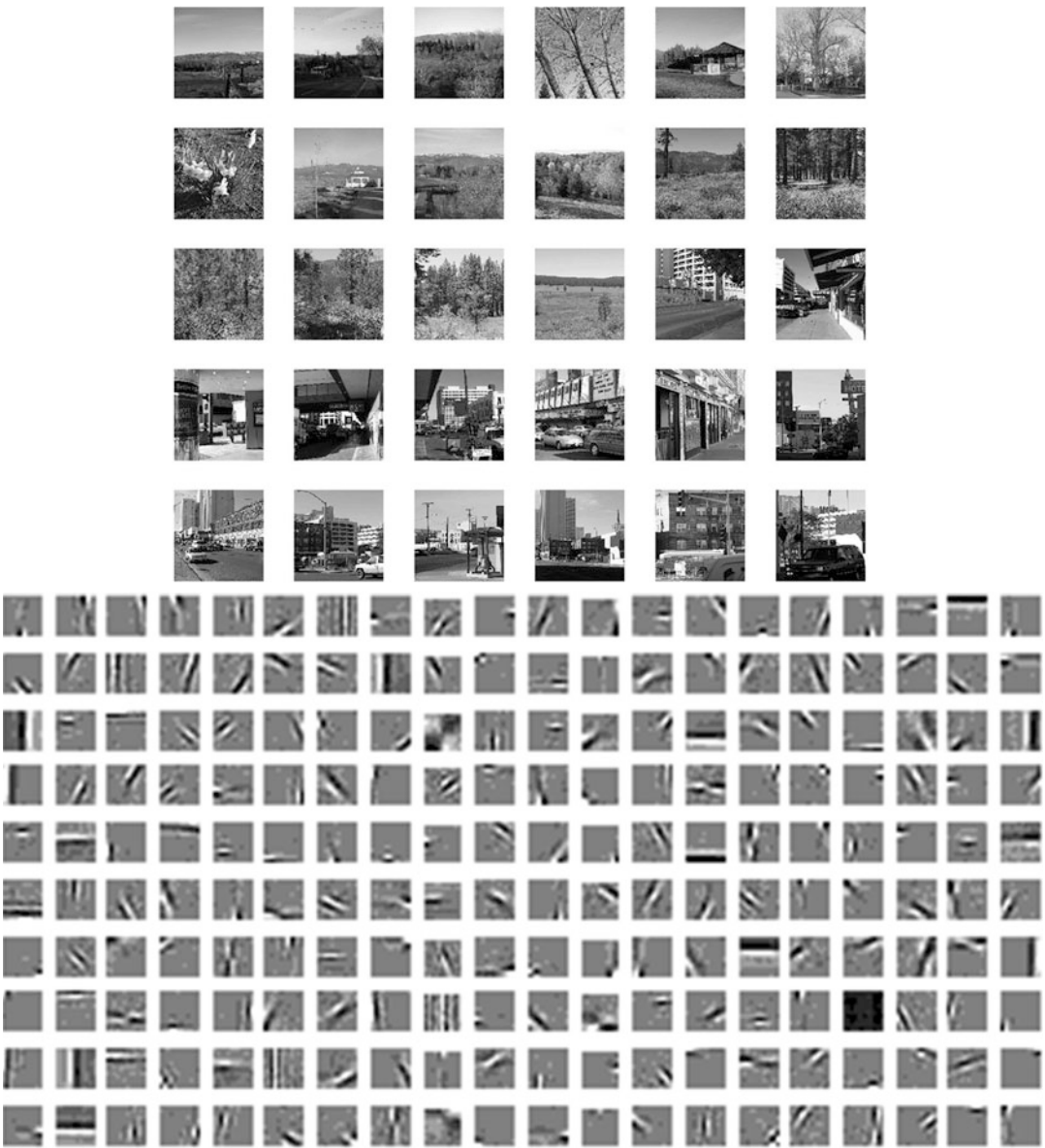
Color Scene Statistics, Chromatic Scene Statistics, Fig. 3 An example of seasonal changes in the natural color scene. (Mountain forest in Sierra Nevada)

the efficient coding hypothesis for both individual neurons and populations of neurons [6].

“Sparse coding,” the concept that neurons encode sensory information using a small number of active neurons at any given second, has been introduced. It increases the storage capacity in associative memories and saves energy. It also makes the structure in natural signals explicit and makes complex data easier to interpret during visual processing [10]. Figure 4 shows an example of a set of basis functions derived from a set of learning images including natural and man-made scenes based on a sparse coding model. Any natural image (scene) can be constructed using combinations of basis

functions. This suggests that the human visual system also constructs the image of a natural scene based on the activity of neurons having receptive fields with different properties.

Probabilistic models are also applied to the chromatic structure of natural scenes. Independent component analysis (ICA) has been applied to natural color images to establish an efficient representation of color in natural scenes. It was shown to produce achromatic and color-opponent basis functions as well as spatiochromatic independent components of cortical neurons. This suggests a relationship between statistics of natural scenes and cortical color processing.



Color Scene Statistics, Chromatic Scene Statistics, Fig. 4 An example of sparse coding created by a procedure in accordance with the method in the paper of

Olshausen and Field [10]. *Right*, image-set for learning; *Left*, derived basis functions

Effect of Color Statistics of a Scene on Color Constancy

Color constancy is a phenomenon in which the stability of the color appearance of an object surface is maintained among variations in the illuminating environment. The central issue is always how humans separate illumination and the surface

of objects, which is a necessary task for achieving color constancy. Identical to other mechanisms in earlier stages of the visual system, color constancy is also associated with the color statistics of environments [4, 9, 11].

A classic approach is the von Kries adaptation. To set a neutral point (or illumination color) for application of the von Kries adaptation, many

assumptions or hypotheses based on scene statistics have been proposed (e.g., average color of a scene: the gray world assumption). In Bayesian decision theory, which is a more statistical approach, the most likely combination of illuminants and surfaces is determined based on prior distributions that describe the probability that particular illuminants and surfaces exist in the world.

Characteristics of a scene statistics also contribute to color constancy [11, 12]. One question is how humans can tell the difference between a white paper under red illumination and a red paper under white illumination if the papers have the same chromaticity. It was suggested that the luminance of the white paper decreases statistically under the red light and that this relationship could be a clue for differentiating the two. Besides, in natural scenes, chromatic variations and the luminance variations aligned with them mainly arise from object surfaces such as the border between different materials, whereas pure or near-pure luminance variations mainly arise from inhomogeneous illumination such as shadows or shading. The human visual system uses these color–luminance relationships and determines the three-dimensional structure of a scene from the natural relationships that exist between color and luminance in the visual scene, material surface, and so on. It is suggested that natural scene statistics could also be the universal basis of color context effects, such as color contrast and color assimilation.

Adaptation to the Color of Natural Scenes

As shown in color constancy, it is well known that color perception adapts to the hue change of the color distribution. Color appearance is also influenced by the saturation or variance of the color distribution of a scene. A pale color patch appears less colorful when it is surrounded by saturated colors, and chromatically selective compression along any direction of chromatic variation occurs after adapting to temporal color variation. These suggest adaptation to the specific color gamut within individual scenes and natural

environments [9, 11]. It is also shown that the impression of a natural image shifts to being less colorful after adaptation to a series of saturated images and vice versa [13]. Moreover, it is suggested that simple statistics, such as the color distribution itself, cannot explain the effect because the adaptation is stronger for natural images than scrambled images. It implies the contribution of higher-level mechanisms, such as scene recognition or cognition.

The timescale of adaptation varies. It can be seconds, minutes, hours, months, years, and possibly a lifetime. The tuning of sensitivities of the color vision mechanism likely continues in all timescales. Evolution and development of the visual function could be considered very long-term adaptation. It is also suggested that more cognitive or social aspects of color perception, such as color category, could be influenced by the color scene statistics of each region. One of the functional meanings of adaptation would be compensating for variations within an observer and between observers as well as within an environment and between environments, in order to maintain a stable and coherent color appearance. The color scene statistics certainly contribute to it. The adaptation shaped by environmental pressure would achieve an efficient transmission and stability of information from the periphery in the visual system to the centers in the brain.

Cross-References

- [Appearance](#)
- [Color Constancy](#)
- [Environmental Influences on Color Vision](#)

References

1. Standard object colour spectra database for colour reproduction evaluation (SOCS). TR X 0012:2004 (ISO/TR 16066:2003), Japanese Standards Association (2004)
2. Wyszecki, G., Stiles, W.S.: *Color Science*. Wiley, New York (1982)
3. Cohen, J.: Dependency of the spectral reflectance curves of the Munsell color chips. *Psychon. Sci.* **1**, 369–370 (1964)

4. Foster, D.H.: Color constancy. *Vision Res.* **51**, 674–700 (2011)
5. Mizokami, Y., Webster, M.A.: Are Gaussian spectra a viable perceptual assumption in color appearance? *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **29**, A10–A18 (2012)
6. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001)
7. Osorio, D., Vorobyev, M.: A review of the evolution of animal colour vision and visual communication signals. *Vision Res.* **48**, 2042–2051 (2008)
8. Webster, M.A., Mizokami, Y., Webster, S.M.: Seasonal variations in the color statistics of natural images. *Network* **18**, 213–233 (2007)
9. Webster, M.A.: Adaptation and visual coding. *J. Vis.* **11**(5), 3, 1–23 (2011)
10. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996)
11. MacLeod, D.I.A., Golz, J.: A computational analysis of colour constancy. In: Mausfeld, R., Heyer, D. (eds.) *Colour Perception: Mind and the Physical World*, pp. 205–246. Oxford University Press, London (2003)
12. Shevell, S.K., Kingdom, F.A.A.: Color in complex scenes. *Annu. Rev. Psychol.* **59**, 143–166 (2008)
13. Mizokami, Y., Kamesaki, C., Ito, N., Sakaibara, S., Yaguchi, H.: Effect of spatial structure on colorfulness adaptation for natural images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **29**, A118–A127 (2012)

Color Scheme

Nilgün Olguntürk

Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Ankara, Turkey

Synonyms

[Color combination](#); [Color harmony](#); [Color palette](#)

Definition

An organized selection and arrangement of colors in design.

Introduction and Classification of Color Schemes

Colors from the color circle are usually combined in a systematic and logical manner to create a color scheme. The number of colors (different hues, saturations and lightnesses) in a scheme could range from two to many. A very basic color scheme would be to use black text on a white background, a common default color scheme in Web design.

A predetermined design idea about the final appearance of two-dimensional designs or products or the ambiance or atmosphere in interiors usually shapes the systematic or logic behind combining different colors into establishing a color scheme. Creating style and appeal, evoking intended feelings (such as in an ambiance), and establishing more practical outcomes all are concerns while deciding on a color scheme.

Different types of schemes are used. These are predominantly based on the selection of colors that are regarded as being harmonious, in other words are aesthetically pleasing when viewed together. This can either be achieved with similarities or with contrasts [1, 2]. As color has three dimensions, namely, hue, saturation, and lightness, color schemes might use colors with similar hues, saturations, or lightnesses, as well as colors with contrasting hues, saturations, or lightnesses.

Color Schemes Using Harmony of Similarities

These schemes use similarity of hues (namely, monochromatic and analogous color schemes), similarity of saturations, or similarity of lightnesses. Monochromatic color schemes use a single hue and obtain a variety of colors with variations in that single hue's saturation and/or lightness levels. For example, in a room, a monochromatic color scheme may use only a certain blue with all its varying lightnesses and darkneses (lightness level) or its paler and vivid (saturation level) versions. Analogous color schemes use neighboring hues in a color circle [3]. The hues chosen in this scheme may incorporate secondary hues (hues that can be produced by adding two primary hues) such as using green, yellow-green, and yellow together. There are

also warm or cool color schemes. These schemes either use all warm hues (e.g., red, orange, yellow) or all cool hues (e.g., blue and green) together.

Schemes using similarity of lightnesses combine different hues with similar lightness levels. For instance, they use only light colors together or only dark colors together. Similarities of saturations might also be used where only weak colors or only pure, vivid colors are combined together. In achromatic color schemes, only achromatic colors (black, white, and grays) are used. In this case if the design or space becomes monotonous, sometimes an accent color is added to these schemes, such as a red or a yellow.

Color Schemes Using Harmony of Contrasts

These schemes use contrasts of hues, contrasts of saturations, or contrasts of lightnesses [4]. Schemes with contrasting hues use opposite hues on the color circle. There are different complementary contrast color schemes depending on the desired number of hues to be used, namely, complementary contrast color schemes (two hues), split-complementary color schemes (three hues), double-complementary color schemes (four hues), etc. Complementary contrast color schemes use two opposing hues on the color wheel, for example, purple and yellow or red and green. These two hues can be used in the design with their many varying lightnesses or saturations. The name complementary contrast derives from afterimage complementaries, where the human visual system tries to compensate the overexposure of a certain hue with its opposing hue in the visual system [5]. When one looks at red for a few minutes, where red is the one and only thing in that person's visual field, and then that person looks at a white surface, that person will see an afterimage of the opposing hue (a light green) on that white surface. Split-complementary color schemes use not the direct opposite hue of a selected hue, but its immediate two neighbors. For example, blue will be used not with its exact opposite orange, but its two immediate neighbors which are yellow-orange and red-orange. Double-complementary color schemes use four hues, mainly two hues next to each other and their direct opposites on the color circle. Thus, yellow and

yellow-orange are used together with purple and purple-blue. Triadic color schemes use hues that are evenly spaced around the color wheel, thus having equal number of hues in between them, such as using green, orange, and purple in a design.

Schemes using contrast of lightnesses use light and dark colors together, and schemes using contrast of saturations use weak and strong colors together. It is also possible to create a color scheme by using achromatic and chromatic colors together.

As there are many colors that human beings can see which can come together in many different combinations as described earlier, the initial intent of the design or the design idea becomes prominent before deciding on the color scheme. The color scheme is chosen usually to best reflect the design intent or the design idea. This design intent or idea may vary from being fully functional to purely aesthetic. Some common design intentions are evoking an emotion, hiding from vision (camouflage), making something visible, or coding. If the design intent is to evoke an emotion, for example, creating a calming ambience, and the design idea is connotations with nature, the color scheme for a place might use similarity of lightnesses and saturations, with hues such as brown, green, and blue. If the design intent is to make things visible, for example, for visually impaired people, then the designer should not use schemes with similar lightnesses. In this case, using dark blue (navy), dark red, and black together would not make the design visible for visually impaired or elderly people.

A color scheme is a selection and arrangement of colors in an organized manner. The selection of a certain color scheme over another one purely depends on the design intent and the design idea of a product or space.

Cross-References

- [Color Circle](#)
- [Color Combination](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Palette](#)
- [Primary Colors](#)

References

1. Chevreul, M.E.: *The Principles of Harmony and Contrast of Colors*. Schiffer, West Chester (1987)
2. Birren, F.: *Principles of Color*. Van Nostrand Reinhold, West Chester (1987)
3. Itten, J.: *The Elements of Color*. In: Birren, F. (ed.). (trans: Van Hagen, E.). Van Nostrand Reinhold, New York (1970)
4. Itten, J.: *Design and Form*. Thames and Hudson, London (1987)
5. Kuehni, R.G.: *Color: Essence and Logic*. Van Nostrand Reinhold, Berkshire (1983)

Color Selection

- [Palette](#)

Color Signals

- [Environmental Influences on Color Vision](#)

Color Similarity Space

- [Psychological Color Space and Color Terms](#)

Color Solids

- [Color Order Systems](#)

Color Spaces

- [Color Order Systems](#)

Color Spreading

- [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)

Color Spreading, Neon Color Spreading, and Watercolor Illusion

Baingio Pinna

Department of Humanities and Social Sciences,
Università degli Studi di Sassari, Sassari, Italy

Synonyms

[Color spreading](#); [Filling-in](#); [Visual illusions](#)

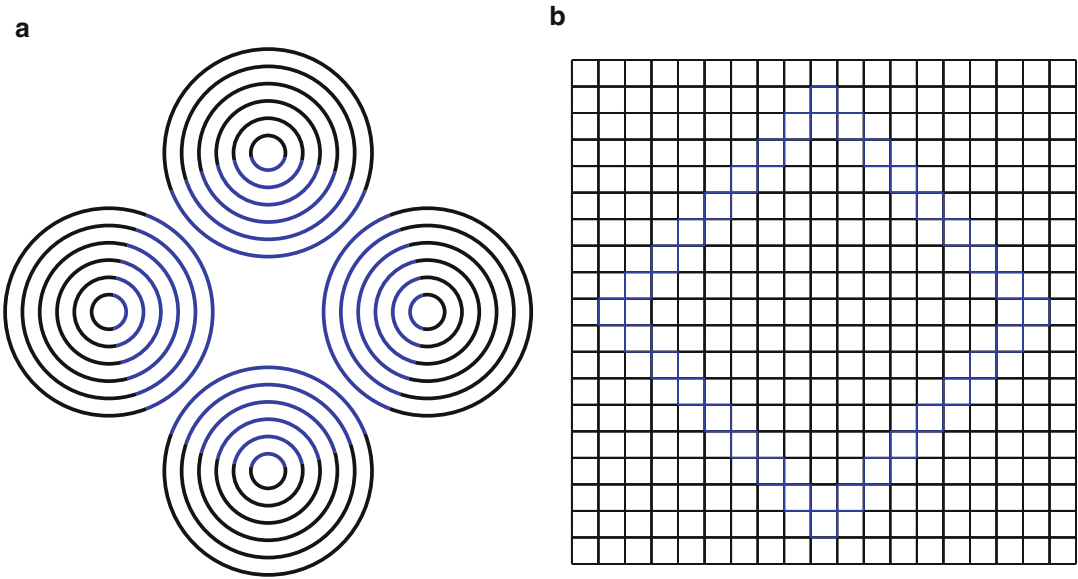
Definition

The color spreading is a long-range assimilative spread of color emanating from a thin colored contour running in the same direction/continuation or being contiguous/adjacent to a darker chromatic contour and imparting a figure-ground effect across a large area. The two main examples of color spreading are the well-known neon color spreading and the watercolor illusion.

Neon Color Spreading

In 1971, Varin [1] studied a novel “chromatic spreading” phenomenon induced when four sets of concentric black circumferences, arranged in a cross-like shape, are partially composed of blue arcs, thus producing a virtual large central blue circle (Fig. 1a). The virtual circle appears as a ghostly transparent circular veil of chromatic translucent diffusion of bluish tint spreading among the boundaries of the blue arcs and filling in the entire illusory circle, induced by the terminations of the black arcs (Fig. 1).

This phenomenon was independently reported a few years later from Varin’s discovery by van Tuijl [2], who named it “neon-like color spreading.” He used a lattice of horizontal and vertical black lines, where segments of different colors (e.g., blue) create an inset diamond shape. The main outcome reveals again a tinted transparent diamond-like veil above the lattice (Fig. 1b).



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 1 Neon color spreading: The central virtual *circle* (a) and the *inset* virtual *diamond*

shape (b) appear as a ghostly overlapping transparent veil of *bluish tint* spreading among the boundaries of the *blue* components

Geometrically, the main geometrical property of all the known cases of neon color spreading is the continuation of one contour, usually black, in a second contour with a different color or, differently stated, a single continuous contour changing from one color to another. Phenomenally, color spreading manifests a coloration and a figural effect described in detail in the following sections.

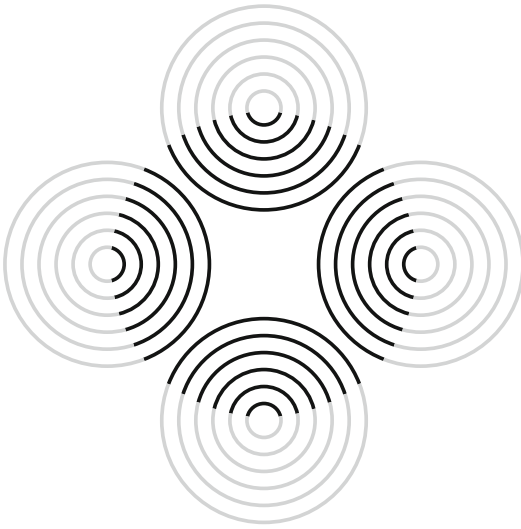
Coloration Effects in Neon Color Spreading

The phenomenology of the coloration effect within neon color spreading depends on the luminance contrast between the two inducing contours and is summed in the next points [2, 3]. (i) The color is perceived as a diffusion of a certain amount of pigment of the inset chromatic segments. (ii) The appearance of the coloration is diaphanous like a smoggy neon upon the background or (under achromatic conditions) like a shadowy, dirty, or filmy transparent veil. (iii) Under conditions where the inset figure is

achromatic and the surrounding inducing elements chromatic, the illusory veil of the inset figure appears tinted not in the achromatic color of the embedded elements, but in the complementary color of the surrounding elements, e.g., the achromatic components appear to spread greenish or bluish illusory colors, respectively, with red or yellow inducers.

Figural Effects in Neon Color Spreading

The apparent coloration of neon color spreading is related to its figural effects. Phenomenally, (i) the illusory neon coloration manifests a depth stratification appearing in front of the inducing elements; (ii) it is also perceived as a transparent film; (iii) by reversing the relative contrast of inset versus surrounding components, the depth stratification reverses accordingly (Fig. 2); (iv) the illusory colored region, under different chromatic conditions, may assume different figural roles by becoming, for example, a superimposed “light,” a “veil,” a “shadow,” or a “fog.”

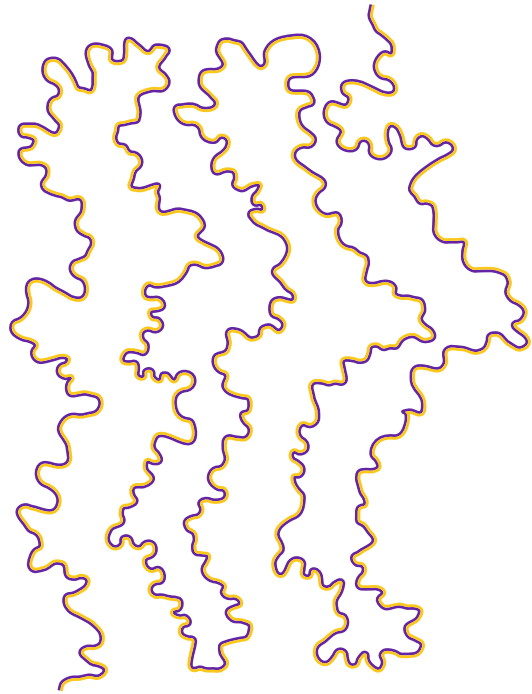


Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 2 By reversing the relative contrast of *inset* versus surrounding components (cf. Fig. 1), the depth stratification of *neon* color spreading reverses accordingly

Watercolor Illusion

The “watercolor illusion” is a long-range spread of color diffusing from a thin colored contour adjacent to a darker chromatic contour and imparting a clear figural effect within large regions [4–9]. In Fig. 3, purple wiggly contours flanked by orange edges are perceived as undefined curved solid shapes, similar to peninsulas emerging from the bottom, evenly colored by a light veil of orange tint spreading from the orange edges.

By reversing purple and orange contours of Fig. 2, the coloration and figure-ground organization are reversed, and, thus, two orangish peninsulas, going from the top to the bottom, connected to the mainland at the top, are now perceived (Fig. 4). Briefly, what appears as illusory colored and segregated in one figure is perceived as empty space without a clear coloration in the other figure and vice versa. Therefore, the peninsulas of Figs. 3 and 4 pop up as totally different objects not referable to the same juxtaposed contours.

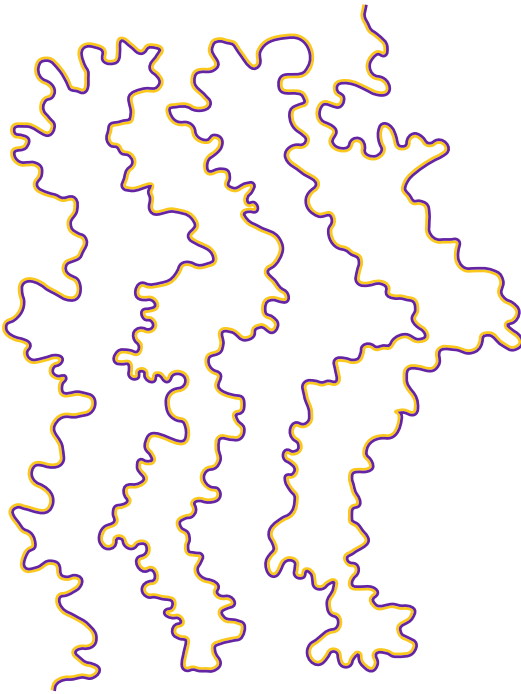


Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 3 Two undulated peninsulas emerging from the *bottom* are perceived illusorily colored by a light veil of *orange tint* spreading from the *orange* edges

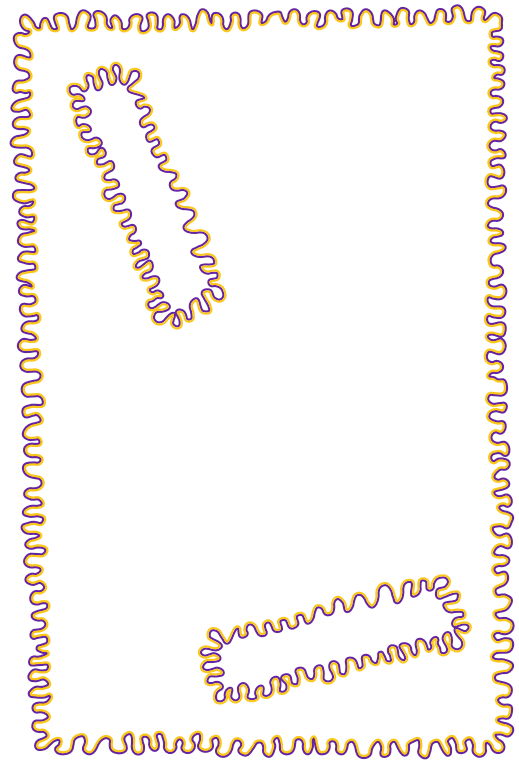
Coloration Effects in the Watercolor Illusion

The phenomenology of the coloration effect in the watercolor illusion reveals the following main attributes:

- (i) As in neon color spreading, the illusory color is approximately uniform. As shown in Fig. 5, the coloration, within the regions where the orange contours are closer, is the same as within the regions where they are distant.
- (ii) The coloration extends up to about 45° .
- (iii) It is complete by 100 ms.
- (iv) Similarly to neon color spreading, all the colors can generate the illusory coloration, as shown in Fig. 6, where an undefined irregular peninsula appears filled with a light blue tint. It should be noted that this peninsula is the Mediterranean Sea when



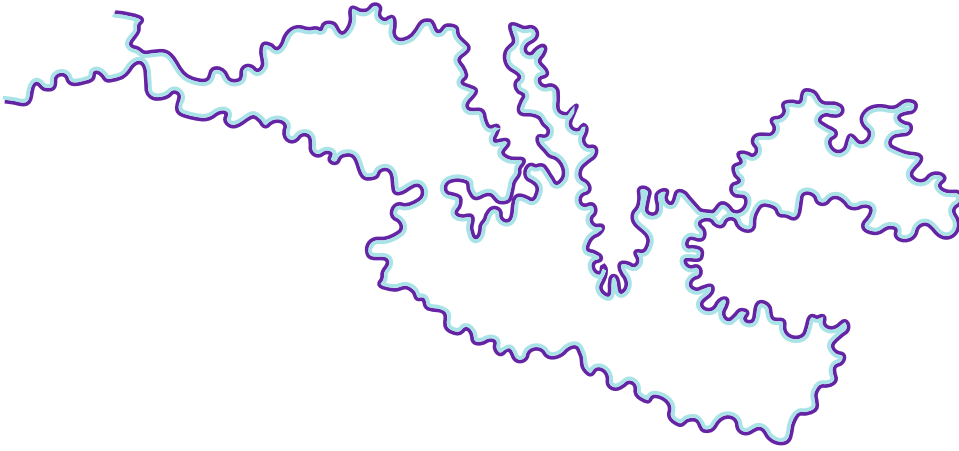
Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 4 By reversing the *purple* and *orange* contours of Fig. 3, the coloration and the figure-ground segregation are reversed: the two peninsulas going from the *top* to the *bottom* and connected to the mainland at the *top*



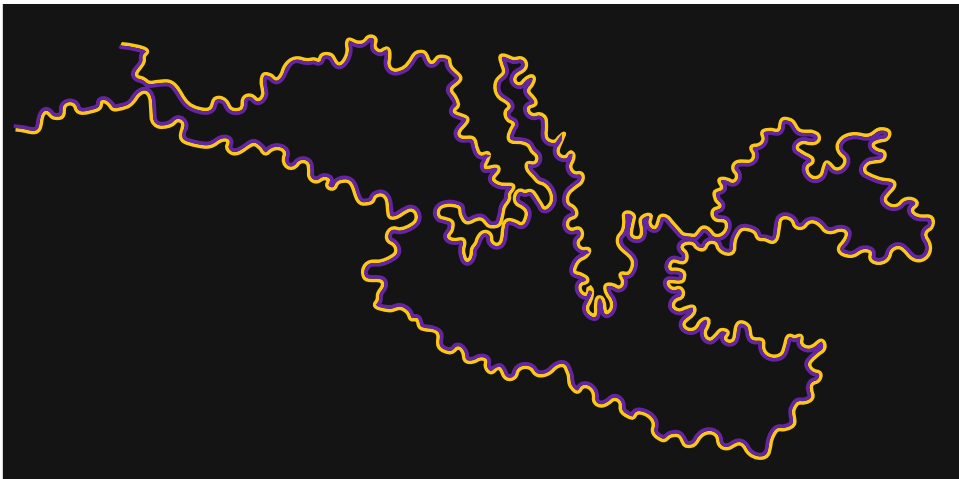
Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 5 The apparent coloration is approximately uniform, i.e., the coloration, within the regions where the *orange* contours are closer, is the same as the regions where they are distant

the two adjacent chromatic contours are reversed.

- (v) The coloration occurs on colored and black backgrounds. In Fig. 7, an undefined irregular peninsula (the Mediterranean Sea when the contours are reversed) appears filled with a purple tint.
- (vi) The optimal contour thickness is approx. 6 arcmin.
- (vii) The effect is stronger with wiggly contours, but it also occurs with straight contours and with chains of dots as shown in Fig. 8 [5, 6].
- (viii) High luminance contrast between inducing contours shows the strongest coloration effect; however, the color spreading is clearly visible at near equiluminance (see Fig. 9) [5, 6].
- (ix) The contour with a lower luminance contrast relative to the background spreads proportionally more than the contour with a higher luminance contrast.
- (x) The color spreads in directions other than the contour orientation.
- (xi) By reversing the colors of the two adjacent contours, the coloration reverses accordingly.
- (xii) Phenomenally, the coloration appears solid, impenetrable, and epiphanous as a surface color.
- (xiii) Similarly to neon color spreading [1–3], the watercolor illusion induces a complementary color when one of the two juxtaposed contours is achromatic and the other chromatic (see Fig. 10) [5]. The inside of the zigzagged annulus appears yellowish.



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 6 All the colors can generate the coloration effect of the watercolor illusion: the undefined irregular peninsula appears filled with a *light blue tint*



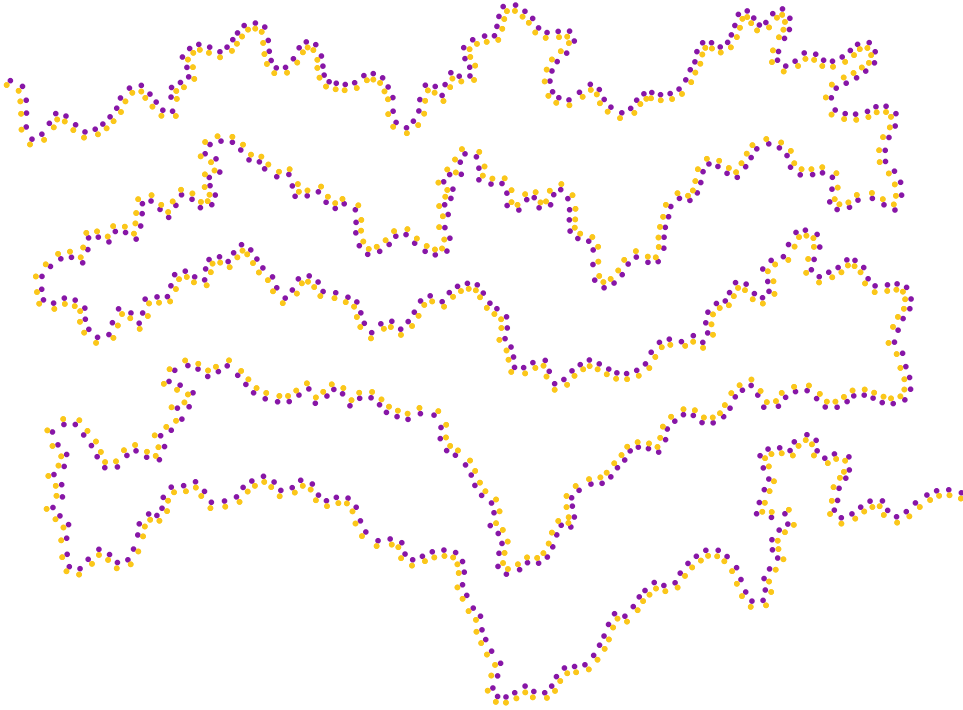
Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 7 The coloration occurs on colored and *black* backgrounds: an undefined irregular peninsula appears filled with a *purple tint*

Figural Effects in the Watercolor Illusion

The main properties of the figural effect are the following:

- (i) The watercolor illusion strongly enhances the “unilateral belongingness of the boundaries” [6, 7, 10].
- (ii) As in neon color spreading, the figural effect of the watercolor illusion is clearly perceived

although it occurs in a different mode of appearance. The figure shows a strong depth segregation and a volumetric rounded and three-dimensional attribute, while the perceived variation of color, going from the boundaries to the center of the object, is seen as a gradient of shading, as if light were reflected onto a volumetric and rounded object. Figure 11 shows undefined, rounded, and volumetric shapes differing from one row



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 8 The watercolor illusion is stronger with wiggly contours, but it also occurs with chains of dots

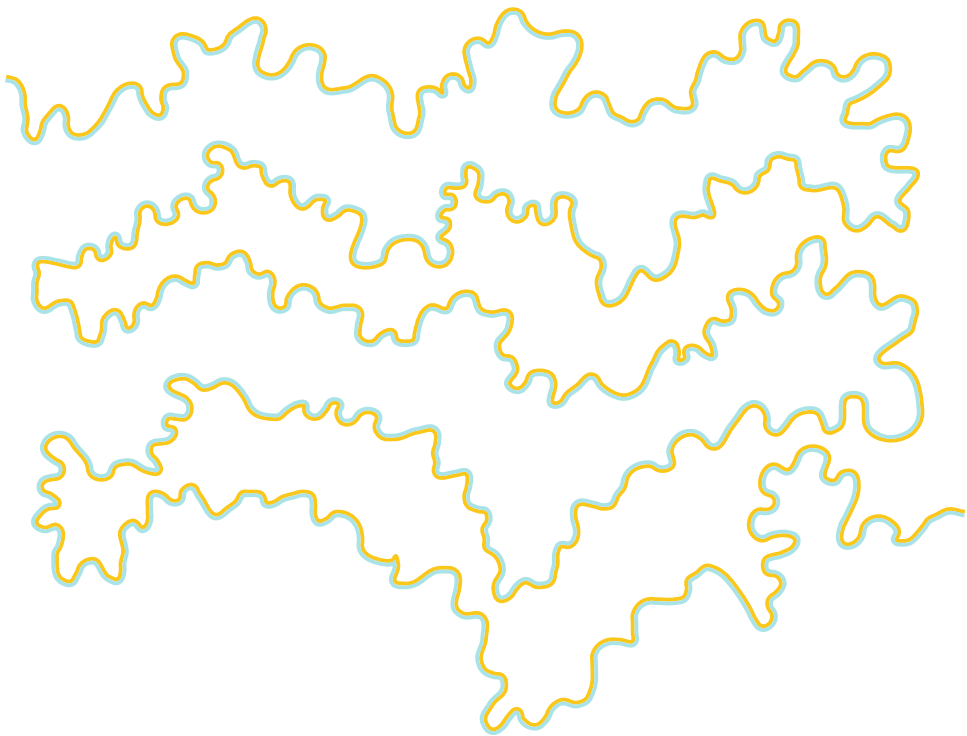
to another on a shapeless empty space due to the unilateral belongingness of the boundaries. The stars are totally invisible.

- (iii) By reversing the colors of the two adjacent contours, the figure-ground segregation reverses accordingly. In Fig. 12, the same elements of Fig. 11, illustrated with reversed purple-orange contours, appear like juxtaposed stars. The undefined shapes differing from one row to another are now invisible.
- (iv) Under the previous conditions, the figure-ground segregation is not reversible and unequivocal.
- (v) The watercolor illusion determines grouping and figure-ground segregation more strongly than the Gestalt principles of proximity, good continuation, *Prägnanz*, relative orientation, closure, symmetry, convexity, past experience, similarity, surroundedness, and parallelism [6, 7, 10]. In Fig. 13, some examples showing the watercolor illusion respectively against and in favor of

surroundedness, relative orientation, good continuation, past experience, and parallelism are illustrated.

- (vi) By reversing the luminance contrast of the background, e.g., from white to black, while the luminance contrast of the contours is kept constant, the figure-ground segregation reverses (Fig. 14) [10]. Going from the bottom to the top of the figure, the crosses become stars. These results are in contrast to Gestalt claim that the currently figural region is maintained even on black/white reversal.

This suggests that the watercolor illusion includes a new principle of figure-ground segregation, the asymmetric luminance contrast principle, stating that, all else being equal, given an asymmetric luminance contrast on both sides of a boundary, the region whose luminance gradient is less abrupt is perceived as a figure relative to the complementary more abrupt region, which is perceived as a background [10].



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 9 High luminance contrast between inducing contours shows the strongest coloration effect; however, the color spreading is clearly visible at near equiluminance

Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 10 The inside of the zigzagged annulus appears yellowish

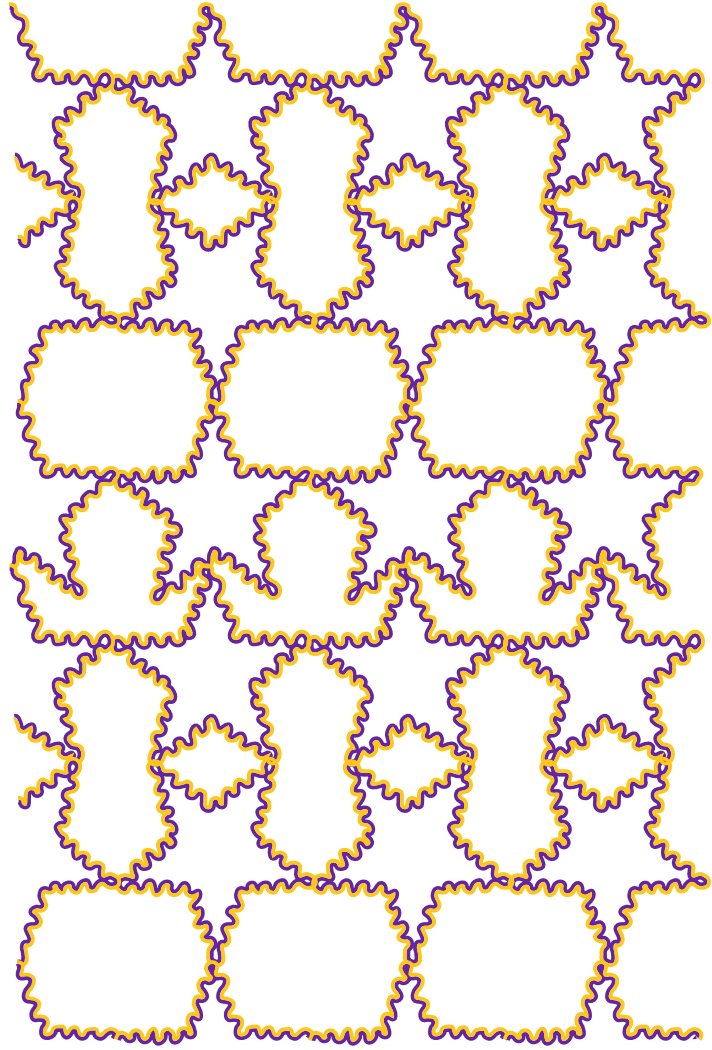


Similarities and Differences In Between the Two Illusions

By summing up the phenomenology of coloration and figural effects in both neon color spreading

and watercolor illusion, the former differs from the latter in both the appearance of the coloration (respectively, transparent vs. solid and impenetrable and diaphanous vs. epiphanous) and in the figural effects (respectively, transparent vs.

Color Spreading, Neon Color Spreading, and Watercolor Illusion,
Fig. 11 Watercolored
 undefined shapes differing
 from one row to another



opaque and dense and appearance as a “light,” a “veil,” a “shadow,” or a “fog” vs. rounded thick and opaque surface bulging from the background).

In spite of these differences, the two illusions are phenomenally similar in their clear color spreading and depth segregation. It is suggested [5] that, while the similarities may depend on the local nearby transition of colors, equivalent in both illusions, the differences may be attributed to the global geometrical boundary conditions, dissimilar in the two illusions. As a matter of fact, while the neon color spreading is elicited by the continuation in the same direction of two

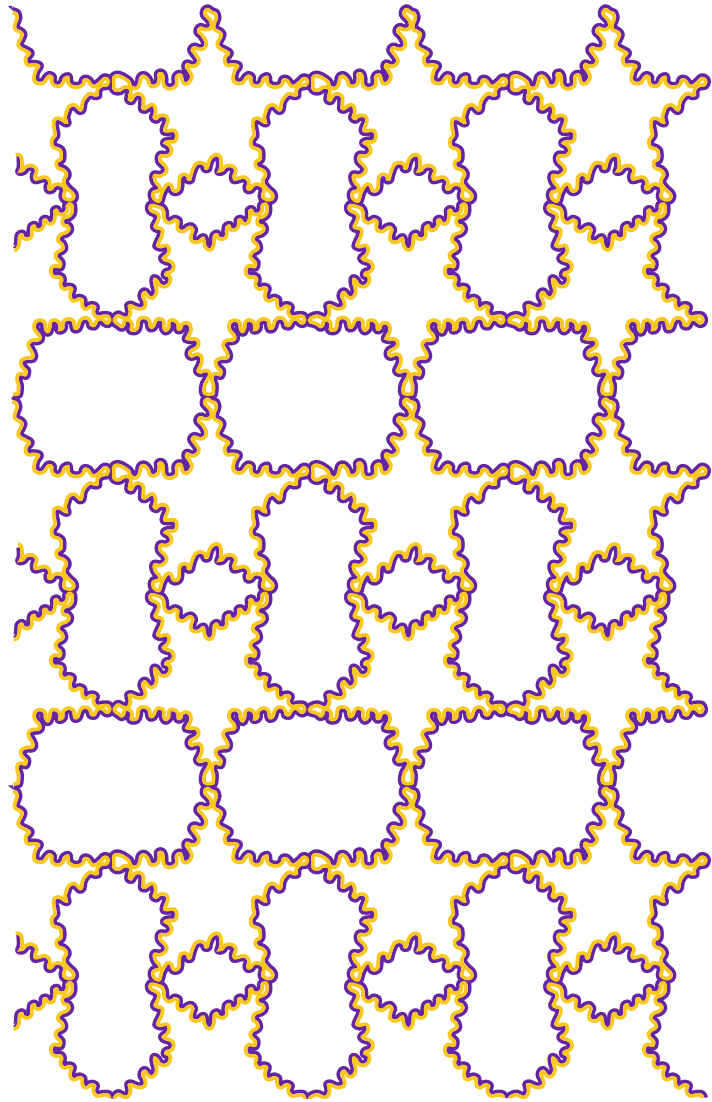
contours of different colors, the watercolor illusion occurs through their juxtaposition.

If this is true, the phenomenal differences between the two illusions can be reduced or eliminated through geometrical variations that bring both phenomena to a common limiting case placed in between and based on the local nearby transitions of colors.

A Limiting Case

Figure 15 shows four conditions that gradually introduce the limiting case. Figure 15a illustrates a

Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 12 The same elements of Fig. 11, illustrated with reversed *purple-orange* contours, appear like juxtaposed stars



C

neon color spreading that represents the starting condition where concentric purple arcs continue by becoming orange. Now, the resulting inset square annulus appears as a transparent veil of orange color not like a ghostly circular veil of translucent color as in Fig. 1. This difference in the color and figural appearance is likely related to the high contrast between the two inducing colors.

Gradual steps toward the final combination of the two illusions in a limiting case are illustrated in Figs. 15b and c. Geometrically, in Fig. 15b, the

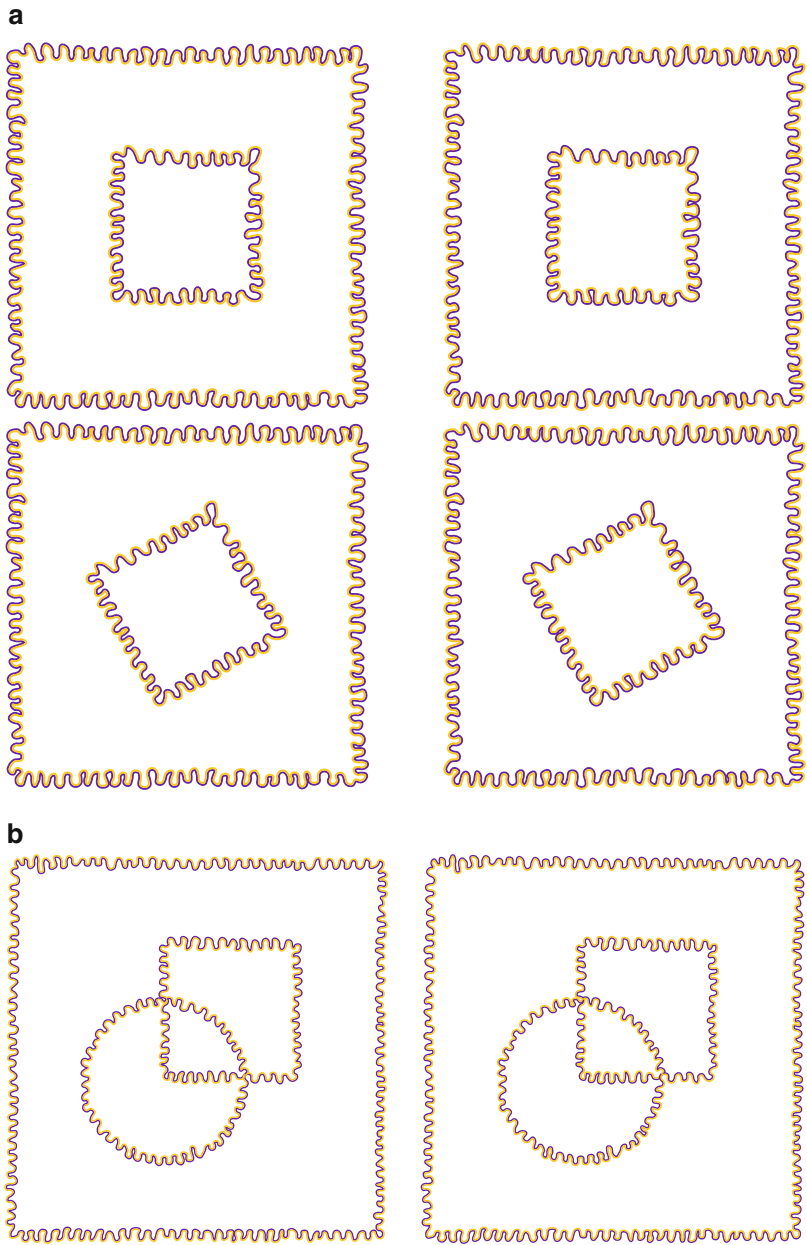
orange inset arcs are reduced to short dashes, creating a condition in between neon color spreading and the watercolor illusion: from the neon color spreading perspective, the inducing elements are contours *continuing* in short dashes (or elongated dots), but from the watercolor perspective, the terminations of the inducing arcs contain *juxtaposed* short dashes. Under these conditions, a coloration effect, not weaker than that of Fig. 15a, is perceived. However, it manifests a poor diaphanous and surface appearance. The illusory figure appears as a fuzzy square annulus, yellowish and brighter than

the background. It is worthwhile to note that the further reduction of the dashes to dots does not change significantly the strength of these effects.

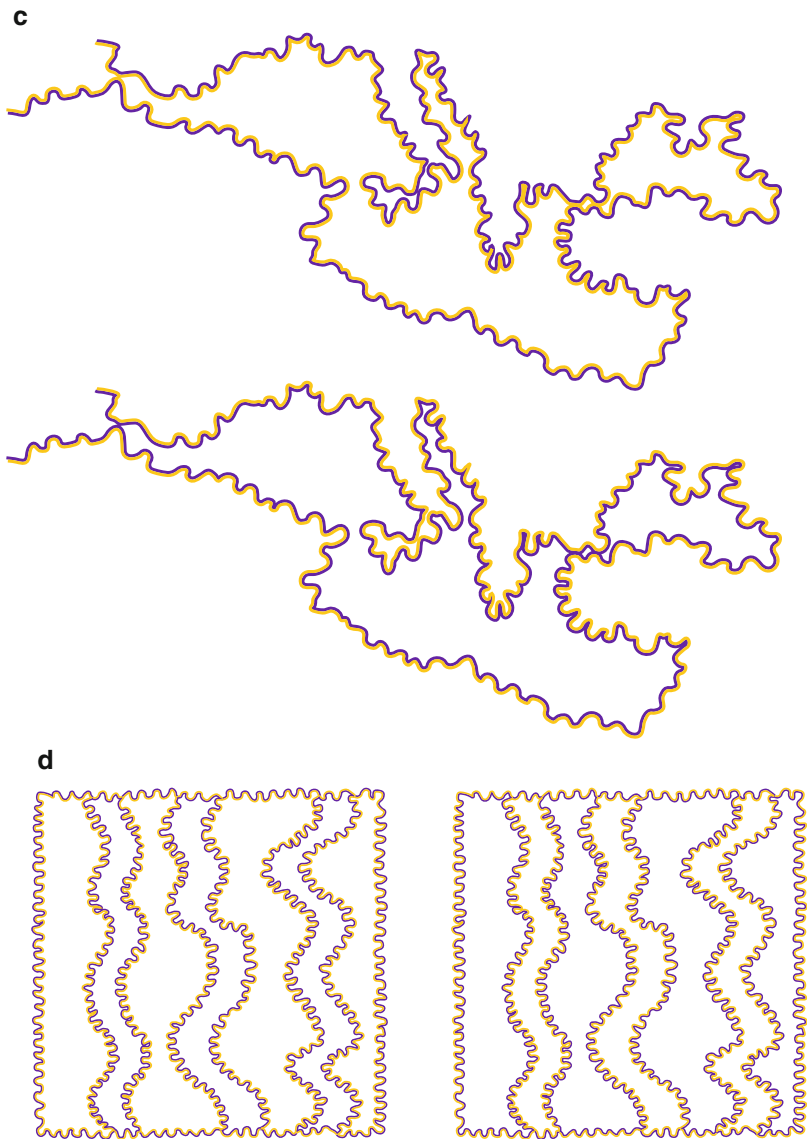
The geometrical reduction in between neon color spreading and the watercolor illusion and opposite to the one of Fig. 15b is illustrated in

Fig. 15c. Under these further conditions, all else being equal, short dashes become the purple arcs of Fig. 15a. Now the coloration effect is weaker than that of Fig. 15a.

Given these geometrical prerequisites, the final step toward the limiting case becomes immediate



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 13 (continued)



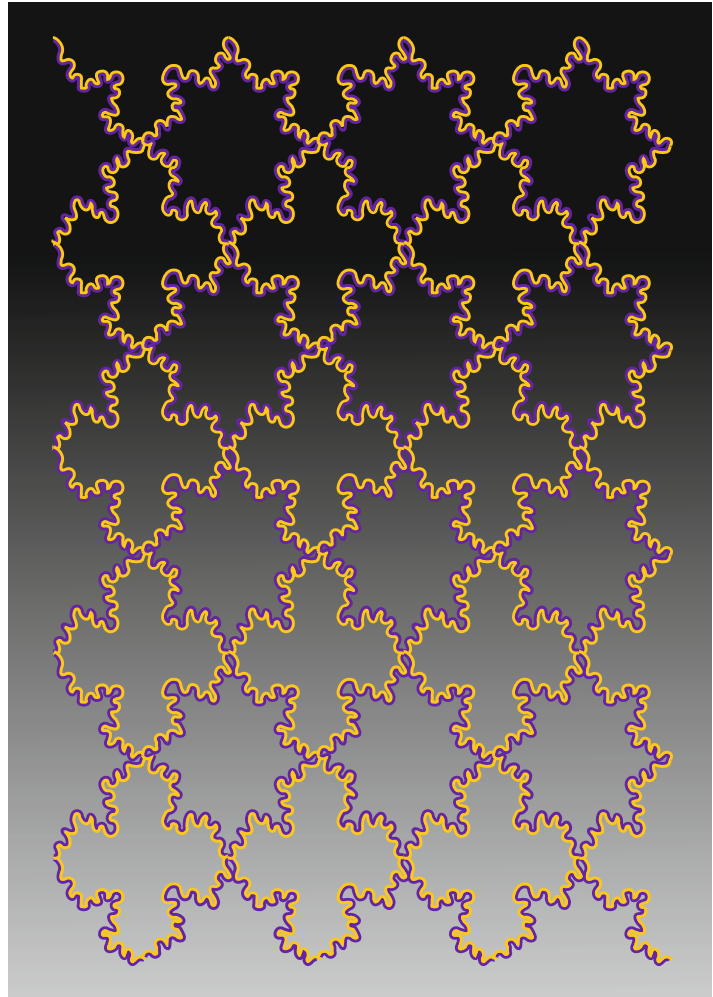
Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 13 Some examples showing the watercolor illusion respectively against and in favor

of surroundedness, relative orientation, good continuation, past experience, and parallelism

and consists in putting together the previous opposite reductions as shown in Fig. 15d. The results show that by reducing both the purple and orange arcs of Fig. 15a to short dashes, the coloration and figural effects do not change significantly [5]. This is corroborated by previous outcomes according to which the watercolor illusion occurs not only by using juxtaposed lines but

also by using juxtaposed chains of dots [6, 7, 10]. Under these conditions both coloration and figural effects become weaker and weaker as the density of the dots becomes sparser and sparser. The two-dot juxtaposition of Fig. 15d can be considered as a true limiting case for neon color spreading and the watercolor illusion. As a matter of fact, (i) the two-dot limiting case can be

Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 14 By reversing the luminance contrast of the background, from *white* to *black*, while the luminance contrast of the contours is kept constant, the figure-ground segregation reverses: going from the *bottom* to the *top* of the figure, the crosses become stars

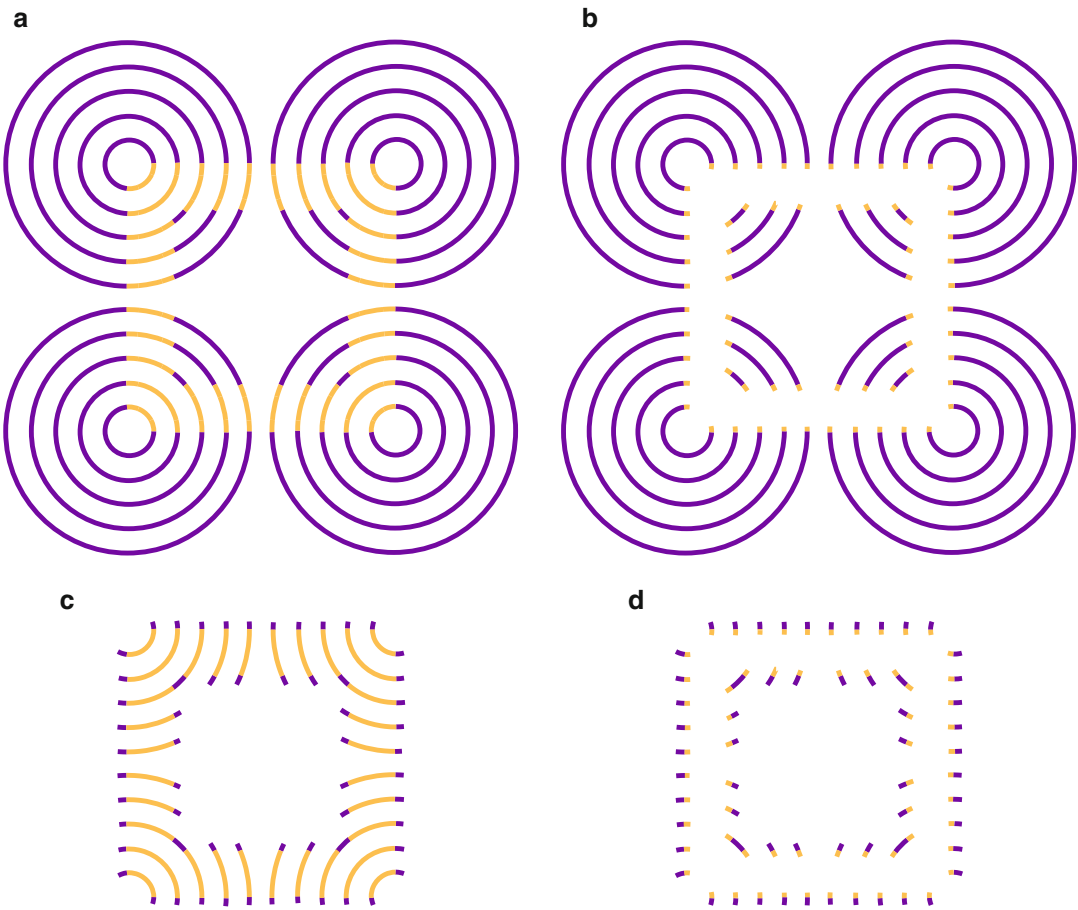


considered as the geometrical common condition, beneath. (ii) The strength of both coloration and figural effects does not change significantly; therefore, the specific mode of appearance of coloration and figural effects in the two illusions is elicited by different local and global distributions of nearby transitions of colors that, in their turn, induce different boundary organizations. (iii) Phenomenally, the differences between the two illusions, where the inner changes are based on continuation and juxtaposition of contours, can now be reconsidered and unified in terms of transition. This is not only a linguistic alternative but also it can bring advantages by providing support for a simple common neural model. (iv) The limiting case can suggest variations of the two illusions that manifest coloration and

figural attributes in between the neon color spreading and the watercolor illusion, as shown in the next section.

Near the Limiting Case

By increasing the width of one of the two juxtaposed contours of the watercolor illusion to such an extent that the contour becomes a surface, the watercolor illusion manifests geometrical properties similar to the neon color spreading and, as a consequence, also shows different coloration and figural effects. The resulting coloration does not assume surface color properties, but properties more similar to the neon color spreading. It



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 15 (a) The *neon* color spreading defined by the continuation of lines of different colors; (b) a condition in between *neon* color spreading and

watercolor illusion, where the *orange inset arcs* are reduced to *short dashes*; (c) a condition in between the two illusions, where the *purple surrounding arcs* of (a) are reduced to *short dashes*; (d) the two-dot limiting case

appears diaphanous like a foggy coloration diffusing everywhere in the background or as a colored light (Fig. 16).

Similarly to the previous condition, the coloration effect of Fig. 17 gives to the illusory star a fuzzy luminous quality. While in Fig. 16 the coloration belongs to the background, in Fig. 17 it belongs to the figure; however, the star does not manifest a strong surface appearance, but its inner surface appears brighter and yellowish and foggy and smooth.

Taken together, these figures suggest that (i) the modes of appearance of coloration are strongly related to boundary conditions that induce specific figural effects; (ii) by changing the boundary

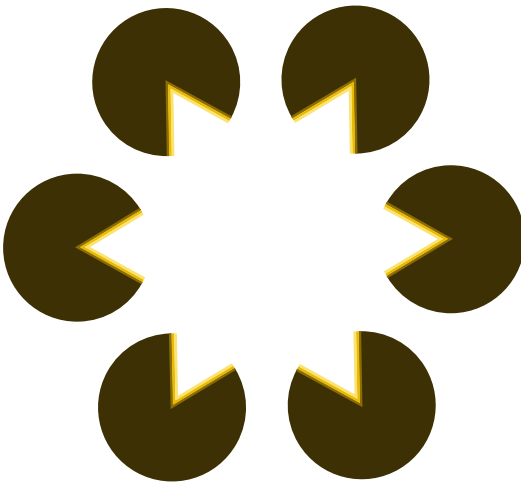
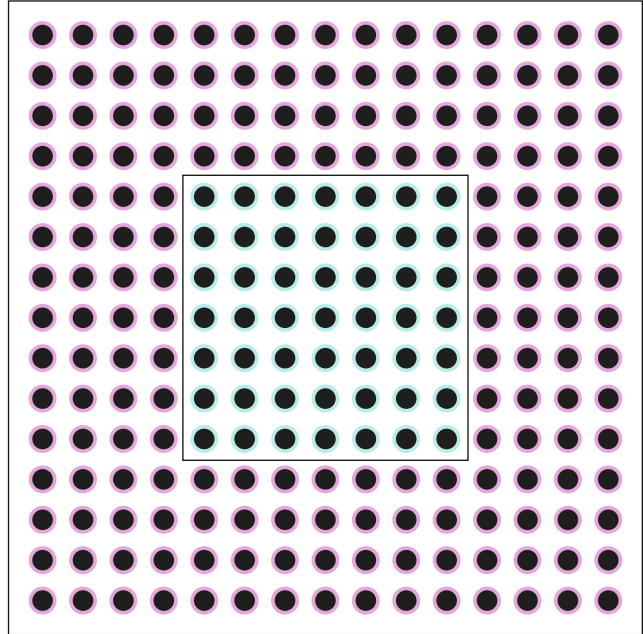
conditions, coloration and figural attributes are perceived more similar to one, to the other illusion, or in between; and (iii) given this variety of appearances on the basis of different conditions, a simpler set of boundary cases, like in the limiting case, can unify both effects using local transitions of colors and can help to explain similarities and dissimilarities of the two illusions.

Neural Mechanisms Underlying the Two Illusions

On the basis of the previous results, coloration and figural effects may derive from parallel processes.

Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 16

A light blue coloration filling in the *inset square* appears surrounded by a red spreading. The coloration effect is not accompanied by a figural effect with a plain volumetric property, but it appears diaphanous like a foggy veil of color



Color Spreading, Neon Color Spreading, and Watercolor Illusion, Fig. 17 The illusory coloration of the star appears fuzzy and luminous and manifests a poor surface appearance

At a feature processing stage, the short-range interaction area around and in between the two dots produces the color spreading common to both illusions, and at a parallel boundary processing stage, the different geometrical

structures in both illusions organize the color spreading to elicit different figural effects. Moreover, the reduction of the neon color spreading and watercolor illusion to a common limiting case can suggest a common and an easier explanation that can be based on the FACADE neural model of biological vision [5]. The model posits that two processes, boundary grouping and surface filling-in [11, 12] substantiated by the cortical interblob and blob streams, respectively, within cortical areas V1 through V4, are responsible of how local properties of color transitions activate spatial competition among nearby perceptual boundaries, with boundaries of lower-contrast edges weakened by competition more than boundaries of higher-contrast edges. This asymmetry induces spreading of more color across these boundaries than conversely. These boundary and surface processes show complementary properties that can also predict how depth and figure-ground effects are generated in these illusions.

Other related findings to both illusions [13–15] showed that neurons in V2 respond with different strength to the same contrast border, depending on the side of the figure to which the border belongs,

implying a neural correlate process related to the unilateral belongingness of the boundaries. Figure-ground segregation may be processed in areas V1 and V2, in inferotemporal cortex and the human lateral occipital complex. Also the color spreading of the two illusions might have its explanation in the cortical representation of borders [9].

Summary

The color spreading is a long-range assimilative spread of color emanating from a thin colored contour running in the same direction/continuation or being contiguous/adjacent to a darker chromatic contour and imparting a figure-ground effect across a large area. Two main examples of color spreading are the well-known neon color spreading and the watercolor illusion. The coloration and the figural properties of the two illusions, studied using phenomenal and psychophysical observations, can be reduced to a common limiting condition, i.e., a nearby color transition called the “two-dot limiting case,” which explains their perceptual similarities and dissimilarities and suggests a common explanation.

Cross-References

- [Assimilation](#)
- [Color Phenomenology](#)
- [Complementary Colors](#)
- [Katz, David](#)
- [Perceptual Grouping and Color](#)

References

1. Varin, D.: Fenomeni di contrasto e diffusione cromatica nell'organizzazione spaziale del campo percettivo. *Riv. Psicol.* **65**, 101–128 (1971)
2. van Tuijl, H.F.J.M.: A new visual illusion: neon-like color spreading and complementary color induction between subjective contours. *Acta Psychologica* **39**, 441–445 (1975)
3. Bressan, P., Mingolla, E., Spillmann, L., Watanabe, T.: Neon colour spreading: a review. *Perception* **26**, 1353–1366 (1997)
4. Devinck, F., Knoblauch, K.A.: Common signal detection model for the perception and discrimination of the watercolor effect. *J. Vis.* **12**, 1–14 (2012)
5. Pinna, B., Grossberg, S.: The watercolor illusion and neon color spreading: a unified analysis of new cases and neural mechanisms. *J. Opt. Soc. Am. A. Opt. Image. Sci. Vis.* **22**, 2207–2221 (2005)
6. Pinna, B., Brelstaff, G., Spillmann, L.: Surface color from boundaries: a new ‘watercolor’ illusion. *Vis. Res.* **41**, 2669–2676 (2001)
7. Pinna, B., Spillmann, L., Werner, J.S.: Anomalous induction of brightness and surface qualities: a new illusion due to radial lines and chromatic rings. *Perception* **32**, 1289–1305 (2003)
8. Werner, J.S., Pinna, B., Spillmann, L.: The brain and the world of illusory colors. *Sci. Am.* **3**, 90–95 (2007)
9. von der Heydt, R., Pierson, R.: Dissociation of color and figure-ground effects in the watercolor illusion. *Spat. Vis.* **19**, 323–340 (2006)
10. Pinna, B.: Watercolor illusion. *Scholarpedia* **3**, 5352 (2008)
11. Komatsu, H.: The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.* **7**, 220–231 (2006)
12. Murakami, I.: Perceptual filling-in. In: *Encyclopedia of Neuroscience*. Springer, Berlin (2008)
13. Zhou, H., Friedman, H.S., von der Heydt, R.: Coding of border ownership in monkey visual cortex. *J. Neurosci.* **20**, 6594–6611 (2000)
14. von der Heydt, R., Zhou, H., Friedman, H.S.: Neural coding of border ownership: implications for the theory of figure-ground perception. In: Behrmann, M., Kimchi, R., Olson, C.R. (eds.) *Perceptual Organization in Vision: Behavioral and Neural Perspectives*, pp. 281–304. Lawrence Erlbaum Associates, Mahwah (2003)
15. Friedman, H.S., Zhou, H., von der Heydt, R.: The coding of uniform color figures in monkey visual cortex. *J. Physiol. (Lond.)* **54**, 593–613 (2003)

Color Statistics

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

Color Stereoscopic Effect

- [Chromostereopsis](#)

Color Stereoscopy

► Chromostereopsis

Color Synesthesia

Berit Brogaard

Department of Philosophy and the Brogaard Lab for Multisensory Research, University of Miami and University of Oslo, Miami, FL, USA

Synonyms

Color synesthesia

Definition

Color synesthesia is a condition in which sensory or cognitive inducers elicit atypical binding of these inducers to concurrent color experiences.

Marks of Color Synesthesia

Synesthesia is a condition in which stimulation in one sensory or cognitive stream involuntarily, or automatically, leads to associated internal or external (illusory or hallucinatory) experiences in a second unstimulated sensory or cognitive system [1–9]. Although most cases of synesthesia are developmental and run in families, acquired cases have also been reported following traumatic brain injury, demyelination, ischemia, tumors, post-traumatic total ocular blindness, and neuropathology involving the optic nerve and/or chiasm [10–12].

Color synesthesia is a special kind of synesthesia that comprises cases of synesthesia in which a noncolored sensory or cognitive stimulus involuntarily leads to internal or external color experiences. The prevalence of color synesthesia is

unknown. Estimates range from 1 in 200 to 1 in 250,000 [13, 14]. Some speculate that color synesthesia may be present in more than 4 % of the population [5].

One of the best-known forms of color synesthesia is grapheme-color synesthesia, in which numbers or letters are seen as colored. But lots of other forms of color synesthesia have been identified, including week-color synesthesia, sound-color synesthesia, taste-color synesthesia, fear-color synesthesia, etc. [5] For lack of space, this entry shall focus primarily on grapheme-color synesthesia.

One mark of color synesthesia is that the synesthetic colors are seen either as projected out onto the world (“projector synesthesia”) or in the mind’s eye (“associator synesthesia”) [15]. Another mark is that it exhibits test-retest reliability [1, 16]: colors identified by the subject as representative of her synesthetic experiences relative to a given stimulus in the initial testing phase are nearly identical to colors identified by the subject as representative of her synesthetic experiences relative to the same stimulus in a retesting phase at a later time (see Fig. 1).

Because of the automatic nature of synesthesia and its test-retest reliability, color synesthesia is not to be confused with memory associations or stereotypical colors of objects. For example, there is no evidence that color synesthetes simply remember the colors of entities or images they were exposed to earlier in their lives or associate stimuli with their stereotypical colors [16].

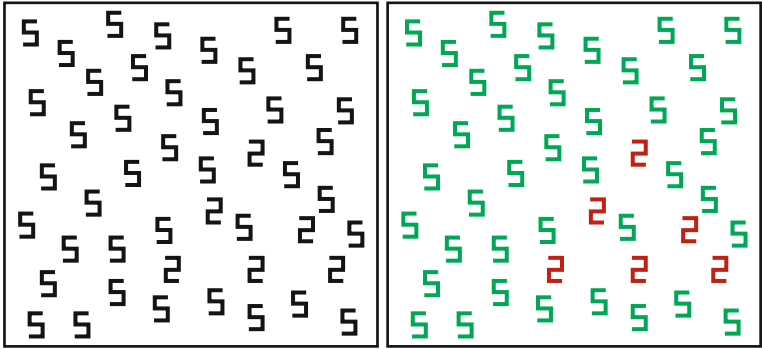
Synesthetic color experience is unique for each synesthete. For example, the letter A may trigger the color red in one grapheme-color synesthete but trigger the color blue in another. In fact, each grapheme has been found to trigger each of the 11 Berlin and Kay colors in different synesthetes (red, pink, orange, yellow, green, blue, purple, brown, black, white, gray). Despite the uniqueness of synesthetic color experience, synesthetic colors sometimes fall into certain clusters. For example, grapheme-color synesthetes tend to associate A with red, E with yellow or white, I with black or white, and O with white [17, 18].

Age/graph	0	1	2	3	4	5	6	7	8	9
3	/	B	Y	G	P	R	Bl	W	Br	R
4	/	B	Y	G	P	R	Bl	W	Br	R
5	Go	B	Y	G	P	R	DBr	W	Br	R
6	Go	B	Y	G	P	R	DBr	W	Br	R
7	B	B	Y	G	P	R	Br	W	Br	R
8	B	B	Y	G	P	R	Bl	W	Br	R

Color Synesthesia, Fig. 1 Example of test-retest reliability of synesthetic experience in one of the St. Louis Synesthesia Lab’s associator grapheme-color synesthetes

from ages 3 to 8 (Go = gold, B = blue, Y = yellow, G = green, P = purple, R = red, Bl = black, DBr = dark brown, Br = brown, W = white)

Color Synesthesia, Fig. 2 When normal subjects are presented with the figure on the *left*, it takes them several seconds to identify the hidden shape. Some grapheme-color synesthetes instantly see the *triangular* shape because they experience the 2 s and the 5 s as having different colors



Low-Level Versus High-Level Perception

An open question about color synesthesia is whether it is a form of low-level or high-level perception. According to Ramachandran and Hubbard [19], synesthesia is a form of low-level perception, a “sensory” phenomenon. As they put it:

Work in our laboratory has shown that synesthesia is a genuine sensory phenomenon. The subject is not just “imagining the color” nor is the effect simply a memory association (e.g., from having played with colored refrigerator magnets in childhood) [19, p. 51].

Some of the evidence listed in favor of treating color synesthesia as a kind of low-level perception is that some grapheme-color synesthetes appear to experience a pop-out effect in visual search paradigms in which some characters elicit synesthetic experience. For example, if a cluster of 2s is embedded in an array of randomly placed 5s, normal subjects take several seconds to find the shape formed by the 2s, whereas grapheme-color

synesthetes who experience a pop-out effect instantly see the shape (see Fig. 2) [7, 20, 21].

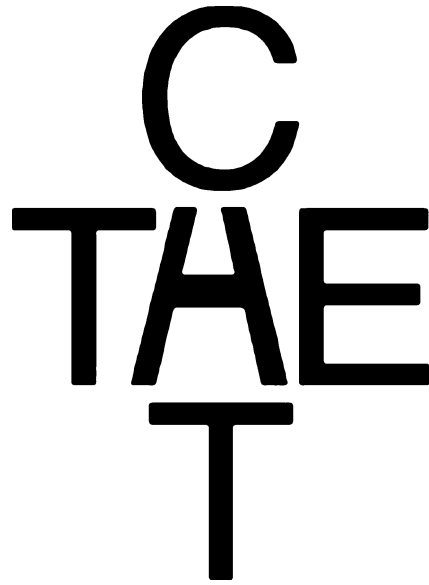
Visual search paradigms are supposed to be indicators of whether synesthetic experience requires focal attention. If synesthetic experience does not require focal attention, then digits with unique synesthetic colors should capture attention, which would lead to highly efficient identification of inducing digits. If, on the other hand, synesthetic experience requires focal attention, then synesthetic colors do not capture attention, and the identification process should be inefficient [22]. Perceptual features must be processed early enough in the visual system for them to attract attention and lead to pop-out and segregation [23, 24]. So the appearance that synesthetic experience can lead to pop-out and segregation indicates that synesthesia is a low-level perceptual phenomenon [7, 19, 20].

While a significant number of grapheme-color synesthetes are more efficient in visual search paradigms than controls, this does not clearly show that attention is not required for synesthetic

experience, however. In one subject PM, it was shown that quick identification of graphemes occurred only when the graphemes that elicit synesthetic experience were close to the initial focus of attention [25]. Smilek et al. [26] used a variation on the standard visual search paradigm to test subject J's search efficiency. J was shown an array of black graphemes on a colored background, some of which induced synesthetic experience. The colored background was either congruent or incongruent with the synesthetic color of the target. The researchers found that J was more efficient in her search when the background was incongruent than when it was congruent. This indicates that the synesthetic colors attracted attention only when they were clearly distinct from the background.

Edquist et al. [22] carried out a group study involving 14 grapheme-color synesthetes and 14 controls. Each subject performed a visual search task in which a target digit differed from the distractor digits in terms of its synesthetic color or its display color. Both synesthetes and controls identified the target digit efficiently when the target had a unique display color, but the two groups were equally inefficient when the target had a unique synesthetic color. The researchers concluded that for most grapheme-color synesthetes, graphemes elicit synesthetic color only once the subject attends to them. This indicates that synesthetic colors cannot themselves attract attention because they are not processed early enough in the visual system.

Another reason to think that not all cases of color experience in grapheme-color synesthesia are forms of low-level perception is that their appearance seems to depend on interpretation of visual experience. In Fig. 3, for instance, synesthetes assign different colors to the middle letter depending on whether they interpret the string of letters as spelling the word "cat" or the word "the." For example, one of our child subjects, a 7-year-old female, experiences the middle letter as red when she reads the word "cat" and the middle letter as brown when she reads the word "the." This suggests that it is not the shape of the letter that gives rise to the color experience but the category or concept associated with the letter [27].



Color Synesthesia, Fig. 3 Synesthetes interpret the middle letter as an A when it occurs in "cat" and as an H when it occurs in "the." The color of their synesthetic experience will depend on which word the grapheme is considered part of

The fact that the very same grapheme can elicit different color experiences in synesthetes depending on the context in which it occurs suggests that synesthetes need to interpret what they visually experience before they experience synesthetic colors. Though Ramachandran and Hubbard [19] argue that grapheme-color synesthesia is a form of low-level perception (a "sensory phenomenon"), they grant that linguistic context can affect synesthetic experience. They presented the sentence "Finished files are the result of years of scientific study combined with the experienced number of years" to a subject and asked her to count the number of "f's" in it. Most normal subjects count only three "f's" because they disregard the high-frequency word "of." Though the synesthete eventually spotted six "f's" she initially responded the way normal subjects do.

Ramachandran and Hubbard [19] suggest that these contextual effects can be explained by top-down factors. Whether this is right, however, will depend on whether color experience processed in early visual areas is indeed affected by top-down factors. If it is not, then top-down

influences cannot explain the contextual effects. A better explanation of contextual influence then may be that interpretation of low-level perceptual information is required for synesthetic experience.

Another explanation of the disagreement about whether color synesthesia gives rise to pop-out effects may bear on the fact that few studies of pop-out effects have properly distinguished between projector synesthesia and associator synesthesia as well as what Ramachandran calls “higher synesthesia” and “lower synesthesia.” Lower grapheme-color synesthesia is synesthesia (either projector or associator) that arises in response to sensory stimuli, whereas higher grapheme-color synesthesia is synesthesia (either projector or associator) that arises in response to thoughts of graphemes. It is possible that the majority of synesthetes are higher synesthetes and that only lower synesthetes experience pop-out effects.

Neural Mechanism

The precise neural mechanism underlying color synesthesia is unknown. One hypothesis, the so-called local cross-activation hypothesis, proposed by Hubbard and Ramachandran, holds that grapheme-color synesthesia arises due to cross-activation between color areas in the visual cortex and the adjacent visual word form area [7, 20, 28]. This suggestion is inspired by the observation that local crossover phenomena can explain other illusory and hallucinatory experiences, such as phantom limb sensations.

A second hypothesis is that color synesthesia may be due to disinhibited feedback from an area of the brain that binds information from different senses [3, 29, 30]. The main piece of evidence cited in favor of this hypothesis comes from an analogous case in which a patient PH reported seeing visual movement in response to tactile stimuli following acquired blindness [29]. As PH was blind, he could not have received the information via standard visual pathways. It is plausible that the misperception was a result of disinhibited feedback from brain regions that receives information from other senses.

The fact that synesthetic experiences can arise when subjects are under the influence of psychedelics provides some further evidence for the disinhibited feedback hypothesis [31]. The synesthetic effect of psychedelic substances may be due to an inhibition of feedback from areas of information binding. It is unknown, however, whether drug-induced synesthesia and congenital synesthesia have the same underlying mechanism.

A third hypothesis is that color synesthesia arises as a result of aberrant reentrant processing [21, 32]. The hypothesis is similar to the disinhibited feedback hypothesis but suggests specifically that high-level information reenters color areas in visual cortex and that it is this form of reentrant information processing that leads to the experience of synesthetic colors. This model would explain why visual context and meaning typically influence which synesthetic colors a grapheme gives rise to [32, 33].

It is plausible that different forms of color synesthesia proceed via different mechanisms. Cases of color synesthesia have been reported in which the visual cortex is not involved in generating synesthetic colors [12, 34]. None of the three aforementioned hypotheses, despite their plausibility in run-of-the mill cases, can explain more unusual cases of color synesthesia.

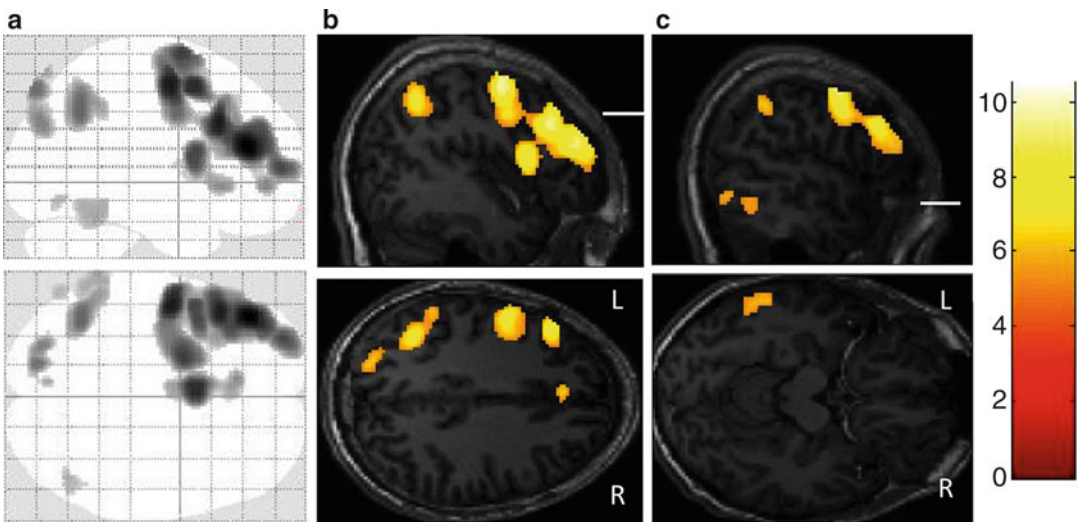
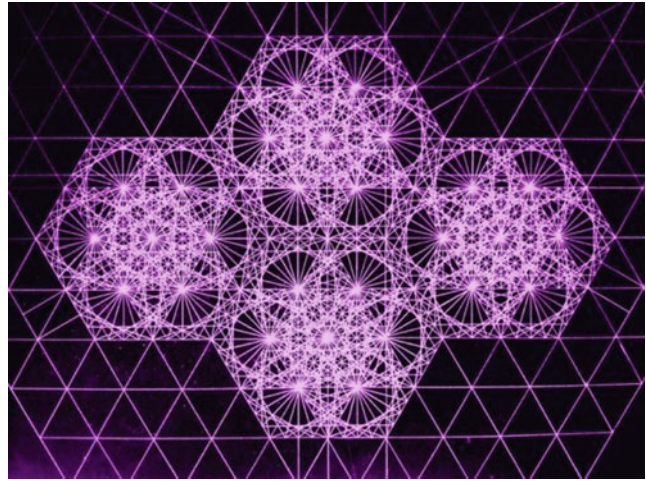
Cognitive Advantages of Color Synesthesia

If pop-out effects require attention to the synesthetic graphemes, grapheme-color synesthesia is unlikely to give subjects much of a cognitive advantage in visual search tests. However, there may nonetheless be cognitive advantages associated with color synesthesia. For example, some case studies suggest that grapheme-color synesthetes may have greater recall ability for digits and written names when compared to non-synesthetes [35, 36].

In rare cases color synesthesia has been associated with extreme mathematical skills. Subject DT, for example, sees numbers as three-dimensional colored, textured forms [34]. His synesthesia gives him the ability to multiply high digits very rapidly. He reports that the product of

Color Synesthesia,

Fig. 4 Image hand-drawn
by subject JP



Color Synesthesia, Fig. 5 Sagittal slices. Activation induced by the image-inducing formula contrasted to non-inducing formulas. The SPM(T) maps were

thresholded at family-wise-error-corrected p-value 0.01 and overlaid on JP's structural T1-weighted MRI which was standardized into MNI-space using SPM8 [12]

multiplying two numbers is the number that corresponds to the shape that fits between the shapes corresponding to the multiplied numbers. Subject DT's color synesthesia also gives rise to extreme mnemonic skills. DT currently holds the European record in reciting the decimal points of the number pi. An fMRI study comparing DT to controls while attempting to locate patterns in number sequences indicated that DT's synesthetic color experiences occur as a result of information processing in nonvisual brain regions, including temporal, parietal, and frontal areas [34].

Brogaard et al. [12] describe a case of a subject, JP, who has exceptional abilities to draw complex geometrical images by hand and a form of acquired synesthesia for mathematical formulas and moving objects, which he perceives as colored, complex geometrical figures (see Fig. 4).

JP's synesthesia began in the wake of a brutal assault that led to unspecified brain injury. A fMRI study contrasting activity resulting from exposure to image-inducing formulas and non-inducing formulas indicated that JP's colored synesthetic images arise as a result of activation in

areas in the temporal, parietal, and frontal cortices in the left hemisphere. The image-inducing formulas as contrasted with the non-inducing formulas induced no activation in the visual cortex or the right hemisphere [12] (see Fig. 5).

These two unusual case studies suggest that at least some forms of color synesthesia can give rise to cognitive advantages in the area of mathematics. As the visual cortex does not appear to be directly involved in generating the synesthetic images in either subject, the two cases also suggest that at least some forms of color synesthesia are best characterized as forms of high-level perception that proceeds via a nonstandard mechanism.

Cross-References

- [Afterimage](#)
- [Ancient Color Categories](#)
- [Appearance](#)
- [Chromatic Contrast Sensitivity](#)
- [Color and Visual Search, Color Singletons](#)
- [Color Constancy](#)
- [Color Contrast](#)
- [Color Phenomenology](#)
- [Color Processing, Cortical](#)
- [Color Psychology](#)
- [Color Synesthesia](#)

References

1. Baron-Cohen, S., Wyke, M., Binnie, C.: Hearing words and seeing colors: an experimental investigation of synesthesia. *Perception* **16**, 761–767 (1987)
2. Cytowic, R.E.: *Synesthesia: A Union of the Senses*. Springer, New York (1989)
3. Grossenbacher, P.G., Lovelace, C.T.: Mechanisms of synesthesia: cognitive and physiological constraints. *Trends Cogn. Sci.* **5**, 36–41 (2001)
4. Hubbard, E.M.: Neurophysiology of synesthesia. *Curr. Psychiatry Rep.* **9**, 193–199 (2007)
5. Hubbard, E.M., Ramachandran, V.S.: Neurocognitive mechanisms of synesthesia. *Neuron* **48**, 509–520 (2005)
6. Hubbard, E.M., Arman, A.C., Ramachandran, V.S., Boynton, G.M.: Individual differences among grapheme-color synesthetes: brain-behavior correlations. *Neuron* **45**(6), 975–985 (2005)
7. Ramachandran, V.S., Hubbard, E.M.: Psychophysical investigations into the neural basis of synaesthesia. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 979–983 (2001)
8. Sperling, J.M., Prvulovic, D., Linden, D.E.J., Singer, W., Stirn, A.: Neuronal correlates of graphemic colour synaesthesia: a fMRI study. *Cortex* **42**, 295–303 (2005)
9. Ward, J., Huckstep, B., Tsakanikos, E.: Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex* **42**, 264–280 (2006)
10. Afra, P., Funke, M., Matsu, F.: Acquired auditory-visual synesthesia: a window to early cross-modal sensory interactions. *Psychol. Res. Behav. Manage.* **2**, 31–37 (2009)
11. Beauchamp, M.S., Ro, T.: Neural substrates of sound-touch synesthesia after a thalamic lesion. *J. Neurosci.* **28**, 13696–13702 (2008)
12. Brogaard, B., Vanni, S., Silvanto, J.: Seeing mathematics: perceptual experience and brain activity in acquired synesthesia. *Neurocase*. (2012, in press)
13. Cytowic, R.E.: Synesthesia: phenomenology and neuropsychology. In: Baron-Cohen, S. (ed.) *Synesthesia: Classic and Contemporary Readings*, pp. 17–39. Blackwell, Oxford (1997)
14. Sagiv, N., Ward, J.: Cross-modal interactions: lessons from synesthesia. *Prog. Brain Res.* **155**, 259–271 (2006)
15. Dixon, M.J., Smilek, D., Merikle, P.M.: Not all synaesthetes are created equal: projector versus associator synaesthetes. *Cogn. Affect. Behav. Neurosci.* **4**, 335–343 (2004)
16. Eagleman, D.M., Kagan, A.D., Nelson, S.S., Sagaram, D., Sarma, A.K.: A standardized test battery for the study of synesthesia. *J. Neurosci. Methods* **159**, 139–145 (2007)
17. Baron-Cohen, S., Harroson, J., Goldstein, L.H., Wyke, M.: Coloured speech perception: is synaesthesia what happens when modularity breaks down? *Perception* **22**, 419–426 (1993)
18. Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., et al.: Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cogn. Neuropsychol.* **22**(8), 1069 (2005)
19. Ramachandran, V.S., Hubbard, E.M.: The phenomenology of synaesthesia. *J. Conscious. Stud.* **10**, 49–57 (2003)
20. Ramachandran, V.S., Hubbard, E.M.: Synaesthesia: a window into perception, thought and language. *J. Conscious. Stud.* **8**, 3–34 (2001)
21. Smilek, D., Dixon, M.J., Cudahy, C., Merikle, P.M.: Synaesthetic photisms influence visual perception. *J. Cogn. Neurosci.* **13**, 930–936 (2001)
22. Edquist, J., Rich, A.N., Brinkman, C., Mattingly, J.B.: Do synaesthetic colours act as unique features in a visual search? *Cortex* **42**, 222–231 (2006)
23. Beck, J.: Effect of orientation and of shape similarity on perceptual grouping. *Percept. Psychophys.* **1**, 300–302 (1966)
24. Treisman, A.: Perceptual grouping and attention in visual search for features and for objects. *J. Exp. Psychol. Hum. Percept. Perform.* **8**(2), 194–214 (1982)
25. Laeng, B., Svardal, F., Oelmann, H.: Does color synesthesia pose a paradox for early-selection theories of attention? *Psychol. Sci.* **15**, 277–281 (2004)

26. Smilek, D., Dixon, M.J., Merikle, P.M.: Synaesthetic photisms guide attention. *Brain Cogn.* **53**, 364–367 (2003)
27. Cytowic, R.E., Eagleman, D.M.: *Wednesday is Indigo Blue*. MIT Press, Cambridge, MA (2009)
28. Hubbard, E.M., Manohar, S., Ramachandran, V.S.: Contrast affects the strength of synesthetic colors. *Cortex.* (2005b, in press)
29. Arnel, K.C., Ramachandran, V.S.: Acquired synesthesia in retinitis pigmentosa. *Neurocase* **5**, 293–296 (1999)
30. Grossenbacher, P.G.: Perception and sensory information in synaesthetic experience. In: Baron-Cohen, S., Harrison, J.E. (eds.) *Synaesthesia: Classic and Contemporary Readings*, pp. 148–172. Blackwell, Malden (1997)
31. Shanon, B.: Ayahuasca visualizations: a structural typology. *J. Conscious. Stud.* **9**, 3–30 (2002)
32. Myles, K.M., Dixon, M.J., Smilek, D., Merikle, P.M.: Seeing double: the role of meaning in alphanumeric-colour synaesthesia. *Brain Cogn.* **53**, 342–345 (2003)
33. Dixon, M.J., Smilek, D.: The importance of individual differences in grapheme-color synesthesia. *Neuron* **45**, 821–823 (2005)
34. Bor, D., Billington, J., Baron-Cohen, S.: Savant memory for digits in a case of synaesthesia and Asperger syndrome is related to hyperactivity in the lateral prefrontal cortex. *Neurocase* **13**, 311–319 (2007)
35. Mills, C.B., Innis, J., Westendorf, T., Owsianiecki, L., McDonald, A.: Effect of a synesthete's photisms on name recall. *Cortex* **42**, 155–163 (2006)
36. Smilek, D., Dixon, M.J., Cudahy, C., Merikle, P.M.: Synesthetic color experiences influence memory. *Psychol. Sci.* **13**(6), 548–552 (2002)

Color Syntax

► Color Combination

Color Trends

Maria Luisa Musso¹, Renata Pompas² and Leonhard Oberascher³

¹Architecture, Design and Urbanism, Color Research Program, University of Buenos Aires, Buenos Aires, Argentina

²Milan, Italy

³FH Joanneum University of Applied Sciences, Graz, Salzburg, Austria

Synonyms

Color directions; Forecast; Temporal collective color preferences; Tendencies

Definition

Expected forthcoming market interest in some specific color shades.

Overview

Sociological Function of Color Trends

Color trends are temporal collective color preferences. They restrain individual- as well as group-specific color preferences and determine people's tastes and judgments over a certain period. Their application is typically not restricted to particular objects, contexts, or aesthetic conventions and is unconstrained by any functional, symbolic, or formal influences. Therefore, even colors with inherently negative symbolic meanings or connotations can become collectively preferred for a certain period. Individuals as well as society more generally perceive color trends as an articulation of "the predominant taste orientation." Anyone who adopts the trend demonstrates that she/he is open-minded, modern, and a member of an ideal (ized) peer group representing the *Zeitgeist*. In the end, using color trends also relieves individuals of mental strain. Those who adopt them avoid stylistic confrontation and guarantee that their choices meet approval [1].

Cyclic Recurrence of Color Trends

Several studies [2–4], which analyzed the rise and fall of collective color preferences in architecture, interior design, and consumer goods during several decades, suggest a cyclic recurrence of color trends. Oberascher [5] points out that color trends must repeat in the long run because the potential gamut for color innovation is limited by the natural boundaries of the perceptual color space. New colors in a strict sense cannot be invented. Technological progress, however, may produce new color appearances. Based on an analysis of predominant color trends in furnishing and interior design in Germany and Austria between 1972 and 1992, he suggests a general model of the cyclic recurrence of collective color preferences. One explanation why certain color groups and ► color combinations are more likely to be

repeated than others might be that the aesthetical evaluation and appreciation of color groups and combinations in general are determined by universal laws of perception and Gestalt psychology. But other factors may play a role. The emergence and spread of new collective color preferences are most probably rooted in a basic human desire (evolutionary, biological, neurophysiological, psychological, sociological) for change, alteration, renewal, and innovation. People's readiness to engage in new and unfamiliar colors increase over time as they become satiated with one trend. Since no new colors can be invented but only selected from the existing perceptual color gamut, those color groups and combinations that have been out of use for the longest time will appear novel. For young people these colors are new; for the older generations, they are a (re) contextualization of their color memories. The model of the cyclic recurrence of collective color preferences does not claim to predict color trends but might be used as a strategic tool for product marketing and management particularly in the furniture, furniture supply, and building material industry.

The Emergence of Color Trends

Historically, fashion has always set its rules and its trends based on a top-down pyramid approach. Every civilized society in its history has been able to use a large number of dye materials to color fibers, textiles, and leather in multiple shades. Where dye materials for some specific colors were lacking, international trade made up for it since they were imported, becoming widespread [6]. It can thus be said that no major color area has been neglected in the evolution of the trends and fashion of upper classes, but some color ranges have prevailed over others from time to time based on aesthetical, symbolic, and social choices imposed by the ruling classes.

A swing in trends occurred as ready-to-wear (prêt-à-porter) clothes and design became more widespread, and consumption democratization developed. Some spontaneous bottom-up trends have emerged since the 1960s. These have affected fashion, resulting in a combined top-down and bottom-up approach that is still

underway, and it currently seems that street fashion is predominant [7].

Instruments to identify and anticipate market requirements were needed in order to properly respond to the new industrial organization of textile, clothing, furnishings, and design industries.

This has led to color trend forecast agencies, where teams of researchers (consisting of stylists, designers, sociologists, psychologists, and market experts) started to analyze consumption and behaviors and to anticipate color trends, ahead of the industrial manufacturing schedule required (2 years).

Not Only Colors, but Also Range Quality

Since then, color trends have renewed their offer season after season, not by changing major color families (red, yellow, green, etc.) but by the "quality of their range" (soft, bright, pastel, dusty, clear, dark, etc.). In the 1990s, surface finishing too became an element inherent to color as it changed its look, value, and "style."

Historically, color trends have encapsulated the "spirit" of the time for many years. Their likelihood of meeting consumption demand results from the fact that they are based on the identification of color trends that are already found in the market, that the industry has adopted, and that the consumer has found within a predefined and limited available range, thus confirming relevant forecasts [8].

What Are Trends?

Trends are the expression of fashion themes and an indication of consumption and behavioral styles. They play two roles: on the one hand they are a monitoring tool for fashion and its multiple manifestations, and on the other hand, they anticipate and express current social orientations.

The life span of one or more trends on the market is very uncertain, from one season to several years.

The emergence of trends depends on historical, social, political, and cultural factors, which develop in the general public regularly and cyclically and influence almost a whole decade.

However, some very successful trends can come forth in a strange and sudden manner, as a

result of occasional factors that have become very popular at a national or international level (a film, an exhibition, a music phenomenon, a product launch, the opening of new shops, some successful places open to the public, etc.).

Individualization of Color Trends

Today's global market, where goods move freely, is unstable and difficult to control because many style orientations coexist simultaneously. Comparative analyses of major color trends often highlight dissimilarities in results, which are due to both the target and the impossibility to develop truly accurate forecasts. That is why color trends no longer play the role of universal consumption indicators; instead, they now act as a particularized anticipation of and focus on the themes intended for a specific market sector.

Companies use color trends to choose corporate colors for communication purposes as they try to develop a visible business identity, based on colors allowing them to stand out from competitors and to take advantage of mainstream tendencies by means of season's colors. Albeit part of the general aesthetic framework, these color trends mark corporate choices and enable consumers to invent their own style through individual combinations.

This is the result of the current historical moment, known as "post-fashion," where there are unclear and often contradictory signs and a fragmented coexistence of all types of different opportunities and orientations.

Creation of Color Trends

Color trends result from a research on ongoing social, cultural, and consumption changes, combining the quality and quantity data collected in order to develop an expected scenario.

The colors suggested by color trends have to play an intense and immediate communication role and must represent first a lifestyle than a consumption style one can identify oneself with. They must belong to the present time, be the continuation of the past, and prove to be able to anticipate the future.

As a matter of fact, the pursuit of novelties and changes cannot deviate much from tradition, from

what consumers have shown to appreciate. Consequently, every color trend will contain some reassuring references in line with the latest successes and some innovations embodying the change.

Work Phases

Pompas points out the following work phases [9]:

- Registration of the latest most successful color trends
- Monitoring and insight of ongoing changes, based on a development process approach
- Collection of quantity and quality data on ongoing changes and their translation into dominant concepts
- Development of a scenario consisting of several images
- Selection of emerging shades and secondary shades
- Organization of a color range or ► [palette](#) that is consistent with dominant concepts
- Suggested ► [color matching](#), consisting in ► [color combinations](#)

Why to Be Aware of Trends

Musso points out about the need to be aware of trends

In the many areas of application and activities that involve the use of color, for instance, in industrial and textile design and in fashion design, only the creators of products who have the right information at the right time can act with advantages in the business environment.

There are many factors that need to be considered when predicting future design and color trends. There is an evolution from one season to the next; there are also important social and economic forces at work. One of the most important factors is the current and future projection of the socioeconomic conditions of the target consumer. Having a clear perception of changes, increasingly accelerated, it is possible to react quickly to the most demanding challenges of today's world.

The need arises from the spread now possible, at all levels, with simultaneity in all parts of the world, of the most important events. Due to the

globalization of the marketplace, these factors are not as segregated by region as in the past. Of course, there are still major cultural influences that affect the interpretation of these factors.

Every new color or design trend starts because of a new influence or change of value perceived by the social group affected by these influences.

There are many reasons for the change of values, one of the most important being a shift in the emotional process of seeing and experiencing these changes. Visual and emotional influences, changes in economic and social circumstances, and perception of new life styles affect the way in which people act in front of design and color. In a global environment of snapshot communications and accelerated pace of social and cultural changes, trends are constantly evolving to reflect those changes. Several organizations and color trend advisory services dedicate themselves to analyze the factors that will influence consumers' emotions. Their members track, analyze, predict, and direct colors and design of the consumer world. These groups base their color and design forecast on trends in fine arts, global events, technological advancements, economical and political circumstances, cultural facts, etc. and on the impact that these combined factors may have on the consumer. The media, opinion makers, specialized fairs, and producers deal with the effects of disclosing these facts and act on the choice of designs and colors. Based on this conceptualization, a new product will come, and it will be suitable to the market. If the designers have accurately identified and understood the customer, they will respond correctly to its own market.

Lifestyle Proposal

In 1985, the style of the products began to take into account the lifestyles [10].

The Advanced Communication Center, in France, divided the French population into five socio-lifestyles which then grew in number, diversifying. This study took into account:

- The places of residence of consumers
- Their behavior and preferences for the indoor environment: materials, objects, furniture, colors, and patterns

- The places where they buy or do their shopping
- The type of information used, newspapers, magazines, television and among other parameters

These data would compose an extremely useful tool to manage the offer of products.

Trends and Revivals

Particularly in the beginning of the twenty-first century, the succession of rapid changes in short and accelerated periods became relevant, leading to the overlapping of different revivals. The change in the nature of the geopolitical and cultural relationship impacts on the representation of the world, in the evolution of the conception and circulation of signs representing a moment. It is therefore essential the conceptual updating of the cultural and aesthetic debate, without neglecting the political, economic, and anthropological impact. It is of great interest therefore to assess how the arts of the image are carriers of elements of exchange, while it is necessary to avoid the danger of uniformity.

To propose a revival involves knowing the reality and the facts that gave rise and its implications. Taking only the outward, without understanding their reasons, leads to superficiality, the mere repetition. Investigating the roots pushes to motivation that is spilling into creativity.

Each of the decade 1970s, 1980s, and 1990s showed its spirit in a complex and multifarious use of color as a clear expression of the social premises of the time in western countries [11]. These trends are reflected in the product catalogs of European and American textile mills and department stores from 1970 to 2013, as well as in the main trade fairs such as Heimtextil (international trade fair for home textiles and commercially used textiles in Frankfurt), Star and Macef (Milan), Paritex and Maison et Objet (Paris), and among others, from 1971 to 2013.

The early 1970s kept the euphoric creativity of the 1960s as regards style. In 1973, however, great structural changes were brought about by the deep crude oil crisis which hit the markets and the economic structures. A feeling of uncertainty and lack of stability got hold of society. In the

field of design, all this resulted in going back to a well-known past in search of the security purported in the everlasting values, either by retrieving old documents or manipulating eclectically their iconography. The desire of running away from reality was shown by all sorts of romantic attitudes such as revivals and retro-proposals. Design turned back to small-scale motifs, geometrical or flowery. The *fleurettes* were monotonously added to all surfaces, and liberty style succeeded in west European countries. The predominant colors accompanying this stylistic shock ranged from medium to dark, grayish colors, non-saturated colors, earthy colors, grayish browns and greens.

The 1980s was the decade of appearance. The utmost egocentricity together with hedonism and obsession for social status was the more significant features in the group engaged in a blind and swift consumerism. It was the yuppie's decade, the decade of the self. The expanding social class was then the new bourgeoisie of the managerial elite, fond of showing off their success and their higher social position. The new style was sophisticated and luxurious, using sumptuous materials and emphasizing the quality in aesthetics. The essential feature was the pluralism, the multiplicity of styles, and the increasing variety of alternatives. The taste of the consumers was the result of their social status, keyed in their lifestyle and in the need of belonging to a social cultural group. It was a decade of a remarkable awareness of color, and thus color was conscientiously made the protagonist. High tech embraced industrial objects made of metal, glass, Formica, and plastic. Black made its appearance, stepping into the picture on its own, or associated with white and red, as well as with metallic colors, silver and gold, so as to emphasize the luxury effect [10, 12].

In the 1990s, the end of the Cold War brought about hope for the end of the nuclear menace. A new trend of thought set forward a deeper awareness of the environment in danger as much as greater concern about nature. Nature and its preservation became one of the main issues to deal with as well as a concern for health care and family life. A new conservative attitude was the most remarkable trend of this period, giving way

to substantial subjects as ecology, the protective home, the enhancement of native roots and traditions, and the return to reassuring values. The new commandments were avoiding pollution, the efficient use of means and resources, the importance of quality instead of quantity, and the respect for nature. Attention was drawn in special toward well-done high-quality work and craftsmanship. The search for true moral values was shown by the use of noble materials, and, in the choice of color, the main one was that of unbleached linen. The favorite colors were those of different types of wood, earth, grain, cereals, straw, sand, stone, and grayish colors that seemed to be worn out. Everyone celebrated nature in their own way: by means of choice of material, color, or subject [13, 14].

Trends for the House of the Century

The trends take into account the desire to expend more time at home, the valorization of empty space and customization of the environment, and the awareness for the preservation of natural resources. The market increasingly focused on products for the home. The decoration is freer, and nothing is definitive. Furniture is polyvalent, with fewer objects. The consumer seeks welfare, looks for quality, and wants a more simplified life. Pragmatism and health are at the order of the day.

In the new millennium, the house gains a new dimension. Tradition and modernity coexist in harmony. Earth, fire, and water are the three basic elements inspiring styles and colors. Individuality and daily rituals of life point the products chosen [14].

The configuration of a universal culture becomes a fundamental sign. The process of globalization tends to the unification of symbols, to the disappearance of diversity. It is therefore crucial to defend the values of each culture, using its imprint, to highlight its mark in every corner of the world, each time the triumph of cosmopolitanism can delete it.

Cross-References

- [Appearance](#)
- [Color Combination](#)

- [Color Preference](#)
- [Dye](#)
- [Palette](#)
- [Pastel Colors](#)

References

1. Oberascher, L.: Colour trends – why we need them. In: Smith, D., Green-Armytage, P., Pope, M., Harkness, N. (eds.) *Proceedings of the 11th Congress of the International Colour Association (AIC 2009)*, CD. Colour Society of Australia, Sydney (2009)
2. Darmstadt, C.: *Farbbewegung in der Architektur-gestaltung*, unpublished manuscript (1985)
3. Koppelman, U., Küthe, E.: Präferenzwellen beim Gestaltungsmittel Farbe. *Marketing, Zeitschrift für Forschung und Praxis* **2**, 113–121 (1987)
4. Schlegel, M.: *Farb- und Materialscouting 1955–2005, Trends 2006–2007, Farbe – Struktur – Oberfläche*. Caparol Farben Lacke Bautenschutz (2006)
5. Oberascher, L.: Cyclic recurrence of collective colour preferences. In: Linton, H. (ed.) *Color Forecasting. A Survey of International Color Marketing*. Van Nostrand, New York (1994)
6. Pompas, R.: *The Trilogy of Precious Dyestuffs from the Mediterranean. Purple, Kermes, and Woad*. Textile Forum, Hannover (2001)
7. Luzzatto, L., Pompas, R.: *I colori del vestire. Variazioni–Ritorni–Persistenze*. Hoepli, Milan (1997)
8. Pompas, R.: La moda e il consumo del colore: quale approccio? In: Rizzo, S. (ed.) *Colore e design tra comunicazione e produzione*. De Ferrari, Genoa (2010)
9. Pompas, R.: *Textile Design. Ricerca–Elaborazione–Progetto*. Hoepli, Milan (1994)
10. Musso, M.L.: Black and status in the ‘80s. In: Kortbawi, I., Bergström, B., Fridell Anter, K. (eds.) *AIC 2008, Colour – Effects & Affects, Proceedings of the Interim Meeting of the International Colour Association, Paper 119*. Swedish Colour Centre Foundation, Stockholm (2008)
11. Musso, M.L.: Colour in the ‘70s, ‘80s, ‘90s, as expression of changes in society. In: Nieves, J.L., Hernández-Andrés, J. (eds.) *AIC Colour 05, Proceedings of the 10th Congress of the International Colour Association*, pp. 1589–1591. Comité Español del Color, Granada (2005)
12. Bayer, A.G.: *Trend Colour Charts*. Bayer AG, Leverkusen, Germany (1974–1990)
13. Musso, M.L.: Ecology and colour in the ‘90s. In: Ye, G., Xu, H. (eds.) *AIC 2007, Color Science for Industry, Proceedings of the Midterm Meeting of the International Color Association*, pp. 158–161. Color Association of China, Hangzhou (2007)
14. Hoechst, A.G., Trevira, GmbH.: *Trends in Living. Trevira Colour Chart*. Hoechst & Trevira, Frankfurt (1981 to 2013)

Color Union

- [Color Combination](#)

Color Uses

- [Functionality of Color](#)

Color Vision Abnormality

- [Tritanopia](#)

Color Vision Screening

- [Color Vision Testing](#)

Color Vision Test

- [Pseudoisochromatic Plates](#)

Color Vision Testing

Galina V. Paramei¹ and David L. Bimler²
¹Department of Psychology, Liverpool Hope University, Liverpool, UK

²School of Psychology, Massey University, Palmerston, North, New Zealand

Synonyms

[Clinical color vision tests](#); [Color vision screening](#); [Diagnosis of defective color vision](#)

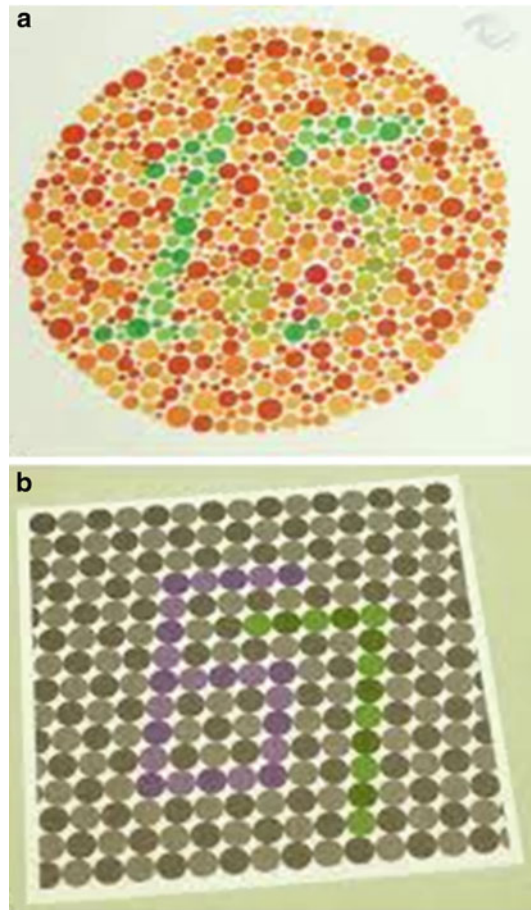
Definition

Color vision testing is the assessment of chromatic discrimination ability and the diagnosis of any perceptual deficiency according to its severity and quality (see Paramei and Bimler, “► [Protanopia](#)”; Paramei and Bimler, “► [Deuteranopia](#)”; Bimler and Paramei, “► [Tritanopia](#)”; Rodríguez-Carmona, “► [Environmental Influences on Color Vision](#)”). Tests vary in sensitivity, specificity, ease of use, and time required for administration [1–4]. Many were designed primarily for vocational screening for congenital deficiency, an issue in any occupation where color-coding conveys information (e.g., railways, aviation, electronics) [5, 6]. Testing is also important for assessing and monitoring acquired color abnormality, appearing as a manifestation of visual-system pathology resulting from ophthalmological diseases (e.g., glaucoma, ocular hypertension), systemic or neurological diseases (e.g., diabetes, Parkinson’s), or the effects of medications or exposure to environmental/occupational toxins (see Paramei, “► [Color Perception and Environmentally Based Impairments](#)”).

Color vision tests fall into several broad categories [1–4]. ► [Pseudoisochromatic plates](#) and arrangement tests both directly address an observer’s confusions between color pairs along a given confusion axis. Both types of tests have the advantage of rapid administration and ease of interpretation, making them suitable for field/epidemiological applications. Generally they distinguish between the protan, deutan, and tritan forms of deficiency. Their results emphasize a dichotomous outcome: whether a subject’s color sensitivity is (vocationally) impaired. Versions of the tests exist for testing color vision in children. More recent computerized developments measure variations in color perception along a continuous range. Matching tests and naming (lantern) tests follow different principles. Below, the most widely used tests are described in more detail.

Pseudoisochromatic Tests

Pseudoisochromatic tests (Fig. 1) ► [Pseudoisochromatic Plates](#) are irregular mosaics of small circles, randomly varying in size and luminance.

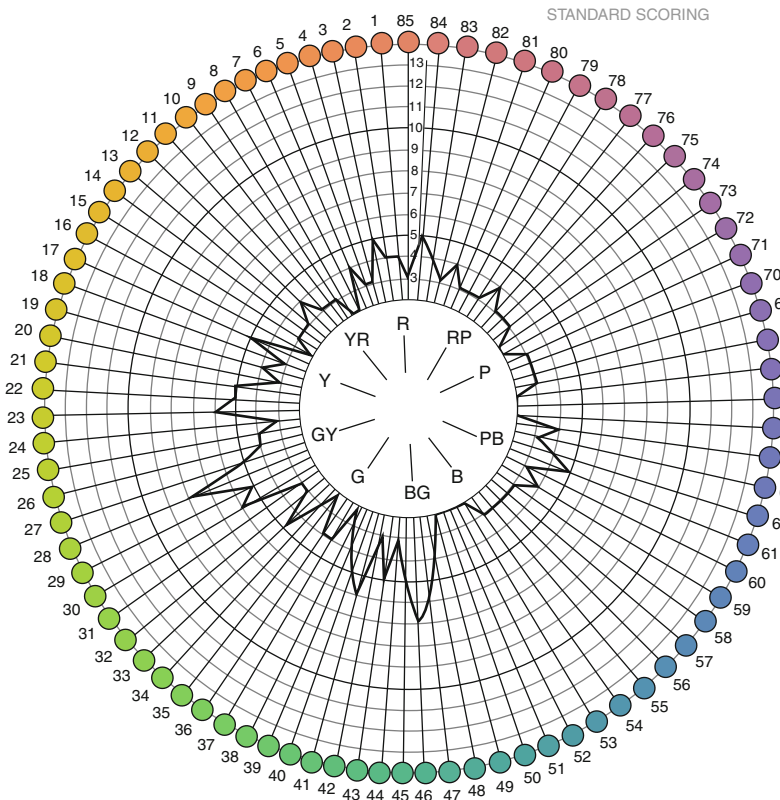


Color Vision Testing, Fig. 1 Examples of the Ishihara pseudoisochromatic plates used for screening for red-green deficiency (*top*) and Hardy-Rand-Rittler plates for screening tritan defects (*bottom*) (Source: Jägle, H., Zrenner, E. Krastel, H., W. Hart. W.: Dyschromatopsias associated with neuro-ophthalmic disease. Ch. 6. In: Schiefer, U., Wilhelm, H., Hart, W. (eds.), *Clinical Neuro-Ophthalmology. A Practical Guide* (2007), Fig. 6.2. Springer Copyright Clearance Center. Licence Number: 3655300610668)

Color differences among the circles demarcate a foreground design (digits, simple geometric forms, or curved lines), so that for a normal trichromatic observer the design stands out by Gestalt fusion from its background. In diagnostic plates the defining color difference disappears for color-deficient observers, and the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction. The spatial fluctuations in the mosaic mask any residual luminance traces of the normal design.

Color Vision Testing,

Fig. 2 Example of the Farnsworth-Munsell 100 Hue test scoring sheet indicating a moderate color discrimination defect (Source: Jameson, K.A. Human potential for tetrachromacy. *Glimpse Journal: The Art + Science of Seeing* 2.3 (2009), Online Supplementary Material, p. 4, Figure 3; <http://www.glimpsejournal.com/2.3-KAJ.html>. Copyright (2009) Kimberly A. Jameson, All Rights Reserved)



The Ishihara test, used most widely, is intended for diagnosis of congenital red-green deficiency (“daltonism”), differentiating its two types, protans and deutans, and severity (mild, moderate, or extreme) [6, 7] (see Paramei and Bimler, “► [Protanopia](#)”; Paramei and Bimler, “► [Deuteranopia](#)”). The Hardy-Rand-Rittler (HRR) test contains additional six plates designed to detect tritan defects and gauge their severity [7, 8]. There also exists a Farnsworth F2 plate designed specifically for revealing tritan abnormality (see Bimler and Paramei, “► [Tritanopia](#)”; Paramei, “► [Color Perception and Environmentally Based Impairments](#)”).

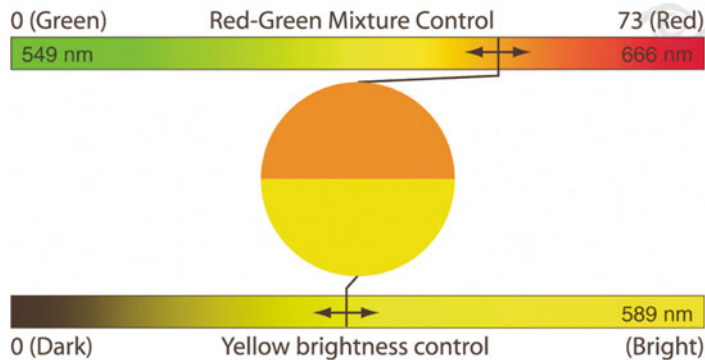
Color Arrangement Tests

Arrangement/panel tests use a set of color stimuli (“caps”) which sample a color circle (see Green-Armytage, “► [Color Circle](#)”) at regular intervals. The subject is requested to arrange them in sequence, so that each color lies between the two colors most similar to it. Transpositions of the

caps, departing from a normal trichromat’s sequence, are recorded as errors. These departures can be plotted graphically to measure the angle of the confusion axis (if any) and summed to quantify the severity of any deficit.

The classical example is the Farnsworth-Munsell 100-Hue (FM100) test [1–4, 8, 9], consisting of 85 caps, which takes 20–30 min to complete. Errors (transpositions) peak in the sectors of the color circle running tangential to the protan, deutan, or tritan confusion axis (Fig. 2). Performance on the FM100 improves with repetition, and as well as measuring color discrimination, it is affected by general nonverbal intelligence [10].

Two shorter versions, the Farnsworth Dichotomous D-15 test and the Lanthony Desaturated D-15d, each contain only 15 movable caps plus a fixed “pilot cap” as the start of the sequence [9] and take about 5 min to complete. The D-15 is designed to diagnose moderate to severe color defects. The D-15d test uses color samples that



Color Vision Testing, Fig. 3 Rayleigh spectral matching in the Nagel anomaloscope (Source: Jägle, H., Zrenner, E. Krastel, H., W. Hart. W.: *Dyschromatopsias associated with neuro-ophthalmic disease*. Ch. 6. In: Schiefer, U.,

Wilhelm, H., Hart, W. (eds.), *Clinical Neuro-Ophthalmology. A Practical Guide* (2007), Fig. 6.5. Springer Copyright Clearance Center. Licence Number: 3655300610668)

are lighter and paler [11]. It was designed specifically to capture mild or subclinical color defects in observers who pass the standard D-15 test. Errors can include diametrical circle-crossing transpositions, indicating the protan, deuten, or tritan confusion axis when plotted graphically (see Fig. 1 in Paramei, “► [Color Perception and Environmentally Based Impairments](#)”). Outcomes of both tests can be summarized as a color confusion index (CCI), where 1.0 corresponds to perfect color arrangement and CCI values greater than 1.0 indicate progressive impairment of color discrimination. These tests are often used in conjunction, though the more sensitive D-15d is widely employed for early detection of mild acquired dyschromatopsias.

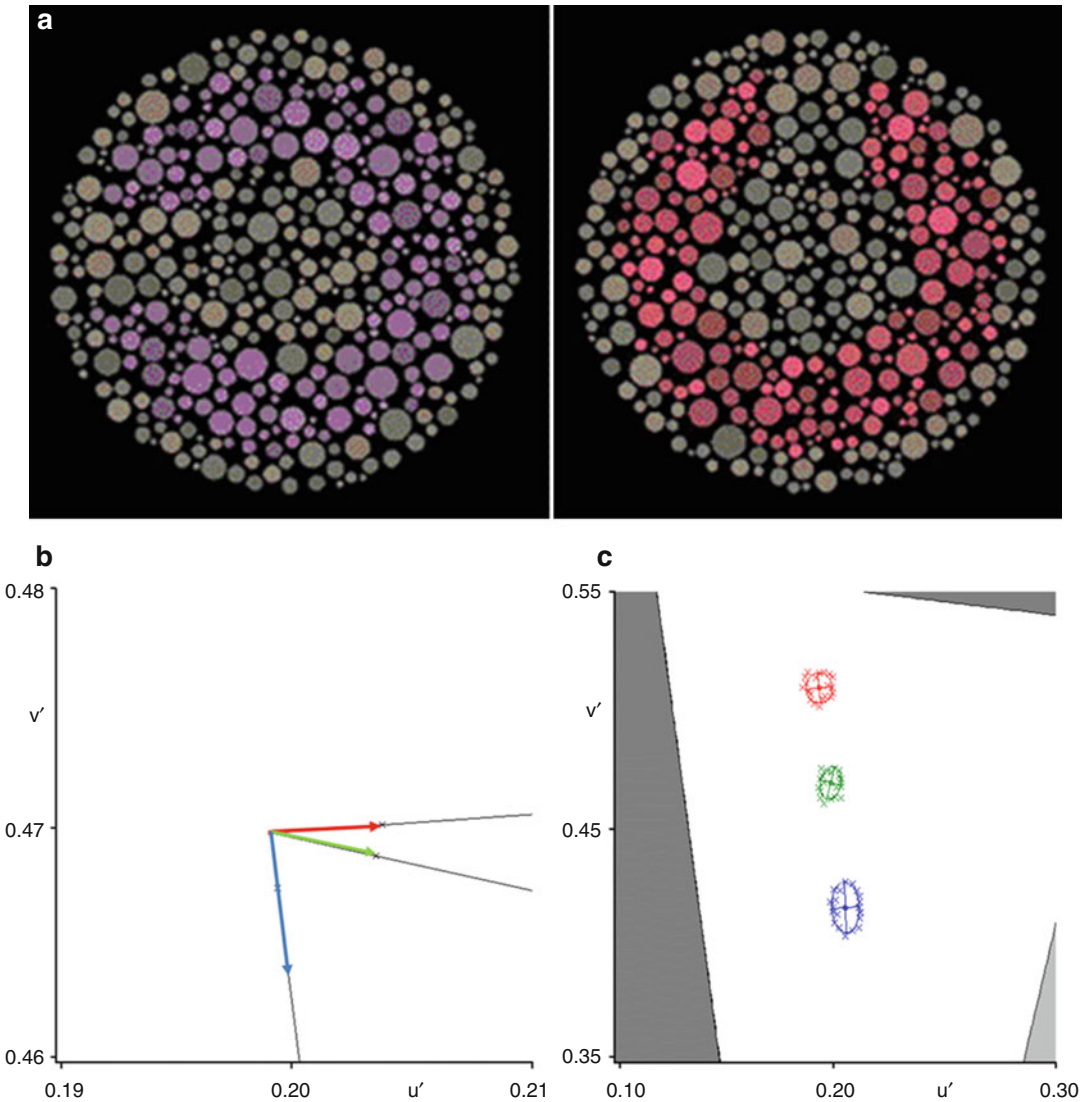
The Lanthony New Color Test [3, 4] comprises four panels of 15 caps each at four levels of saturation, to examine color similarities at four different scales. Like the D-15d, it can be used to track the progression of acquired dyschromatopsias.

Although not strictly an arrangement test, the City University Test [2–4] is derived from the D-15. It is a forced-choice test consisting of 10 panels, each presenting four colored dots in a quincunx around a central dot; the subject has to indicate which of the four colors most closely resembles the central one. The colors are selected so that protan, deutan, or tritan deficiencies affect which dot is subjectively most similar to the center.

Anomaloscopes

Within color-matching tests, the “gold standard” of color deficiency diagnosis are anomaloscopes, which present colors as monochromatic light rather than on a computer monitor or via reflective pigments. Compared to the pseudoisochromatic and panel tests, anomaloscopy requires a skilled examiner.

The Nagel anomaloscope is intended for assessment of red-green discrimination (Fig. 3). The observer views a 2° hemipartite circle, where one half is yellow light (589 nm), while the other half-circle mixes red (666 nm) and green light (549 nm) – known as the Rayleigh equation [1–3]. These wavelengths differ only in their relative stimulation of L- and M- cones (S-cones being unresponsive in this spectral range) (see Stockman, “► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)”). A normal trichomat’s setting is characterized by a green/red ratio around 41 (on the scale between 0 and 73) and a very narrow range of such settings upon retest. Anomalous trichomats accept a wider than normal range of mixed lights as indistinguishable from the yellow, with the range – from narrow to very broad – characterizing mild, moderate, or extreme impairment. A dichromat can match *any* red/green ratio to the yellow light by adjusting the luminance of the latter, which a protanopic dichromat dims if the mixed light is mainly red whereas a green-dominated light requires a more luminant



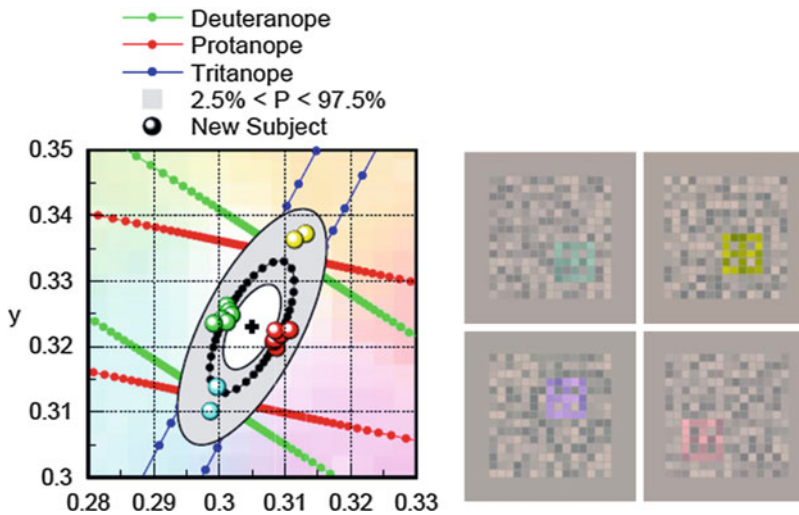
Color Vision Testing, Fig. 4 The Cambridge Colour Test. (a) An illustration of chromatic targets, Landolt “C,” embedded in the luminance noise background. (b) Confusion vectors (in CIE 1976 $u'v'$ chromaticity diagram) along which the chromaticity is varied in the CCT Trivector test: Protan (red), Deutan (green), and Tritan (blue). The origin of the vectors indicates chromaticity coordinates of the neutral background ($u' = 0.1977$,

$v' = 0.4689$). (c) Examples of chromatic discrimination ellipses for normal trichromats: Ellipse 1 (middle), Ellipse 2 (top), Ellipse 3 (bottom); crosses indicate raw discrimination vectors, fitted ellipses are shown by solid lines (Figures 3a, c; Source: Mollon, J.D., Regan, B.C.. Cambridge Colour Test. Handbook (Cambridge Research Systems Ltd., 2000), p. 4. Permission has been obtained from Prof. John D. Mollon, who holds the copyright)

yellow; the opposite is observed for a deuteranope (see Paramei and Bimler, “► Protanopia”; Paramei and Bimler, “► Deuteranopia”).

The Moreland anomaloscope serves to assess tritan discrimination. One half of the 2° hemipartite circle is a cyan standard (480 nm

light tinged with a small admixture of 580 nm), which must be matched by mixing indigo (436 nm) and green lights (490 nm) in the other half-circle, known as the Moreland equation [12]. Decreasing discrimination along the tritan confusion lines increases the range of mixtures



Color Vision Testing, Fig. 5 An illustration of the Colour Assessment and Diagnosis (CAD) test (*right*). Direction-specific, color-defined moving stimuli must be detected against a background with random, dynamic luminance contrast (Source: Barbur, J.L., Rodríguez-Carmona, M.: Variability in normal and defective colour

vision: consequences for occupational environments. In: Best, J. (ed.) *Colour Design: Theories and Application*, pp. 24–82. Woodhead Publishing, Philadelphia (2012). P. 52, Fig. 2.13(a, b). The authors hold the copyright; permission has been obtained from Prof. John Barbur)

which perceptually match the standard. Notably, at 8° only complete tritanopes accept the full range of color mixtures as a match to the cyan standard (see Bimler and Paramei, “► [Tritanopia](#)”).

Computerized Tests

More recently, computerized equivalents of pseudoisochromatic tests have become common, displaying a series of colored mosaics in which the elements vary in luminance spatially as well as dynamically to leave only chromatic cues. Employed on a calibrated monitor under strict psychophysical protocols, the tests allow precise measurement of chromatic sensitivity.

In the Cambridge Colour Test (CCT) [13] (Fig. 4a) the design in each display is a stylized letter “C” with a four-way choice for the orientation for the letter’s open side (see Paramei and Bimler, ► [Protanopia](#); Paramei and Bimler, “► [Deuteranopia](#)”; Bimler and Paramei, “► [Tritanopia](#)”). The magnitude of the color difference demarcating each design (i.e., the difficulty of the choice) varies interactively in response to the observer’s ongoing performance,

to specify the direction in the color plane of elevated thresholds, and to “bracket” their discrimination in the frame of the CIE ($u'v'$) 1976 chromaticity diagram (see Schanda, “► [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)”). Outcomes are chromatic discrimination thresholds along the protan, deutan, and tritan confusion lines (Trivector subtest (Fig. 4b)) and elongation/orientation parameters for three MacAdam ellipses (Ellipses subtest (Fig. 4c)). CCT normative data for normal trichromats for eight life decades track the impact of age upon chromatic discrimination [14]. The CCT has been employed in numerous clinical studies for differential diagnostics of relative damage to chromatic pathways (e.g., [15]).

The Colour Assessment and Diagnosis (CAD) test [5, 6] employs spatiotemporal luminance contrast masking. Direction-specific, color-defined moving stimuli must be detected against a background of random, dynamic luminance contrast (Fig. 5, right). Chromatic sensitivity is measured in 16 color directions in the CIE (x,y) 1931 chromaticity diagram. From these, mean thresholds are computed for thresholds of the red-green and

blue-yellow systems (Fig. 5, left), ► [Color Vision, Opponent Theory](#) to diagnose accurately the subject's class of color vision (i.e., normal, protan-, deutan-, tritan-like congenital loss or acquired color deficiency). The CAD test has been extensively applied for screening in occupations with color-intensive visually demanding tasks, in particular in aviation [5, 6]. The CAD units of chromatic sensitivity are based on the mean thresholds measured in 333 young normal trichromats. CAD data on protan, deutan, and tritan thresholds for normal trichromats across the lifespan have recently been obtained (in press).

Cross-References

- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE u', v' Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision, Opponent Theory](#)
- [Deutanopia](#)
- [Environmental Influences on Color Vision](#)
- [Protanopia](#)
- [Pseudoisochromatic Plates](#)
- [Tritanopia](#)

References

1. Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L. G. (eds.): *Congenital and Acquired Color Vision Defects*. Grune and Stratton, New York (1979)
2. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
3. Dain, S.: Clinical colour vision tests. *Clin. Exp. Optom.* **87**, 276–293 (2004)
4. Cole, B.L.: Assessment of inherited colour vision defects in clinical practice. *Clin. Exp. Optom.* **90**, 157–175 (2007)
5. Barbur, J.L., Rodríguez-Carmona, M.: Variability in normal and defective colour vision: consequences for occupational environments. In: Best, J. (ed.) *Colour Design: Theories and Application*, pp. 24–82. Woodhead Publishing, Philadelphia (2012)
6. Rodríguez-Carmona, M., O'Neill-Biba, M., Barbur, J. L.: Assessing the severity of color vision with

implications for aviation and other occupational environments. *Aviat. Space Environ. Med.* **83**, 19–29 (2012)

7. Birch, J.: Identification of red-green colour deficiency: sensitivity of the Ishihara and American Optical Company (Hard, Rand and Rittler) pseudo-isochromatic plates to identify slight anomalous trichromatism. *Ophthal. Physiol. Opt.* **30**, 667–671 (2010)
8. Foote, K.G., Neitz, M., Neitz, J.: Comparison of the Richmond HRR 4th edition and Farnsworth-Munsell 100 Hue Test for quantitative assessment of tritan color deficiencies. *J. Opt. Soc. Am. A* **31**, A186–A188 (2014)
9. Farnsworth, D.: Farnsworth-Munsell 100-hue and dichotomous tests for color vision. *J. Opt. Soc. Am.* **33**, 568–578 (1943)
10. Cranwell, M.B., Pearce, B., Loveridge, C., Hurlbert, A.: Performance on the Farnsworth-Munsell 100-Hue test is significantly related to non-verbal IQ. *Invest. Ophthalmol. Vis. Sci.* **56**, 3171–3178 (2015)
11. Lanthony, P.: The desaturated panel D-15. *Doc. Ophthalmol.* **46**, 185–189 (1978)
12. Moreland, J.D., Young, W.B.: A new anomaloscope employing interference filters. *Mod. Probl. Ophthalmol.* **13**, 47–55 (1974)
13. Mollon, J.D., Regan, B.C.: *Cambridge Colour Test Handbook*. Cambridge Research Systems Ltd. (2000). <http://www.crs Ltd.com/tools-for-vision-science/measuring-visual-functions/cambridge-colour-test/>
14. Paramei, G.V., Oakley, B.: Variation of color discrimination across the life span. *J. Opt. Soc. Am. A* **31**, A375–A384 (2014)
15. Feitosa-Santana, C., Paramei, G.V., Nishi, M., Gualtieri, M., Costa, M.F., Ventura, D.F.: Color vision impairment in type 2 diabetes assessed by the D-15d test and the Cambridge Colour Test. *Ophthal. Physiol. Opt.* **30**, 717–723 (2010)

Color Vision, Opponent Theory

Sophie Wuerger¹ and Kaida Xiao²

¹Department of Psychological Sciences, University of Liverpool, Liverpool, UK

²Department of Psychological Sciences, Institute of Psychology, Health and Society, University of Liverpool, Liverpool, UK

Synonyms

[Color-opponent processing](#); [Colour-opponent processing](#); [Hue opponency](#); [Opponent color theory](#)

Definition

Opponency in human color vision refers to the idea that our perceptual color mechanisms are arranged in an opponent fashion. One mechanism, the red-green mechanism, signals colors ranging from red to green; the other one, the yellow-blue mechanism, signals colors ranging from yellow to blue. This opponency is often referred to as hue opponency, as opposed to cone opponency.

this mechanism (Fig. 2) and this mechanism is at an equilibrium. The colors for which the yellow-blue mechanism is at an equilibrium are called “unique red” or “unique green,” depending on which side of the neutral point they lie. Colors that silence the red-green opponent mechanisms are called “unique yellow” or “unique blue.” Hurvich and Jameson [2] were the first to use a hue cancellation procedure to find the equilibria points (null responses) of these opponent mechanisms.

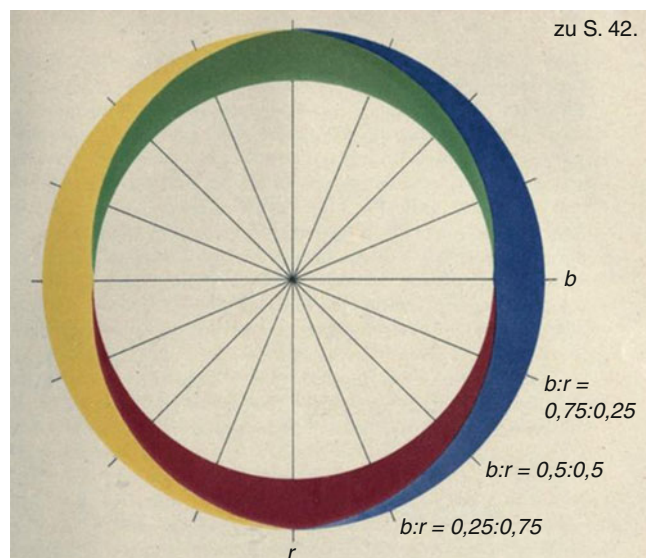
Behavioral Evidence for Color-Opponent Processing

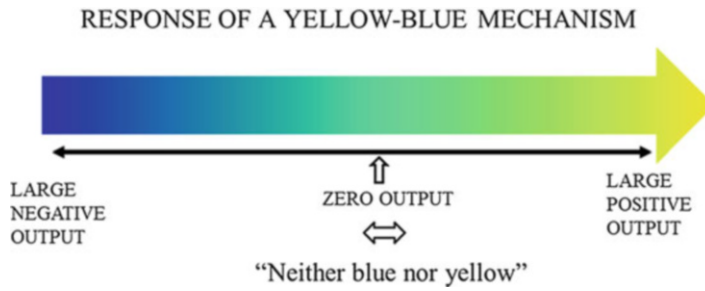
Hering [1] was the first to notice that some pairs of colors, namely, red and green and yellow and blue, cannot be perceived at the same time. He named these pairs of colors “Gegenfarben” [opponent colors] since they are mutually exclusive colors; in Hering’s original figure (Fig. 1), this mutual exclusivity is conveyed by the lack of overlap between red and green and between yellow and blue. The idea is that these opponent colors constitute the end points of the two chromatic mechanisms. For example, the putative yellow-blue mechanism signals colors from yellow to blue; if the stimulus contains neither yellow nor blue, then this stimulus elicits no response in

Linearity and Constancy of the Color-Opponent Mechanisms

Krantz and colleagues [3, 4] tested the linearity of these color-opponent mechanisms and concluded that the red-green opponent mechanism is approximately linear but the yellow-blue mechanism exhibited significant deviations from linearity and additivity. These basic results have been replicated by numerous studies, either using a hue cancellation procedure [5, 6] or a modified hue selection task [7], as shown in Fig. 3. Here the task of the observer was to choose that patch of light that appears “neither yellow nor blue” (to obtain unique red and green) or “neither red nor green” (to obtain unique yellow and blue). Figure 4 shows the unique hue settings for 185 color-

Color Vision, Opponent Theory, Fig. 1 Hering [1], page 42 *ibid.* The opponent colors, *red* versus *green* and *yellow* versus *blue*, are mutually exclusive. This idea is conveyed in Hering’s original figure by showing that there is no overlap between *red* and *green* and between *yellow* and *blue*





Color Vision, Opponent Theory, Fig. 2 The response of a putative yellow-blue mechanism is shown. When this mechanism is at equilibrium, i.e., produces zero output in response to a stimulus, then this stimulus is – by

definition – perceived as “neither blue nor yellow.” Stimuli that are perceived as “neither blue nor yellow” are defined as “unique red” or “unique green,” depending on which side of the neutral gray point they are located



Color Vision, Opponent Theory, Fig. 3 Modified hue selection task to obtain unique red [7]. The task of the observer is to identify the patch which appears “neither yellow nor blue.” Stimuli that are perceived as “neither blue nor blue” are defined as “unique red” or “unique green,” depending on which side of the neutral gray point they are located

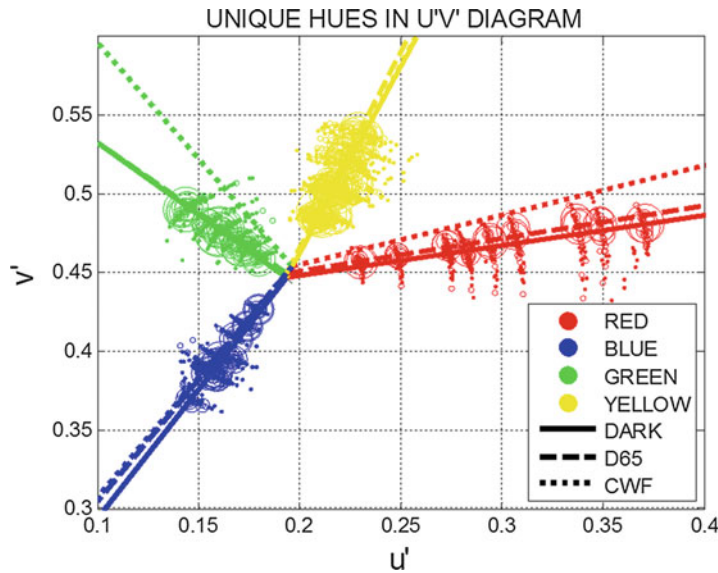
normal observers using the hue selection task [8, 9], plotted in an approximately uniform u', v' chromaticity diagram. All experiments were conducted under three different viewing conditions: either under dark conditions (only source of illumination was the gray monitor background), under adaptation to D65 (daylight simulator), or under adaptation to typical office light (CWF). The symbols and the solid lines (first principal component) denote the unique hues

under the dark viewing condition; dashed and dotted lines are the first principal components under adaptation to D65 or CWF.

Figure 4 demonstrates the two major features of the color-opponent mechanisms: (1) Consistent with Krantz and colleagues, the red-green opponent mechanism was found to be approximately linear, that is, unique yellow and blue are colinear [3]; the yellow-blue opponent mechanism on the other hand is not a single linear mechanism [4], that is, unique red and green do not lie on a line through the neutral gray background. Therefore, one needs to either postulate a highly nonlinear yellow-blue mechanism or, which is more likely, two separate unipolar mechanisms, one signaling yellow and the other one signaling blue. (2) The red-green opponent mechanism is fairly color constant in comparison to the yellow-blue opponent mechanism; the unique yellow and blue settings are not affected by the changes in the ambient illumination (solid, dashed, and dotted lines are coincident), while large shifts in unique hue settings are observed for the yellow-blue equilibria. Unique green settings, in particular, shift toward yellow when viewed under typical office light (CWF).

Physiological Basis for Opponent Hue Processing: Hue Opponency Versus Cone Opponency

The physiological basis of these color-opponent (hue-opponent) mechanisms is still unclear.

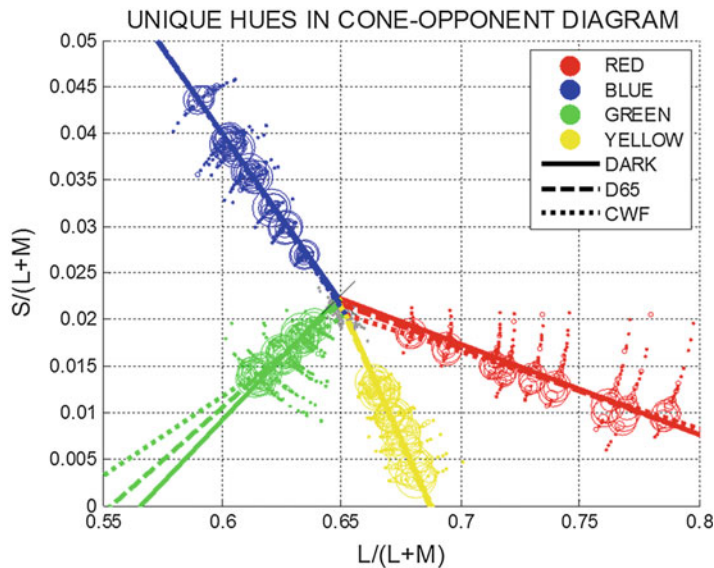


Color Vision, Opponent Theory, Fig. 4 Unique hue settings for 185 color-normal observers are plotted in an approximately uniform u',v' diagram. A hue selection task was used to obtain the unique hues [7]. Symbols are proportional to the number of data points and are only shown for the dark viewing condition for clarity. For each unique

hue, a total of 1,665 data points are shown (185 observers \times 9 saturation/lightness combinations). The *solid*, *dashed*, and *dotted lines* indicate the first principal component for the dark, D65 (daylight simulator), and CWF (office light) viewing conditions

Color Vision, Opponent Theory, Fig. 5

Unique hue settings for 185 color-normal observers are plotted in a cone-opponent diagram. For details, see Fig. 4



We know that the early cone-opponent mechanisms that originate in the retina and are inherited by the lateral geniculate nucleus [10] are not the neural substrate of the behaviorally measured hue opponency. This is easily seen by replotting the

unique hue settings in a cone-opponent space. Figure 5 shows the data in an isoluminant Boynton-MacLeod diagram (BML); the $L/(L + M)$ axis denotes the L-M cone-opponent mechanism; the $S/(L + M)$ denotes the S-(L + M) cone-

opponent mechanism. Apart from scaling, the BML diagram and the DKL space (Derrington-Krauskopf-Lennie space) are identical for isoluminant stimuli. The BML diagram is not intended to be a uniform chromaticity diagram, but the basic features of the color-opponent mechanisms can also be seen here. (1) It is clear that the color-opponent mechanisms (unique hue lines) are not aligned with the cone-opponent axes ($L/(L + M)$; $S/(L + M)$). The red-green equilibria (line connecting unique yellow and blue) are not parallel to the $S/(L + M)$ cone axis; the yellow-blue equilibria (line connecting unique red and green) are not aligned with the $L/(L - M)$ axis. In reference to the gray background ("x" in the middle of Fig. 5), unique blue requires, in addition to the S-cone input, also an M-cone input; unique yellow requires a negative S-cone input and a positive L-cone input. Similarly, unique green requires a negative S-cone and a positive M-cone input. Unique red is the only hue that is at least approximately aligned with the $L/(L + M)$ axis; but even here a systematic negative S-cone input is required. This confirms that the early cone-opponent mechanisms (depicted by the axes in Fig. 5) do not constitute the neural mechanisms underlying the observed hue opponency (lines connecting the unique hues). Further recombinations of the early opponent mechanisms must occur between the LGN and the primary visual cortex or within the visual cortical areas.

Conclusions

Firstly, the color-opponent mechanisms obtained using behavioral measures such as hue cancellation are not aligned with the cone-opponent mechanisms that have been identified in the retina and the lateral geniculate nucleus. It may therefore be more appropriate to refer to these mechanisms as hue-opponent and cone-opponent mechanisms, respectively. Secondly, the red-green opponent mechanism (yielding unique yellow and blue) is an approximately linear mechanism. In contrast, the yellow-blue opponent mechanism (yielding unique red and green) cannot be modeled as a

single linear opponent mechanism since unique red and green do not lie on a line through the neutral gray origin. The most parsimonious explanation is to postulate two separate yellow-blue mechanisms; when at equilibrium, one of them signals red, the other one green. Thirdly, while the red-green opponent mechanism is almost completely color constant (unique yellow and blue settings are invariant under changes in ambient illumination), the equilibria point of the yellow-blue mechanism change under changes in ambient illumination: unique green in particular undergoes a major shift toward yellow when viewed under CWF in comparison to simulated daylight (D65). We speculate that this failure of constancy for unique green might have the same neural origin as the relatively large interobserver variability found in unique green settings [11, 12].

Cross-References

- [Adaptation](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Constancy](#)
- [Color Processing, Cortical](#)
- [Cone Fundamentals](#)
- [Unique Hues](#)

References

1. Hering, E.: *Grundzüge der Lehre vom Lichtsinn*. Julius Springer, Berlin (1920)
2. Jameson, D., Hurvich, L.: Some quantitative aspects of an opponent-colors theory. I. Chromatic responses and spectral saturation. *J. Opt. Soc. Am.* **45**, 546–552 (1955)
3. Larimer, J., Krantz, D., Cicerone, C.: Opponent-process additivity. I: red/green equilibria. *Vision Res.* **14**, 1127–1140 (1974)
4. Larimer, J., Krantz, D., Cicerone, C.: Opponent-process additivity. II: yellow/blue equilibria and nonlinear models. *Vision Res.* **15**, 723–731 (1975)
5. Webster, M.A., Miyahara, E., Malkoc, G., Raker, V.E.: Variations in normal color vision. II. Unique hues. *J. Opt. Soc. Am. A* **17**, 1545–1555 (2000)
6. Werner, J.S., Wooten, B.R.: Opponent chromatic mechanisms: relation to photopigments and hue naming. *J. Opt. Soc. Am.* **69**, 422–434 (1979)

7. Wuerger, S.M., Atkinson, P., Cropper, S.J.: The cone inputs to the unique-hue mechanisms. *Vision Res.* **45**, 3210–23 (2005)
8. Wuerger, S.: Colour constancy across the life span: evidence for compensatory mechanisms. *PLoS One* **8**, e63921 (2013)
9. Xiao, K., Fu, C., Mylonas, D., Karatzas, D., Wuerger, S.: Unique hue data for colour appearance models. Part II: chromatic adaptation transform. *Color. Res. Appl.* **38**, 22–29 (2013)
10. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (1984)
11. Kuehni, R.G.: Unique hues and their stimuli – state of the art. *Color. Res. Appl.* **39**, 279–287 (2014)
12. Mollon, J.D., Jordan, G.: On the nature of unique hues. In: Murray, I., Carden, D., Dickinson, C. (eds.) *John Daltons colour vision legacy*, pp. 381–392. Taylor and Francis, London (1997)

Color Wheel

► Color Circle

Colorant

- Dye
 - Pigment, Ceramic
-

Colorant, Environmental Aspects

Harold Freeman
Fiber and Polymer Science Program, North
Carolina State University, Raleigh, NC, USA

Synonyms

[Eco-/genotoxicity of synthetic dyes](#); [Environmental chemistry of synthetic dyes](#); [Structure property relationships](#)

Definition

The environmental characteristics of colorants (dyes and pigments) encompass topics pertaining

to the impact of such compounds on human health and the environment. Regarding human health, the concern is the potential for colorants to exhibit genotoxicity, namely, to cause adverse interactions between DNA and various compounds that produce a heritable change in the cell or organism. Heritable changes in humans include birth defects, carcinogenesis, teratogenesis, and other types of diseases. Mutations caused by molecular interactions with DNA are generally viewed as the first events in the onset of carcinogenesis. Consequently, screening methods have been developed to determine the mutagenic potential of colorants. Such tests include in vitro methods that use micro-organisms (e.g., bacteria) or isolated tissues as substitutes for whole animal (in vivo) studies. The potential genotoxicity of dyes for various applications came to the forefront in the 1960s when it was found that azo dye manufacturing involving benzidine and 2-naphthylamine as precursors contributed to bladder cancers among plant workers. This outcome led to extensive testing of azo dyes and their aromatic amine precursors for mutagenic and/or carcinogenic potential. Regarding the environment, the key concern is the potential for colorants to harm aquatic life (plants or animals) or pollute drinking water. This topic became a matter of concern because as much as 15 % of the colorants produced can be lost during their manufacture and end-use application [1]. In the case of dyes, the principal source of losses is water-soluble colorants remaining in dyebaths following textile dyeing processes. Consequently, methods pertaining to the treatment of industrial wastewater prior to the release of effluents have been developed, along with pollution prevention methods, as key components of environmental stewardship.

Overview

The environmental properties of colorants are often determined by employing a battery of tests that are concerned principally with the potential for mutagenicity, carcinogenicity, and aquatic toxicity. An overview of progress in these areas is presented.

Mutagenicity

Some dyes are known to exhibit mutagenicity, and the most commonly used test for assessing mutagenicity potential is the reverse mutation assay employing specially engineered *Salmonella* bacteria. The standard *Salmonella*/mammalian microsome assay, often called the Ames test, is now the most widely used protocol as an initial screening test procedure for assessing the mutagenic potential of new experimental colorants. It was introduced in 1975 by Ames and co-workers [2], who observed that most mutagens were also carcinogens and that the extent of a compound's mutation of DNA was related to its carcinogenic potential, due to the susceptibility of DNA to chemical mutagenesis. Nowadays, the correlation between mutagenic compounds in the Ames test and carcinogenicity seems to be >60 %.

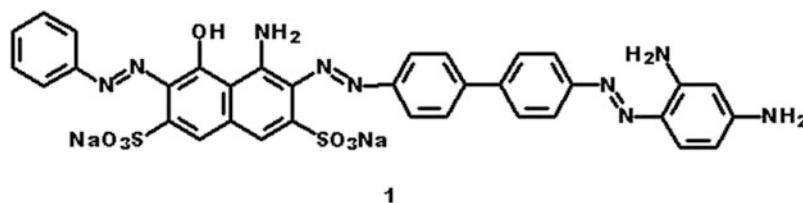
The Ames test is an in vitro method that commonly uses one or more strains of *Salmonella typhimurium*. There are six strains of *Salmonella typhimurium* that are widely used in a mutagenicity test and are designed to detect different types of mutations involving colorants. TA98 and TA1538 are sensitive to frameshift mutagens, TA100 and TA1535 are used to detect base pair substitution mutations, and TA97 and TA1537 are used for base pair substitution and some frameshift mutations, which sometimes cannot be detected with the former strains. These strains cannot produce the amino acid histidine, an essential component for growth. Thus, the bacteria are unable to multiply unless a mutagen causes the proper type of reverse mutation in the histidine gene. Mutagenic activity can be measured quantitatively by simply counting the number of colonies present after incubating *Salmonella* bacteria with several doses of the test compound and other necessary test additives for a standard length of time. The change in the bacteria that permits

growth is called a reverse mutation and the colonies that form are called revertant colonies. Generally, the test compound is considered mutagenic when the number of revertant colonies is more than twice that of the base count. Based on the number of the revertant colonies produced, the compound is characterized as a nonmutagen, a weak mutagen, or a strong mutagen. A solvent is added to facilitate adequate mixing, and an enzyme mixture may also be added to metabolize the test compound to enhance the sensitivity of the test, since bacteria do not have some of the enzymes present in mammalian tissues. The enzyme mixture is important because many compounds are mutagenic once they are metabolized in the liver and other tissues. An exogenous metabolic system, usually rat liver microsomal enzyme system, S9, is used in the *Salmonella* microsome assay.

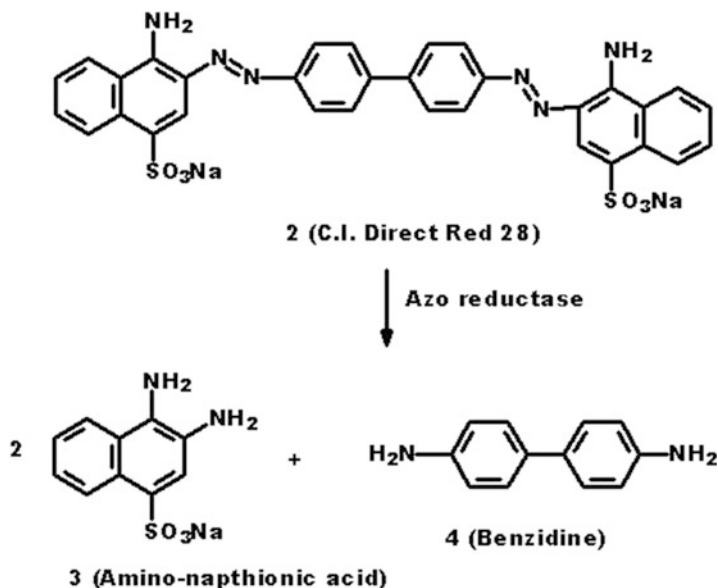
Benzidine-based dyes such as CI Direct Black 38 (Fig. 1) are carcinogenic in several mammalian species. However, certain benzidine-based dyes (e.g., 1; CI Direct Red 28; Congo Red) are not mutagenic in the standard *Salmonella* microsome test. Consequently, the standard test was modified in 1982 to insure the liberation of the parent diamine (cf., 2→4; Fig. 2) and the maximum possible mutagenic activity in each of the benzidine-based azo dyes studied [3]. This is an important protocol for benzidine-based dyes and is known as the preincubation assay under reductive conditions, a test that employs hamster liver S9 which is believed to be richer in reductase enzymes.

As a follow up to positive Ames/Prival outcomes, in vitro testing such as a gene mutation assay involving mammalian cells or the mouse lymphoma test, a chromosome aberration assay, has been recommended [4]. This approach complements the mutagenic evaluation of a compound

Colorant, Environmental Aspects, Fig. 1 Molecular structure of CI Direct Black 38 (1)



Colorant, Environmental Aspects, Fig. 2 Reductase enzyme cleavage of CI Direct Red 28 liberating carcinogenic benzidine (4)

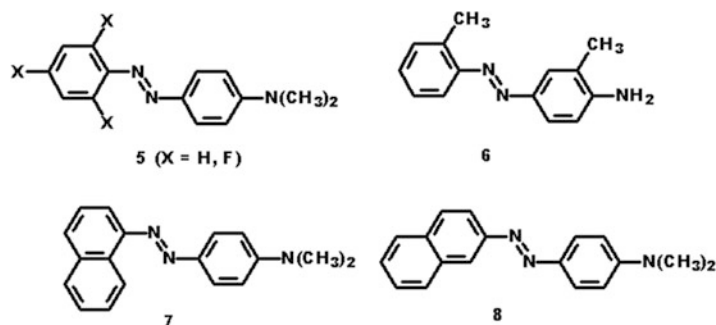
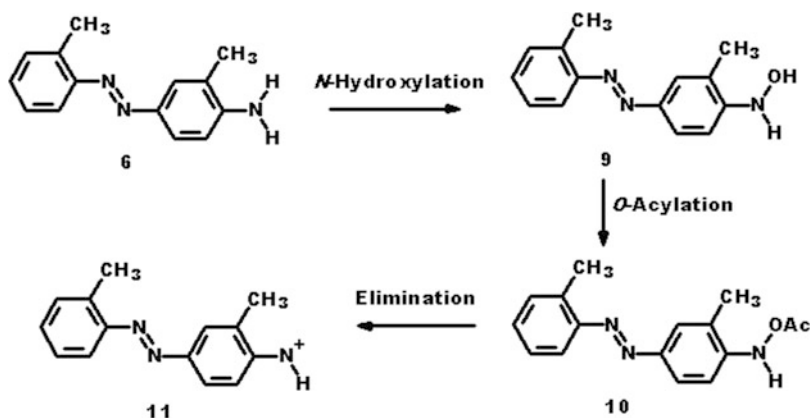
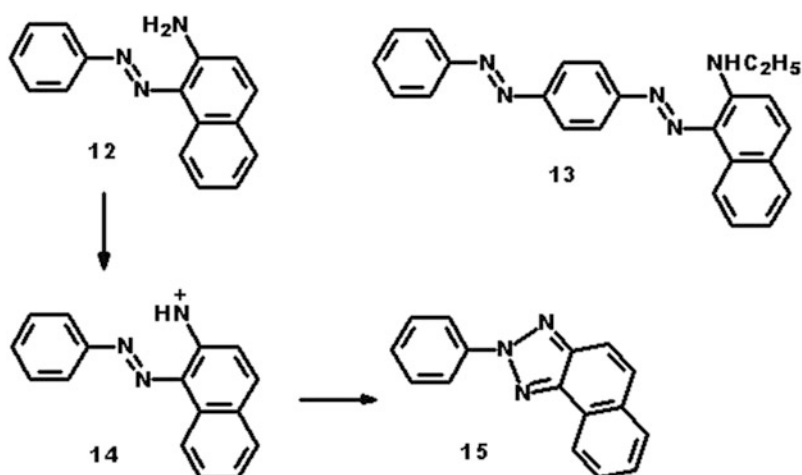


providing a more accurate prediction of its mutagenic properties and carcinogenic potential.

Carcinogenicity

Although a variety of synthetic colorants have been studied, the majority of our knowledge and major concerns in this area arises from the recognition of azo dye manufacturing involving benzidine and 2-naphthylamine as a source of bladder cancer in humans. This soon led to extensive evaluation of azo dyes and their precursors for carcinogenic potential. The volume of results from work in this area became so substantial and diverse that a concerted effort to digest it has been made [5]. A key goal was to determine (1) whether the production of tumors in animal studies was sufficient to designate a dye as a human carcinogen and (2) the reliability of the published literature. This inspired an assessment of published data on a group of 97 representative azo colorants associated with the dyestuffs industry, for the purpose of determining the scientific standing of the experimental work in this area and the potential human hazard associated with these compounds. This assessment contributed to a set of recommended guidelines for carcinogenicity testing that cover chemical purity of test compounds, number of animals tested and survival rates, study controls, dose levels, route of administration,

duration of experiments, pathological considerations, number of species, and evidence of human carcinogenicity. Application of the guidelines to published data afforded six categories of animal carcinogens and noncarcinogens: (1) Class A – the colorant was tested in at least two species of animals, resulting in the generation of nonconflicting data, using a sufficient number of animals (25 per group) and having a sufficient number of survivors for about two-thirds their expected lifetime. Repeated injections and urinary bladder implantations should not be regarded as meaningful in the context of chemically induced neoplasia. For a positive response, a target organ was identified and the induced tumors diagnosed as malignant. (2) Class B – the colorant was tested according to Class A guidelines but meaningful data was generated for one species only. (3) Class C – testing of the colorant involved an insufficient number of test animals and/or too many premature deaths occurred for results to be conclusive. For a positive result or trend, an increase in the number of animals with malignant tumors was observed but a target organ not identified. (4) Class D – the colorant was tested in multiple animal assays, which individually did not permit a conclusion but taken together provide sufficient evidence for an opinion. (5) Class E1 – the colorant was tested in a method considered inappropriate for

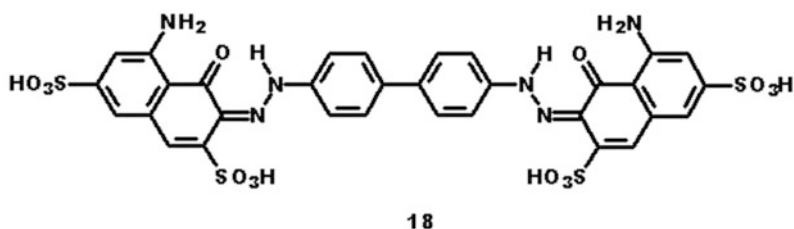
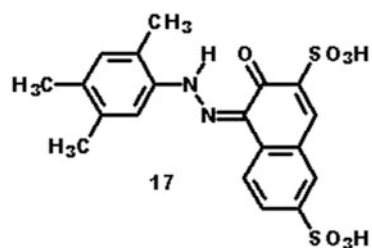
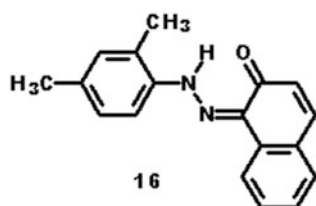
Colorant, Environmental Aspects,**Fig. 3** Carcinogenic 4-aminophenylazo disperse dyes**Colorant, Environmental Aspects,****Fig. 4** Metabolism of a 4-aminophenylazo dye leading to an electrophilic species**Colorant, Environmental Aspects,****Fig. 5** Noncarcinogenic 2-aminoazo dyes (12–13) and metabolism leading to triazole (15) formation

evaluation of chemical carcinogenicity, e.g., by bladder implantation or using too few animals. (6) Class E2 – there was insufficient data to permit a rational judgment regarding carcinogenicity.

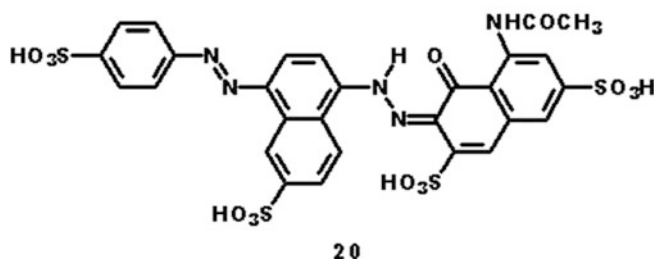
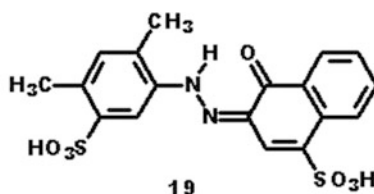
Structure-activity relationships associated with azo dye carcinogenicity produced the groupings

depicted in Figs. 3, 4, 5, 6, 7, and 8 [6]. Group 1 comprises carcinogenic azo dyes that are hydrophobic (lipophilic) colorants having a 4-aminophenylazo structure (5–8; Fig. 3) [7]. As illustrated in Fig. 4, the 4-amino group is susceptible to a series of metabolic steps leading to an

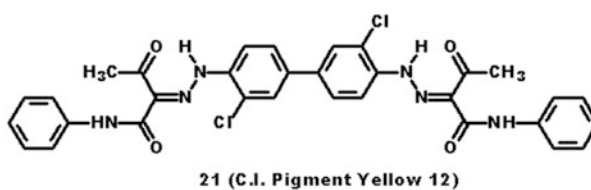
Colorant, Environmental Aspects, Fig. 6 Azo dyes liberating a carcinogenic aromatic amine upon reductive cleavage



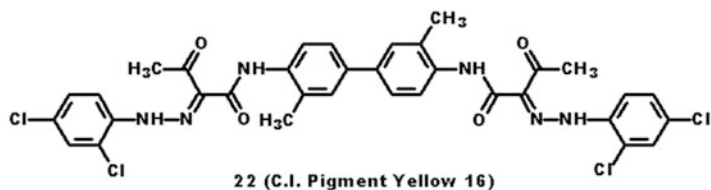
Colorant, Environmental Aspects, Fig. 7 Azo dyes liberating noncarcinogenic sulfonated amines upon reductive cleavage



Colorant, Environmental Aspects, Fig. 8 Examples of noncarcinogenic highly insoluble azo pigments



21 (C.I. Pigment Yellow 12)



22 (C.I. Pigment Yellow 16)

electrophilic nitrenium ion (**6**→**11**) that can react with DNA. This illustration is important because it serves as a reminder that azo bond cleavage is not an absolute requirement for genotoxicity. Group 2 examples in Fig. 5 show that placement of the amino group *ortho* to the azo bond rather than *para* affords a nitrenium ion having an internal mechanism for deactivation (cf., **12**→**15**). Group 3 comprises azo dyes producing a carcinogenic aromatic amine upon cleavage of the azo bonds that are designated as carcinogenic, examples of which are shown in Fig. 6. In these examples, 2,4-dimethylaniline, 2,3,4-trimethylaniline, and benzidine are produced by cleavage of the azo bonds in **16**–**18**. The properties of this dye grouping led to a ban on commercial products containing dyes derived from any of the family of 22 aromatic amines (now 24) known to be carcinogenic [8]. Similarly, dyes metabolized to benzidine are now classified by the IARC (International Agency for Research on Cancer) as belonging to Group 1 (carcinogenic to humans) [9, 10].

Group 4 azo dyes are metabolized to noncarcinogenic sulfonated aromatic amines and are regarded as safe to use (**19**–**20**). These structural types include FD&C Red 4 (**19**; Fig. 7), as well as FD&C Yellow 5 (Tartrazine, CI Food Yellow 4) and FD&C Yellow 6 (CI Food Yellow 3) [11]. Due to its benign nature, CI Food Black 2 was the first black dye used for ink-jet printing and remains the prototype for designing improved black dyes for this end-use area [12].

Insoluble organic pigments such as **21**–**22** characterize Group 5 colorants, which are noncarcinogenic (Fig. 8). They are not readily metabolized to their diamine precursors or absorbed into mammalian systems. The principal requirement for safety among organic pigments is the control of the concentration of unreacted (free) aromatic amine precursors in the commercial product [13].

Aquatic Toxicology

Aquatic toxicology is the study of the effects of chemicals such as colorants on aquatic organisms. The associated tests are used to measure the degree of response produced by exposure to

various concentrations of dyes and pigments and may be conducted in the laboratory or field. Laboratory tests encompass four general methods: (1) static test (organisms are placed in a test chamber of static solution), (2) recirculation test (organisms are placed in a circulated test solution), (3) renewal test (organism is placed in a static test solution that is changed periodically), and (4) flow-through test (organisms are placed in a continuously flowing fresh test solution).

In the USA, the Clean Water Act (cf., Title 40 of the US Code of Federation Regulations 100–140, 400–470), which is administered by the Environmental Protection Agency (EPA), limits water pollution from industrial and public sources, stresses the importance of controlling toxic pollutants, and encourages investigations leading to waste-treatment technologies [13]. Similarly, the Office of Pollution Prevention and Toxics (OPPT) implements the Toxic Substances Control Act (TSCA) as part of its responsibility for evaluating and assessing the impact of new chemicals and chemicals with new uses to determine their potential risk to human health and the environment. Testing recommended includes acute and chronic toxicity for assessing environmental effects and bioconcentration, biodegradation, physicochemical properties, and transport/transformation studies for assessing environmental fate [14, 15]. Ecosystem-related toxicological testing includes acute and chronic toxicity tests employing species such as fish, *Daphnia*, earthworms, and green algae. The overall goal is to determine the potential for changes in the composition of plant or animal life, abnormal number of deaths of organisms (e.g., fish kills), reduction of reproductive success or the robustness of a species, alterations in the behavior of distribution of a species, and long-lasting or irreversible contamination of components of the physical environment (e.g., surface water). Specifically, acute toxicity tests are short-term tests designed to measure the effects of toxic agents on aquatic species during a short period of their life span. Acute toxicity tests evaluate effects on survival over a 24- to 96-h period. Chronic toxicity tests are designed to measure the effects of toxicants to aquatic species over a significant portion of the organism's life cycle,

typically one-tenth or more of the organism's lifetime. Chronic studies evaluate the sublethal effects of toxicants on reproduction, growth, and behavior due to physiological and biochemical disruptions.

The most common static tests are performed with daphnids, mysid shrimp, amphipods, and fish (e.g., fathead minnow, zebrafish, and rainbow trout). Renewal tests are often used for life-cycle studies using *Ceriodaphnia dubia* (7 days) and *Daphnia magna* (21 days). Longer-term studies are usually performed with these tests. Flow-through tests are generally considered superior to static tests, as they are very efficient at retaining a higher-level of water quality, ensuring the health of the test organisms. Flow-through tests usually eliminate concerns related to ammonia buildup and dissolved oxygen usage as well as ensure that the toxicant concentration remains constant.

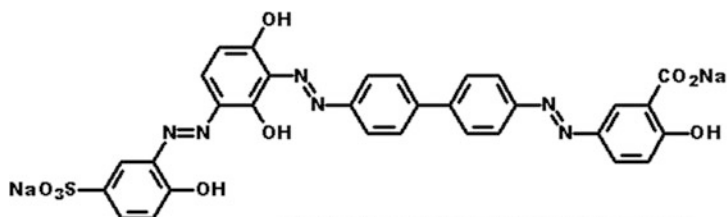
TL₅₀ values (aqueous concentrations at which 50 % of the test organisms survive after a 96-h exposure) from a static bioassay using fathead minnow and a diverse group of commercial synthetic dyes were measured [16]. The most toxic dyes were Basic Violet 1 (Methyl Violet), with a TL₅₀ of 0.047 mg/L, and Basic Green 4 (Malachite Green) at 0.12 mg/L, Disperse Blue 3 at 1 mg/L, Basic Yellow 71 at 3.2 mg/L, Basic Blue 3 at 4 mg/L, Acid Blue 113 at 4 mg/L, Basic Brown 4 at 5.6 mg/L, Mordant Black 11 at 6 mg/L, Acid Green 25 at 6.2 mg/L, and Acid Black 52 at 7 mg/L. These commercial products included 29 of the 47 dyes with TL₅₀ >180 mg/L, many that were bisazo or trisazo dyes. Three dyes had TL₅₀ values between 100 and 180 mg/L, and the remaining 14 had values <100 mg/L. These results are consistent with the frequent lack of correlation between dye structures leading to genotoxicity and those leading to aquatic toxicity. Whereas benzidine-based dyes such as CI Direct Black 38 (**1**) and Direct Blue 6 (**18**) are designated as carcinogenic, they did not exhibit acute toxicity when tested with aquatic organisms, along with their counterparts **23–24** – both of which are Cu complexes (Fig. 9). On the other hand, all of the cationic (basic) dyes and anthraquinone disperse and acid dyes (**25–31**; Fig. 10) were significantly, if not extremely, toxic in this assay. Interestingly,

bisazo dye Acid Blue 113, unlike the bisazo direct dyes, and Cr-complexed dye Acid Black 52 and its unmetallized precursor, Mordant Black 11 (**32–34**; Fig. 11), were also significantly toxic. - Water-insoluble vat dyes (**35–37**; Fig. 12) did not exhibit acute toxicity in this assay. In a study involving the sparingly water-soluble monoazo dye Disperse Red 1, aquatic toxicity was observed in a variety of animals, with *Daphnia similis* the most sensitive species in acute tests and *Pseudokirchneriella subcapitata* the most sensitive species in chronic tests [17]. Clearly, more studies are needed in this area in order to establish clear correlations between the molecular structures of colorants and their aquatic toxicity. While the picture is somewhat clearer regarding cationic dyes and vat dyes, it is much less so regarding sulfonated azo dyes.

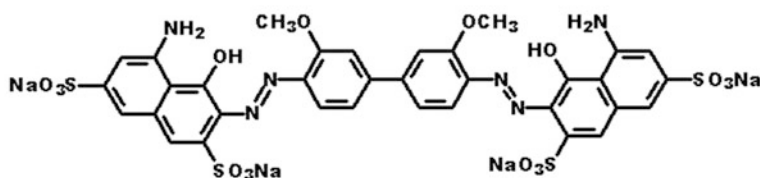
Bioconcentration studies are performed to evaluate the potential for a chemical to accumulate in aquatic organisms that may subsequently be consumed by higher trophic-level organisms including humans. The extent to which a chemical is concentrated in tissue above the level in water is referred to as the bioconcentration factor (BCF). It is widely accepted that the octanol/water partition coefficient (e.g., Log K_{ow} or Log *P*) can be used to estimate the potential for nonionizable organic chemicals to bioconcentrate in aquatic organisms.

With regard to water quality concerns, a variety of methods have been developed to treat wastewater prior to industrial discharges, in order to remove colorants that are clearly visible at very low levels (e.g., 1 ppm). The methods employed involve chemical and biological agents for reduction and oxidation, physical adsorption, membrane filtration, precipitation, and recycle and reuse. The latter methods are especially important since dyes are engineered to be stable compounds under end-use conditions and thus are completely decomposed with considerable difficulty. Also, care must be used when choosing a wastewater decoloration method. In several countries, effluents from biological treatment are decolorized using chlorine for simultaneous pathogen removal. It has been found, however, that the chlorination of wastewater containing certain disperse dyes can generate mutagenic compounds

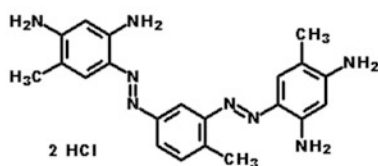
Colorant, Environmental Aspects, Fig. 9 Examples of benzidine and benzidine congener-based dyes found negative in aquatic toxicity testing



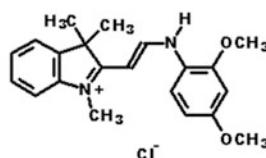
23 (C.I. Direct Brown 95; Cu Complex)



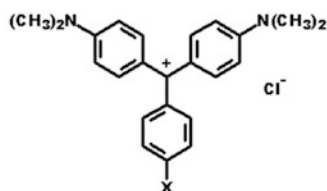
24 (C.I. Direct Blue 218; Cu Complex)



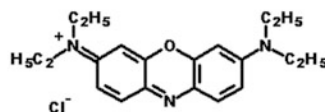
25 (C.I. Basic Brown 4)



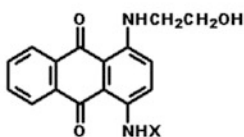
26 (C.I. Basic Yellow 11)



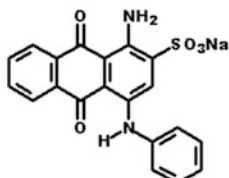
27 (X = H, N(CH₃)₂)



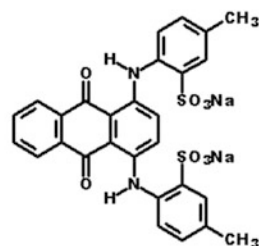
28 (C.I. Basic Blue 3)



29 (X = CH₃, CH₂CH₂OH)



30 (C.I. Acid Blue 25)

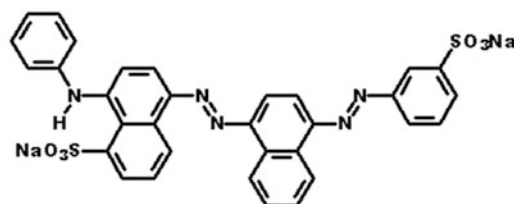


31 (C.I. Acid Green 25)

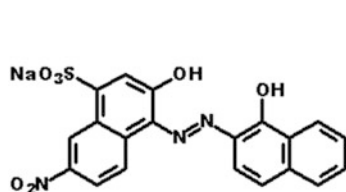
Colorant, Environmental Aspects, Fig. 10 Examples of cationic, disperse, and acid dyes exhibiting aquatic toxicity

Colorant, Environmental Aspects,

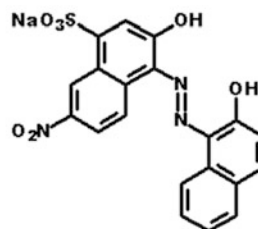
Fig. 11 Examples of bisazo and metal-complexed acid dyes exhibiting aquatic toxicity



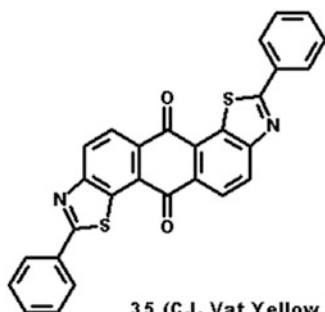
32 (C.I. Acid Blue 113)



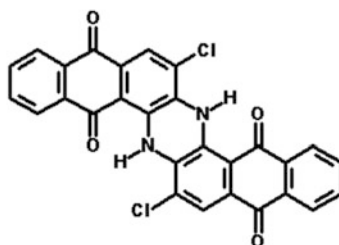
33 (C.I. Mordant Black 11)



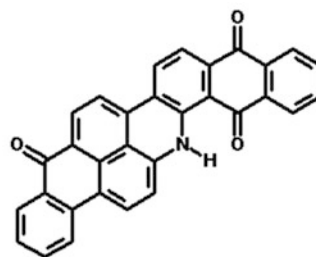
34 (C.I. Acid Black 52 - 2:3 Complex)



35 (C.I. Vat Yellow 2)



36 (C.I. Vat Blue 6)



37 (C.I. Vat Green 3)

Colorant, Environmental Aspects, Fig. 12 Water-insoluble vat dyes found negative in aquatic toxicity testing

known as phenyl benzotriazoles (PBTAs; **38**). These chlorinated compounds have been observed in drinking water. Similarly, the chlorination of commercial Disperse Red 1 (**39**; Fig. 13) produced higher mutagenicity than the untreated dye.

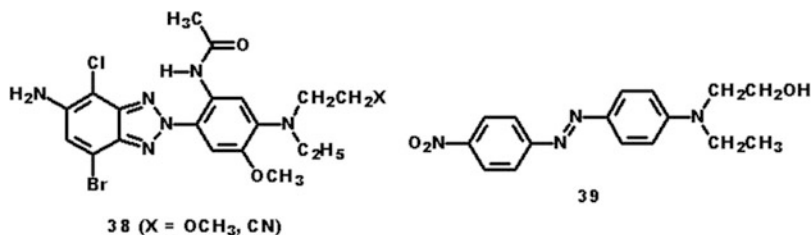
Future Directions

Bearing in mind that organic colorants include naturally occurring and synthetic compounds and that the major concern generally lies with

the latter types, there is renewed interest in exploring natural dye-based coloration because such dyes are widely regarded as biodegradable and of low inherent toxicity. In fact, various natural dyes are part of a normal diet (e.g., beta-carotene and lycopene, both of which are important antioxidants). Although few natural colorants have shown commercial viability (e.g., indigo, madder, logwood), they continue to find success in niche areas such as arts and crafts and enjoy exploration for potential use in mainstream products. It is worthwhile pointing out that the use of natural dyes as colorants for textiles often introduces the

Colorant, Environmental Aspects,

Fig. 13 Structures of compounds 38–39



need for additives known as mordants, to help bond these dyes to fibers, since natural dyes have very low affinity for textiles. Mordants such as Cu and Cr are toxic to aquatic life and pose human health risks, making Al and Fe better choices. Regarding synthetic colorants, which encompass water- and/or solvent-soluble organic dyes and insoluble organic and inorganic pigments, they will continue to be the primary agents for adding color to substrates such as natural and synthetic fibers, paper, plastics, petroleum products, printing inks, hair, and food, drug, and cosmetic products. Due to their practically insoluble nature, pigments have very low bioavailability; hence much of the attention in the area of environmental concerns has fallen on synthetic dyes. Among the synthetic dyes, azo dyes lie at the forefront in this area, and the design of such dyes will continue to take into consideration the environmental properties of the required precursors and the metabolites formed in mammalian systems.

Cross-References

- [Colorant, Environmental Aspects](#)
- [Colorant, Textile](#)
- [Dye](#)

References

1. Zollinger, H.: *Color Chemistry: Synthesis, Properties, and Applications of Organic Dyes and Pigments*, "Analysis, Ecology, and Toxicology of Colorants". Wiley-VCH, Switzerland (2003)
2. Maron, D.M., Ames, B.N.: Revised methods for the *Salmonella* mutagenicity test. *Mutat. Res.* **113**, 173 (1983)
3. Prival, M.J., Mitchell, V.D.: Analysis of a method for testing azo dyes for mutagenic activity in *Salmonella typhimurium* in the presence of flavin mononucleotide and hamster. *Mutat. Res.* **97**, 103 (1982)
4. Hunger, K. (ed.): *Industrial Dyes: Chemistry, Properties, and Applications*, "Health and Safety Aspects". Wiley-VCH, Germany (2003)
5. Longstaff, E.: An assessment and categorisation of the animal carcinogenicity data on selected dyestuffs and an extrapolation of those data to an evaluation of the relative carcinogenic risk to man. *Dyes Pigments* **4**, 243–304 (1983)
6. Gregory, P.: Azo dyes: structure-carcinogenicity relationships. *Dyes Pigments* **7**, 45–56 (1986)
7. Weisburger, E.K.: Cancer-causing chemicals. In: LaFond, R.E. (ed.) *Cancer-The Outlaw Cell*. American Chemical Society, Washington, DC (1978)
8. German ban on use of certain azo compounds in some consumer goods. *Text. Chem. Color.* **28**(4), 11–13 (1996)
9. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Some Aromatic Amines, Organic Dyes, and Related Exposures, vol. 99, 692 pp (2010)
10. Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Coglian, V.: Some aromatic amines, organic dyes, and related exposures. *Lancet Oncol* **9**(4), 322–323 (2008)
11. Marmion, D.M.: *Handbook of U.S. Colorants: Foods, Drugs, Cosmetics, and Medical Devices*. Wiley, New York (1991)
12. Carr, K.: Dyes for ink jet printing. In: Freeman, H.S., Peters, A.T. (eds.) *Colorants for Non-textile Applications*. Elsevier Science BV, Amsterdam (2000)
13. Code of Federal Regulations, Title 40, Protection of Environment, US EPA, amended (1972)
14. Fogle, H.C.: Regulatory toxicology: U.S. EPA/chemicals TSCA. In: M.J. Derelanko, M.-A. Hollinger (eds.) *Handbook of Toxicology*, CRC Press, Boca Raton (1995)
15. Adams, W.J., Rowland, C.D.: Aquatic toxicology test methods. In: Hoffman, D.J., Rattner, B.A., Burton Jr., G.A., Cairns Jr., J. (eds.) *Handbook of Ecotoxicology*. Lewis Publishers (CRC Press), Boca Raton (2003)
16. Little, L.W., Lamb III, J.C.: Acute Toxicity of 46 Selected Dyes to the Fathead Minnow, *Pimephales promelas*. The University of North Carolina Wastewater Research Center, Department of Environmental Sciences and Engineering, Chapel Hill (1972)
17. Vacchi, F.I., Ribeiro, A.R., Umbuzeiro, G. A.: Ecotoxicity Evaluation of CI Disperse Red 1 Dye, SETAC North America 31st Annual Meeting. Portland, 7–11 Nov 2010

Colorant, Halochromic

► [Colorant, Thermochromic](#)

Colorant, Natural

Thomas Bechtold

Research Institute for Textile Chemistry and
Textile Physics, Leopold Franzens University of
Innsbruck, Dornbirn, Vorarlberg, Austria

Synonyms

[Natural colorants](#); [Natural dyes](#); [Pigments](#)

Definition

The term natural colorant comprises all kind of materials available from natural sources which are able to impart color to matter.

The main principle of color development for natural colorants is the specific absorption of light in the wavelength region of 400–700 nm [1].

Other principles on color formation may be based upon physical effects, for example, refraction of light (rainbows), interference (feathers of peacocks), or electron excitation (electroluminescence) [2].

Sources for natural colorants include minerals (red ochre, $\alpha\text{-Fe}_2\text{O}_3$), plant material (e.g., flavonoids from Canadian golden rod), and animal-based dyes (e.g., indigoid colorants from selected mollusc species).

Colored pigments are applied as finely divided solid particles which remain in an insoluble state during their application and use. Water-soluble or oil-soluble natural colorants can be found in food and beverage applications. A colorant that is used as dye must exhibit a specific substantivity towards that substrate. A dye is thus able to be sorbed at the surface of the material.

According to structure, organic natural colorants can be divided into several main groups,

namely, flavonoid dyes (including anthocyanins), substances containing naphthoquinoid and anthraquinoid groups, indigoid dyes, tannins, carotenoids, and chlorophylls [3].

The colorant content in natural sources is relatively low, usually of the order of a few % of the dry material. Thus, considerable amounts of raw material have to be processed.

Natural colorants are usually obtained by means of aqueous extraction of plant material; the use of solvents is less common as substantial amounts of plant material have to be extracted. As a result of the nonselective extraction process, natural colorants usually contain a number of different color-providing chemical substances. The classification of a dye plant or animal source to a certain group of colorant usually is based on the properties of the most important species of dye present in the extract.

Major applications of natural colorants are as the coloring elements in food, as dyes for traditional textile dyeing eco-textiles, and as pigments for cosmetics [4].

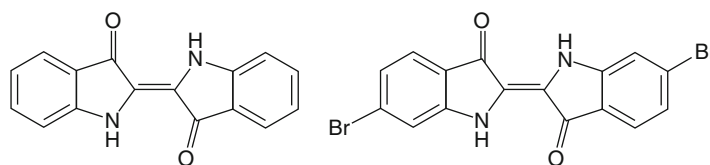
Natural colorants can be applied in different ways. In many cases the extracted dye can be used directly (e.g., food and selected textile applications), but often mordants are used to increase dye adsorption (e.g., for textile dyeing and dye-lake formation). Important mordanting substances are tannins or metal salts.

Major Classes of Natural Colorants

As mentioned, as a result of the extraction of natural colorants from native sources, with few exceptions, a number of colored substances are present in the extract, and, therefore, natural colorants are classified on the basis of the main constituent, which is of relevance for a given application. Important dyes with known chemical structure have been assigned Colour Index (C.I.) Generic Names and/or C.I. Constitution Numbers. The major classes of colorant are discussed below.

Indigoid Dyes

The two most important natural dyes of indigoid structure are indigo (C.I. Natural Blue 1) obtained from plant sources and Tyrian purple

Colorant, Natural,**Fig. 1** Structure of indigo (left) and Tyrian purple (right)**Colorant, Natural, Table 1** Indigo plants

Plant	Botanical name	Climate	Region	Precursor
Indigo plant	<i>Indigofera tinctoria</i> L.	Tropical	India, Africa, and North, Central, and South America	Indican indigo- β -D-glucoside
Woad	<i>Isatis tinctoria</i> <i>Isatis indigotica</i>	Temperate	Europe China	Isatan B indoxyl-5-ketogluconat
Dyer's knotweed, Ai	<i>Polygonum tinctorium</i> AIT	Subtropical and temperate	Europe, Japan	Indican indigo- β -D-glucoside

(6,6'-dibromoindigo, C.I. 75800) from the Muricidae family, e.g., spiny dye murex (*Bolinus brandaris*) or banded dye murex (*Hexaplex trunculus*).

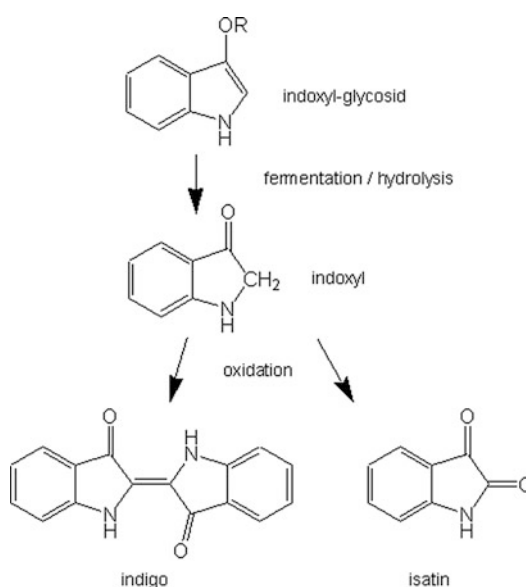
The structures of these two dyes are given in Fig. 1.

Tyrian purple is found in extracts from the hypobranchial glands of the gastropod mollusc. While the applicatory properties and fastness of the colorant were generally acceptable, the use of this dye ceased due to the low dye content in the mollusc.

Indigo is the most relevant blue natural dye and different plants have been cultivated for indigo production all over the world. As a general rule, the indigo is present in the plant in the form of different colorless indoxyl glycosides as precursors. Selected plants for indigo production are summarized in Table 1.

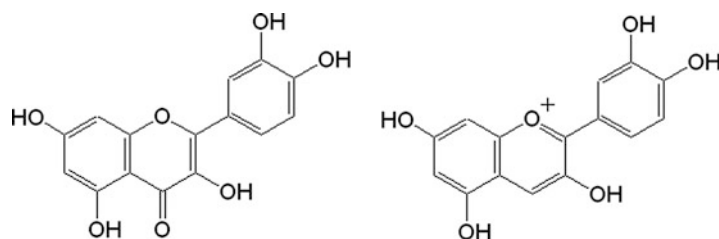
To obtain indigo the precursor is first hydrolyzed enzymatically to obtain the intermediate indoxyl, which then rapidly oxidizes in the presence of air oxygen to indigo. The precipitated solid indigo dye then can be collected. The oxidation of indoxyl to yield the yellow isatin is an unwanted side reaction, which can cause substantial losses in indigo yield. A general scheme is presented in Fig. 2.

Besides the use as insoluble indigo colorant, indigoid dyes usually are applied as vat dyes. In vat dyeing the insoluble dye (indigo) is firstly converted to the alkali soluble leuco form, which

**Colorant, Natural, Fig. 2** Main steps in production of natural indigo (R = glycosidic bond carbohydrate)

exhibits substantial substantivity towards protein fibers (wool, silk) and cellulosic fibers (cotton, flax, regenerated cellulose fibers). The adsorbed reduced indigo dye (yellow) is then oxidized to generate the blue-colored insoluble form of the colorant using air or chemical oxidants. For deep shades, the procedure of immersion into the leuco dyebath and subsequent oxidation in air is repeated several times.

Dye reduction can be achieved by use of reductants and alkali, such as sodium dithionite

Colorant, Natural,**Fig. 3** Basic structures of quercetin (flavonoid, *left*) and cyanidin (anthocyanidin, *right*)**Colorant, Natural, Table 2** Plant sources for flavonoids

Plant	Botanical name	Main colorants		Part of plant
Weld	<i>Reseda luteola</i>	Luteolin	C.I. Natural Yellow 2	Plant except roots
		Apigenin		
Roman chamomile	<i>Chamaemelum nobile</i>	Apigenin	C.I. Natural Yellow 1	Flower
Onion	<i>Allium cepa</i>	Quercetin		Outer shell of fruit
Black oak	<i>Quercus velutina</i>	Quercetin	C.I. Natural Yellow 10	Bark

($\text{Na}_2\text{S}_2\text{O}_4$) and NaOH ; alternatively, anaerobic microbial reduction can be employed. Microbiological dye reduction is practiced in traditional handicraft dyeing, but up to now is not applicable for dyeing with indigo in denim production (jeans).

Flavonoids and Anthocyanins

Flavonoids contribute to the color of many food products and thus are consumed by humans in the form of fruits and food. The color gamut provided by these colorants ranges from the yellow flavonoids (e.g., flavonols, chalcones, aurones) to orange–red–purple anthocyanins. In plants many flavonoids are present as glycosides; anthocyanins are also glycosides, while the respective aglycones are named anthocyanidins. Representative structures of flavonoids and anthocyanidins are given in Fig. 3.

Yellow flavonoid dyes can be extracted from a high number of plant sources; selected representatives are given in Table 2 [5].

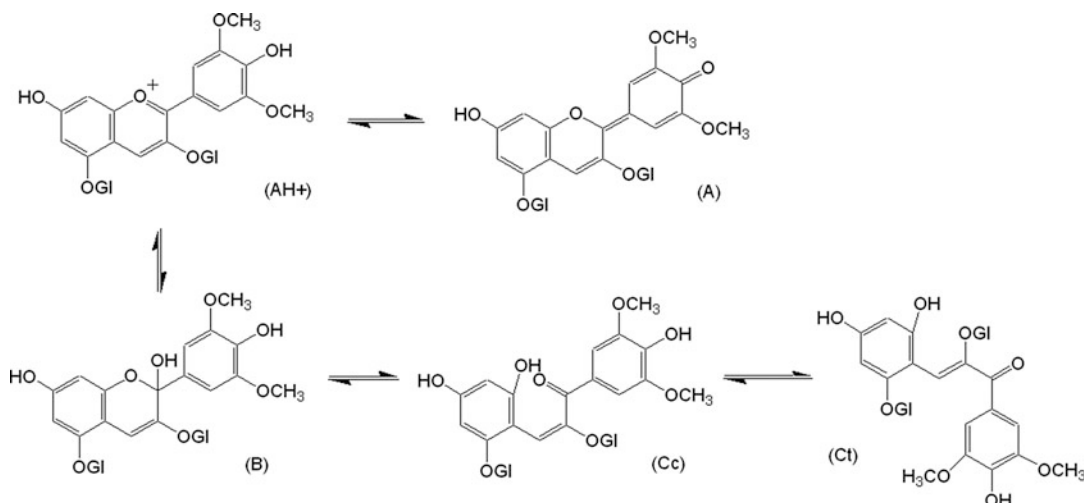
For textile dyeing, flavonoids can be used as direct dyes or as mordant dyes in combination with metallic mordants such as Fe or Al salts. When Fe, Sn, or Cu salts are employed as mordants, metal complexes are formed and a distinct change in color is observed, e.g., olive shades are obtained when an onion extract is applied in combination with Fe mordant.

The range of colors of anthocyanin dyes is of high coloristic interest for application as food colorants and for general purposes of coloration. The limited stability of the molecules can be improved by co-pigmentation, metal complexation, and the use of additives. Dependent on solution pH, anthocyanins rearrange to form different species with different absorption spectra. Some of these products are colorless which explains the propensity of the colorants to fade.

A general scheme describing the pH-dependent rearrangement of an anthocyanin structure is given in Fig. 4.

At pH 1.0, orange–red/violet oxonium/flavylium ions are present which are of greatest stability (Fig. 4, AH^+). With increase in pH to between 4 and 6, the colored quinoid base appears (Fig. 4a). At the same time the formation of the colorless hemiketal (Fig. 4b) gains importance. The hemiketal B is sensitive to tautomerization and ring opening, thereby leading to colorless cis- and trans-chalcone forms (Fig. 4, Cc and Ct) [6].

Stabilization of the dye molecules can be achieved by co-pigmentation. By inter- and intramolecular complex formations, self-association, and metal complex formation, the stability of the dyes can be increased; in many cases the color becomes bluer (bathochromic shift) and absorbance is increased (hyperchromic shift).

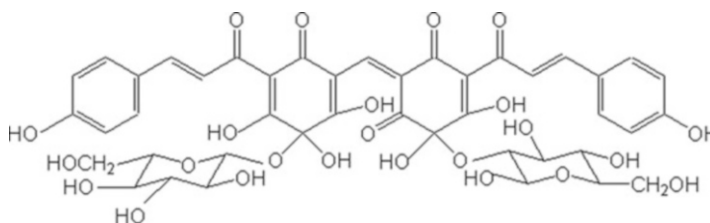


Colorant, Natural, Fig. 4 pH-dependent structural transformation of anthocyanins (malvidin 3,5,-diglycoside)

Colorant, Natural, Table 3 Plant sources for anthocyanins

Plant	Botanical name	Main colorants		Part of plant
Elder	<i>Sambucus nigra</i>	Cyanidin glycosides		Fruit
Vine	<i>Vitis vinifera</i>	Malvidin glycosides		Fruit
Privet	<i>Ligustrum vulgare</i>	Malvidin, cyanidin, delphinidin glycosides	C.I. Natural Black 5	Berries
Hollyhock	<i>Alcea rosea</i>	Malvidin, delphinidin glycosides		Flowers

Colorant, Natural, Fig. 5 Structure of carthamin



Anthocyanins are found in many intensively colored fruits and berries. Examples are summarized in Table 3.

Quinoid, Naphthoquinoid, and Anthraquinoid Dyes

Important yellow, orange, and red dyes belong to this group of colorants.

The red–orange quinoid dye carthamin (C.I. Natural Red 26) can be obtained by extraction of blooms of Safflower (*Carthamus tinctorius*) (Fig. 5).

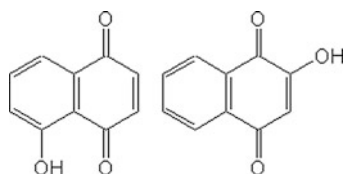
Two important representatives of naphthoquinone dyes are lawson (2-hydroxy-1,4,-naphthoquinone) and juglon (5-hydroxy-1,4,-naphthoquinone) (Fig. 6).

Lawson (C.I. Natural Orange 6) is extracted from henna (*Lawsonia inermis*). The plant grows best in tropical savannah and arid zones. The leaves contain lawson in the form of its glycoside, which hydrolyses and releases the aglycone, which then is oxidized to the quinone form. Lawson is of significant interest in hair dyeing. As a small molecule, its diffusion rate is

sufficiently high to impart acceptable levels of color under the gentle conditions employed for hair dyeing. When henna leaves are mixed with leaves of the indigo plant, the so-called black henna is obtained. In the application as hair dye, lawson and indigo are fixed on the hair and brown-black shades are obtained.

Anthraquinone dyes can be extracted from plant sources (madder, *Rubia tinctorum*, C.I. Natural Red 8; *Hedge bedstraw*, *Gallium mollugo*, Natural Red 14) [7] or from animals (Kermes, *Kermes vermilio*, C.I. Natural Red 3; cochineal, *Dactylopius coccus*, C.I. Natural Red 4). Anthraquinoid dyes also have been extracted from fungi and lichen [8].

Most plant-based anthraquinoid dyes are extracted from the roots of the plants. The colorants are present as glycosides which hydrolyze during storage or extraction to the corresponding anthraquinoid dye. Representative structures (Alizarin, 1,2-dihydroxyanthraquinone,



Colorant, Natural, Fig. 6 Structures of lawson (*left*) and juglon (*right*)

pseudopurpurin

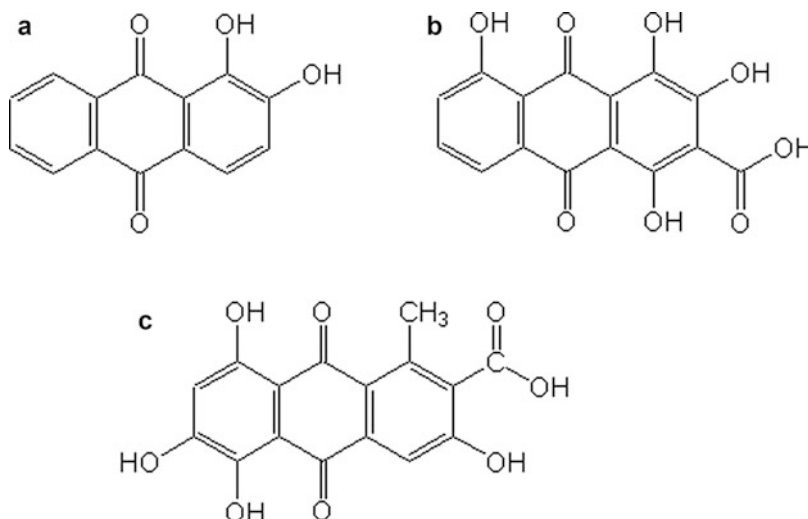
3-carboxy-1,2,4-trihydroxyanthraquinone) are shown in Fig. 7. Through the high number of hydroxyl groups that neighbor the quinoid system, these compounds are able to form stable metal complexes with, for example, aluminum or calcium. Thus, in traditional textile dyeing with madder extracts, mordanting with metal salts is used to improve dye fixation and fastness. The high fastness properties of such dyeings makes them valuable for textile applications, although care has to be exercised to ensure that harmful components such as lucidin (1,3-dihydroxy-2-hydroxymethylanthraquinone) have been removed, e.g., through an oxidation step.

Kermes, cochénille, and lac are important examples of insect-based natural colorants. In the cases of kermes and cochénille, the parasitic insects live on host plants from where the female insects are collected and dried. Lac is obtained from the secretions of the lac insect. Due to the high quality (light fastness and wash fastness) of red textile dyeings, these dyes were of high importance in traditional dyeing. The dyes still are produced in considerable amounts for cosmetic applications and as food colorants (Table 4).

Tannin-Based Dyes

Tannins are a complex group of polyphenolic compounds. Gallotannins comprise the structural

Colorant, Natural, Fig. 7 Structures of alizarin (a), pseudopurpurin (b), and kermesic acid (c)

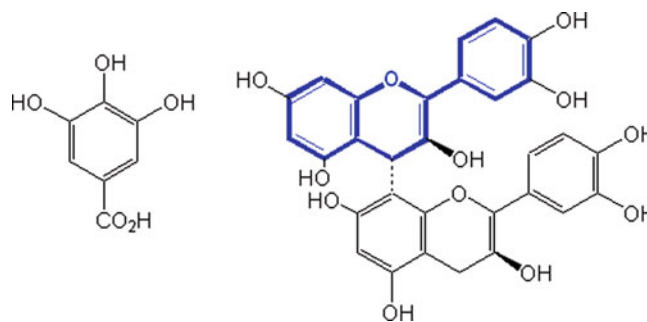


Colorant, Natural, Table 4 Insect dyes

Insect	Name	Main colorants		Host plant
Kermes	<i>Kermes vermilio</i>	Kermesic acid	C.I. Natural Red 3	Kermes oak
Cochenille	<i>Dactylopius coccus</i>	Carminic acid	C.I. Natural Red 4	Nopal cactus
Lac insect	<i>Kerria lacca</i>	Laccaic acids	C.I. Natural Red 25	Not specific

Colorant, Natural,

Fig. 8 Structure of gallic acid and a condensed tannin ((+)-catechin-(+)-catechin, B3) flavan structure marked blue)

**Colorant, Natural, Table 5** Plant sources for gallotannins and tannin agents (condensed tannins)

Plant	Name		Main colorants	Content/%	
Aleppo gall on oak tree	<i>Quercus infectoria</i>	Gallnut	Turkey tannin	50–70	
Sicilian sumac	<i>Rhus coriaria</i>	Leaves, twigs	Gallotannin	23–35	C.I. Natural Brown 6
Sticky alder tree	<i>Alnus glutinosa</i>	Bark	Gallotannin	20	
Pomegranate	<i>Punica granatum</i>	Pomegranate fruit bark	Gallotannin	28	C.I. Natural Yellow 7
Scots pine	<i>Pinus sylvestris</i>	bark	Tannin agent	17	
Cutch/catechu	—	Bark, leaves	Tannin agent	—	C.I. Natural Brown 3
Tea	<i>Camellia sinensis</i>	Leaves	Tannins agent	25	

unit of gallic acid esterified with sugar molecules (hydrolysable tannins), while tannins contain flavan as general structural element, by condensation complex polyphenolic structures are formed (condensed tannins) (Fig. 8).

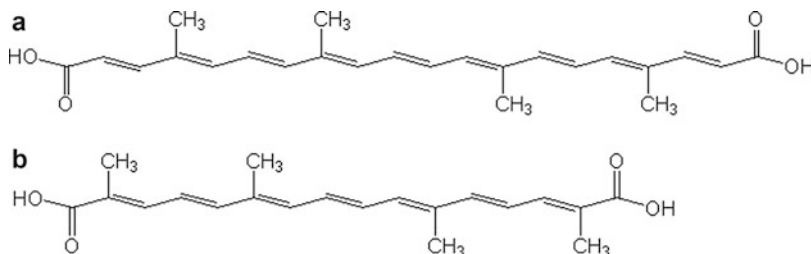
Tannins are found in all parts of plant, e.g., bark, wood, leaves, floral parts, and gallnut. In contrast to other natural colorants, the tannin content in plant material is much higher and, depending on source, can reach levels of 50 % of the plant mass (Table 5).

A gallnut is formed as response to the egg deposition by a gall wasp, e.g., on a leaves and twigs of an oak tree.

The term catechu is used for a number of plant extracts with high content in tanning agents, obtained from different plants, e.g., mangroves, conifers, cutch, and acacia.

Tannins and tannin agents form intensively colored complexes with metals such iron and copper; thus, color development usually is coupled to use of metal mordants.

Colorant, Natural,
Fig. 9 Structure of bixin
 (a) and crocetin (b)



Tannins also are used as mordants in place of metal mordants to increase the extent of adsorption and fastness of other natural colorants.

Carotenoid Dyes

A number of yellow, orange, or red pigments are found in many plants and animals. Animals are not able to synthesize carotenoids, which they have to obtain from their food.

Carotenoids are polyisoprenoids; hydrocarbon carotenoids are carotenes, while oxygen-containing molecules are named xanthophylls. Characteristically, carotenoids comprise conjugated double bonds. Typical examples are α -carotene (carrots, *Daucus carota*; red palm oil), lutein (green leafy vegetables, broccoli, corn), bixin (annatto seeds, *Bixa orellana*, C.I. Natural Orange 4), capsanthin (paprika seed, *Capsicum annum*), lycopene (tomato, *Lycopersicon esculentum*), and crocetin (saffron, *Crocus sativus*, C.I. Natural Yellow 6) (Fig. 9).

From annatto seeds and saffron flowers, orange dyes can be obtained by aqueous extraction. The water-soluble colorants can be used to dye wool, silk, and cellulose fibers. Both dyes also are used as food colorants.

A major group of other carotenoids exhibit poor solubility in water, and extracts are obtained using solvents of low polarity (hexane, oil). As a result, their use as food colorants is dependent on the oil content of the product as in many cases the oily phase will contain the dye.

Chlorophylls

While chlorophylls represent the most abundant pigments in nature, their use as colorants is limited due to low chemical instability and high production costs. Different natural chlorophylls have

been identified (chlorophyll *a*, *b*, *c*, *d*, and *e*). While the extraction of the green dye from plant leaves, algae, appears of high interest for the coloration of food, cosmetics, and textiles, difficulties in obtaining pure products and the rapid modification via endogenous plant enzymes to brown-green products prevent the simple, direct use of chlorophyll as a colorant.

Chlorophyllines are semisynthetic, metal-chelated chlorophyll derivatives which exhibit higher stability and are water soluble in many cases. After hydrolytic treatments and replacement of the central magnesium by copper, sodium copper chlorophyllin can be obtained. In Europe the green complex can be used as colorant for sweets, ice cream, and cheese. In the United States the use of copper chlorophyllin is more restricted, a possible use being in dentifrice.

The use of copper chlorophyllin for the coloration of paper, textiles, and leather is under consideration, although the Cu content in the product has to be considered.

Applicatory Aspects

As many natural colorants are readily soluble in water, this is the main solvent used to extract the dye from a source. Due to the low content of colored material in the natural source, the amounts of extracts are limited. Use of solvent extraction such as ethanol or hydrocarbons is limited to special purposes as considerable amounts of treated material, e.g., plant residues have to be deposited.

Concentration by solvent evaporation has to be considered carefully for energy reasons. Another technique to obtain concentrated dye preparations is precipitation as a dye lake by formation of insoluble complexes, e.g., by addition of calcium, iron, or aluminum salts.

The dye lakes can be used directly as pigment dyes for paints and cosmetics.

For textile dyeing purposes, the extracted natural colorant can be used either directly without recourse to mordants. Attachment of the colorant to the substrate is then based on H-bonding, dipole interactions, and van der Waals' forces.

Mordants are auxiliary chemicals used to increase dye adsorption and fixation on the substrate. The mordant can be applied in a separate bath (pre-mordanting), added directly to the dye containing bath (meta-mordanting), or used as an aftertreatment (post-mordanting).

While in traditional dyeing with natural colorants, many different heavy metal salts were used (e.g., copper, tin, chromium), environmental legislation has restricted the use of heavy metals in textile processes, and as a result, mainly iron-, aluminum-, and tannin-based mordants may be used nowadays [9].

Cross-References

- [Colorant, Natural](#)
- [Colorant, Textile](#)
- [Coloration, Mordant Dyes](#)
- [Coloration, Textile](#)
- [Dye, Metal Complex](#)

References

1. Nassau, K.: The fifteen causes of color: the physics and chemistry of color. *Color. Res. Appl.* **12**(7), 4–26 (1987)
2. Zollinger, H.: *Color Chemistry – Syntheses, Properties, and Applications of Organic Dyes and Pigments*. Wiley VCH, Weinheim (2003)
3. Schweppe, H.: *Handbuch der Naturfarbstoffe*. Ecomed Verlagsges, Landsberg/Lech (1993)
4. Bechtold, T., Mussak, R. (eds.): *Handbook of Natural Colorants*. Wiley, Chichester (2009)
5. Bechtold, T., Mahmud-Ali, A., Mussak, R.: Chapter 31. Natural dyes from food processing wastes – usage for textile dyeing. In “Waste Management and Co-product in Food Processing”, pp. 502–533. Ed. Keith W. Waldron, Woodhead Publishing, Cambridge (2007). ISBN 1 84569 025 7
6. McClelland, R.A., McGall, G.H.: Hydration of the flavylum ion. 2. The 4'-hydroxyflavylum ion. *J. Org. Chem.* **47**, 3730–3736 (1982)
7. Derksen, G.C.H.: Red, Redder, Madder – Analysis and isolation of anthraquinones from madder roots (*Rubia tinctorum*). Dissertation Wageningen University, Wageningen (2001). ISBN 90-5808-462-0
8. Räisänen, R.: Anthraquinones from the Fungus *Dermocybe sanguinea* as Textile Dyes, Dissertation, Department of Home Economics and Craft Science, University of Helsinki, Helsinki, ISBN 952-10-0537-9, (2002)
9. Cardon, D.: *Natural Dyes – Sources, Tradition, Technology and Science*. Archetype Publications, London (2007)

Colorant, Nonlinear Optical

Maria Manuela Marques Raposo

Department of Chemistry, University of Minho, Braga, Portugal

Synonyms

[Dye functional](#)

Definition

Light is composed of an electromagnetic field that can interact with the elementary charges in matter, whose response can in turn influence the behavior of the other light waves. When light passes through any material, its electric field induces changes in the polarization of the material's molecules. In “linear” materials the degree of electron displacement, characterized by the linear polarizability α , is proportional to the strength of the applied electric field.

The distinguishing characteristic of nonlinear optical colorants is that their polarization response to optical waves depends nonlinearly on the applied electric field strength. This can result in the emission of new radiation fields which are altered in phase, frequency, polarization, or amplitude relative to the incident optical radiation. Many of these effects are sensitive to specific

characteristics of the local optical properties and interfaces. Multi-photon absorption can also result in electronic excitations that for a given incident light beam are more strongly localized in space than those resulting from linear absorption processes.

Nonlinear optical materials continue to attract the interest of both industrial and academic researchers due to their many versatile applications in the domain of optoelectronics and photonics.

Historical Background

Over the past decades, a variety of materials have been investigated for their nonlinear optical properties, including inorganic materials, coordination and organometallic compounds, liquid crystals, organic molecules, and polymers [1–10]. Nonlinear optics was born in 1875 when John Kerr discovered that the refractive index of glass and organic liquids could be altered by an applied electric field with the induced changes being proportional to the square of the applied field strength. Kerr measured, for the first time, the third-order nonlinear susceptibility, or what is today called the *Kerr effect* or the *electro-optical effect*. This discovery was shortly followed in 1883 by the observation of a similar but linear electric field effect in quartz; this latter process now referred to the *Pockels effect*. These nonlinear effects had limited use until the invention of the laser in 1960 by Maiman and, in the following year, the observation by Franken et al. of second harmonic generation in quartz. Following these events, the field of nonlinear optics developed explosively throughout the 1960s, highlighted by the work of Bloembergen and coworkers in exploring the full range of nonlinear optical responses of material systems and that of Buckingham and colleagues in exploring nonlinear processes in atoms and molecules of interest to chemists [11].

Initially, optical communications technology used components almost exclusively based on inorganic NLO materials. At the time it was thought that these compounds were more suitable

than organic compounds due to their greater chemical, photochemical, and physical stability and ease of processing. The two main classes of materials investigated were inorganic crystals of lithium niobate (LiNbO_3) and barium titanate (BaTiO_3) or semiconductors such as gallium arsenide (GaAs) or zinc selenide (ZnSe). Electro-optical devices that use lithium niobate have been on the market for several decades. However, these crystals have several drawbacks: high-quality single crystals are difficult to grow, are expensive, and are not easy to incorporate into electronic devices. During the 1980s it became clear that organic materials might be a better choice for use in nonlinear optical applications. A lot of organic compounds exhibit extremely high and fast nonlinearities, much better than those observed in inorganic crystals. In addition, due to the versatility of organic synthesis, their nonlinear optical properties can be modified depending on the desired application. Furthermore, organic chromophores can be incorporated into a variety of macroscopic structures such as crystals, Langmuir-Blodgett (LB) films, self-assembled films, poled polymers [1, 2, 8, 9], zeolites [12], and nanofibers [13].

At present, several organic systems (molecular as well macroscopic) with sufficiently high nonlinearities have been developed. The focus of the research in this area seems to be shifting to the optimization of a variety of other parameters, i.e., thermal and chemical stability and optical loss. Also, new approaches to design efficient nonlinear optical materials have lately emerged [7, 9].

Principles and Origin of Nonlinear Optical (NLO) Effects

In “linear” materials the degree of electron displacement, characterized by the linear polarizability α , is proportional to the strength of the applied electric field. Nonlinear optical effects arise from the interaction of electromagnetic fields in various dielectric media to produce new fields altered in phase, frequency, amplitude, or other propagation characteristics relative to the incident optical fields with a response that depends nonlinearly

on the applied electric field strength. When a beam of light propagates through a material, the electric field associated with the incident beam can provoke small displacements of the electrons within the material. This results in an induced dipole moment (μ_{ind}) which, at low electric field strengths, is linearly proportional to the amplitude of the applied electric field E , the proportionality factor being the first-order polarizability α . However, at high electric field strengths, higher-order terms become significant, and the induced dipole moment can be expanded in a Taylor series in powers of the applied electric fields (Eq. 1):

$$\mu_{\text{ind}} = \alpha E + \beta E^2 + \gamma E^3 + \dots \quad (1)$$

in which β represents the second-order polarizability or first-order hyperpolarizability and γ the third-order polarizability or second-order hyperpolarizability. This description may be expanded to the macroscopic regime of the bulk media with the first-order susceptibility and the second- and third-order nonlinear susceptibility, respectively (Eq. 2):

$$P_{\text{ind}} = \chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \dots \quad (2)$$

When the polarization of the incident and induced fields is taken in account, the above relations taken on a matrix form with the first and second hyperpolarizabilities (or equivalently the second- and third-order susceptibilities) being three and four rank tensors.

For practical applications, organic second-order NLO materials can be considered to be dipolar molecules organized into noncentrosymmetric lattices. Therefore, high β values are a necessary but not sufficient condition in order to obtain efficient NLO second-order materials [5, 8, 9].

Nonlinear Optical Effects and Their Practical Applications

The main NLO effects associated with linear and nonlinear susceptibilities are [1, 7, 9, 14–16]:

- First-order susceptibility $\chi^{(1)}$: *Refraction* is the change in direction of a light wave due to a change in its phase velocity. This phenomenon is the foundation of most geometric optical effects and has applications in optical fibers and lenses.
- Second-order susceptibility $\chi^{(2)}$:
 - (i) *Optical second harmonic generation (SHG)* is the nonlinear conversion of two photons of frequency ω into a single photon of frequency 2ω ($\omega + \omega \rightarrow 2\omega$) which, in the electric dipole approximation, requires a noncentrosymmetric medium. This effect was first demonstrated by Franken et al. [14] in 1961 and has applications in the frequency doubling of lasers. More recently, SHG-imaging has developed through the last decade as a progressively popular analytical technique especially for high-resolution optical microscopy for biological imaging. In fact nonlinear optical imaging has revolutionized microscopy for the life sciences due to the capacity of this technique to produce high-resolution images from deep inside biological tissues [15].
 - (ii) *Frequency mixing*: an NLO material may be used to add frequencies of two input waves to produce a single wave, whose frequency is their sum: $\omega_1 + \omega_2 \rightarrow \omega_3$ – or difference: $\omega_1 - \omega_2 \rightarrow \omega_3$. This effect has application in the frequency conversion of lasers and various nonlinear spectroscopic techniques.
 - (iii) *Parametric amplification*: an NLO material may transmit two light waves of different frequencies, with one light wave leaving the material amplified at the expense of the other. This effect allows the production coherent light at wavelengths for which no lasers are available and is frequently used in high-resolution spectroscopy.
 - (iv) *Linear electro-optical effect (Pockels effect)* is the change of refractive index of an NLO material which occurs on the

application of an electric field, the extent of the change being related to the strength of the field ($\omega + 0 \rightarrow \omega$). This effect has application in waveguides and electro-optical modulators.

– Third-order susceptibility $\chi^{(3)}$:

- (i) *Kerr effect* is a change in the refractive index of a material in response to an applied electric field. This effect has applications in optical transistors and image processing.
- (ii) *Third harmonic generation*: light of frequency ω enters the material and leaves with frequency 3ω , ($\omega + \omega + \omega \rightarrow 3\omega$). This effect has applications in optical image processing and scanning microscopy [7].
- (iii) *Two-photon absorption* is the induced electronic transition provoked by the near simultaneous absorption of two incident photons. This effect was first studied theoretically by Maria Goeppert-Mayer already in the 1931 and is often used in nonlinear microscopy to allow deeper penetration of the incident light, avoid photobleaching, and increase spatial resolution [16].

Nonlinear Optical Materials: Classes, Advantages, and Limitations

The search for novel chromophores for optoelectronics and nonlinear optics (NLO) is one of the main goals of modern materials science and physics. Their practical applications require not only an appropriate design but also the relevant macroscopic properties of newly established materials. The even susceptibilities χ_2 , χ_4 , etc., are only nonzero in materials lacking a center of symmetry. Therefore, the arrangements of the molecules on a macroscopic level are vitally important to NLO activity. For example, if a crystalline material with a large value of β crystallizes in a centrosymmetric structure, the nonlinear responses of the individual molecules will cancel each other out, and there will be no resultant NLO activity. The same applies to an organic

compound with an amorphous structure. If there is no order in the molecule at all, the statistical average of the NLO responses of the molecules will be zero or nearly zero, and again, the material will not be an active NLO material. Thus to use the strong hyperpolarizabilities often found in organic molecules in a bulk phase, the constituent molecules must be somehow noncentrosymmetrically ordered. The main types of ordered assemblies that have been investigated for use in NLO materials are the following [5, 8–13].

Inorganic

These materials were the first to be studied since the observation by Franken et al. [14] of second harmonic generation in quartz. Several ionic crystals such as potassium dihydrogen phosphate (KDF= KH_2PO_4), lithium niobate LiNbO_3 , barium sodium niobate ($\text{Ba}_2\text{NaNb}_5\text{O}_{15}$), and β -barium borate ($\text{BBO}=\text{BaB}_2\text{O}_4$) were developed and are currently used for several optical applications such as frequency converters, electro-optical modulators, and optical switches. The main advantages of inorganic materials are their high stability, high mechanical strength, and high nonlinear optical susceptibilities. However, the growth of these crystals is time consuming (frequently requiring 1–8 weeks) and the crystals tend to be hygroscopic, and expensive, somewhat difficult to integrate into electronic devices and are often limited by the low response speeds.

Coordination and Organometallic Complexes

In the last decade, organometallic and coordination metal complexes have occupied a relevant role in the area of NLO chromophores due to an additional flexibility, when compared to organic chromophores, owing to the presence of metal-ligand charge transfer (MLCT) transitions usually at relatively low energy and of high intensity, tunable by virtue of the nature, oxidation state, and coordination sphere of the metal center. The main classes of SHG NLO coordination and organometallic

complexes include derivatives with monodentate nitrogen donor ligands (amines, stilbazoles, pyridines), chelating ligands (Schiff bases, bipyridines, phenanthrolines, terpyridines), alkenyl, vinylidene and cyclometallated ligands, macrocyclic ligands (porphyrins and phthalocyanines), metallocene derivatives, and chromophores with two metal centers (Fig. 1) [11].

Organic

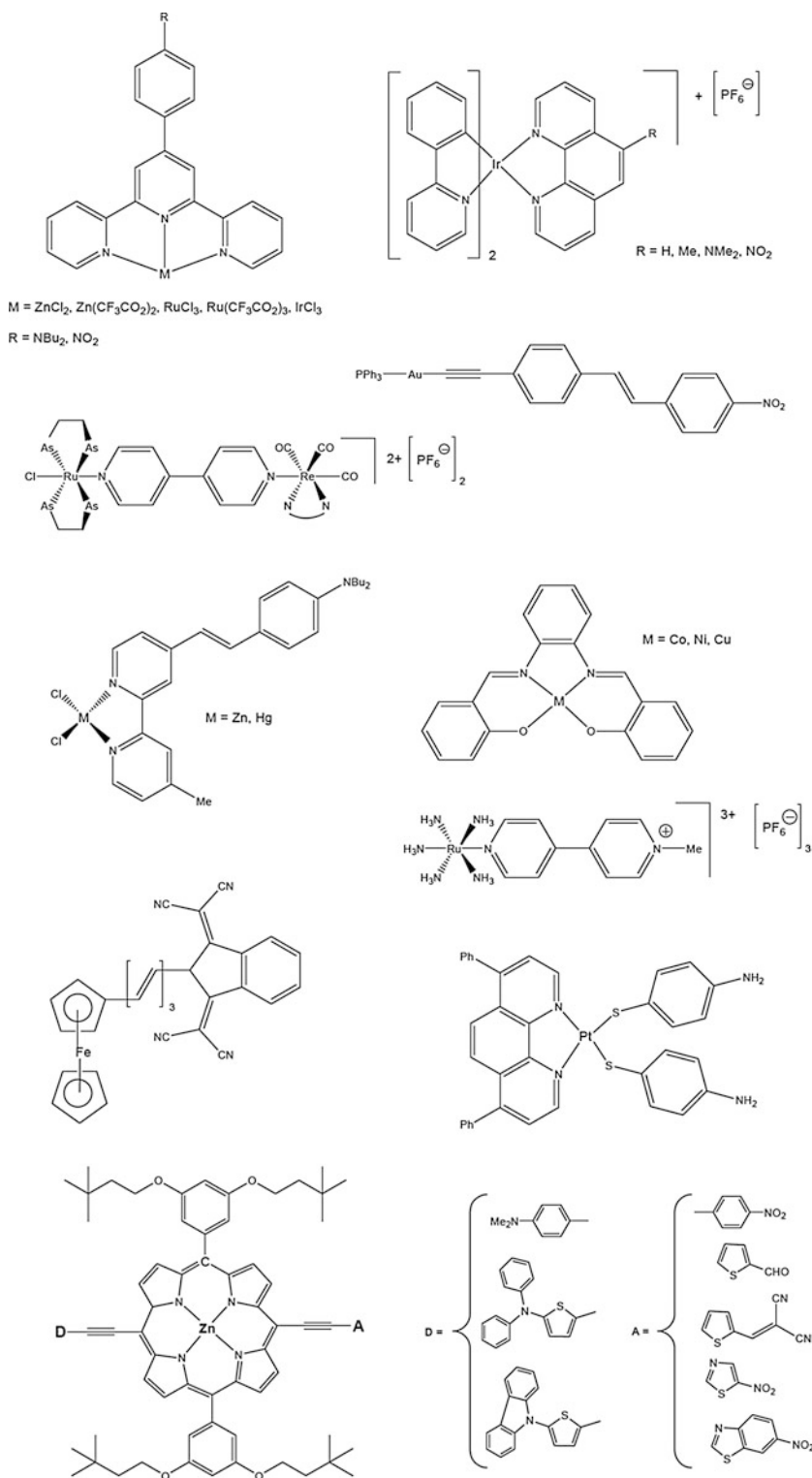
Dipolar and octopolar organic NLO chromophores (Fig. 2) have several advantages when compared to inorganic materials: (i) ultrafast response times, low dielectric constants, better tailorability and processability, as well as large NLO responses, due to the high electronic polarization of the π electrons of the molecules instead of the distortions of the atoms in the crystal lattice; (ii) easy synthesis and functionalization, allowing the optimization of their structural characteristics leading to the maximization of their NLO properties; and (iii) can be used as monocrystals and films or incorporated into liquid crystals, zeolites, or fibers. On the other hand, they have also several disadvantages such as (i) lower mechanical strength, (ii) lower photochemical stability, (iii) the propensity to acquire defects during crystal growth, and (iv) a somewhat limited working temperature range. Over the past three decades, great progress has been made in the development of new organic donor-acceptor π -conjugated systems [5, 8, 9] being driven by their chemical and thermal stability as well as the ease of synthesis and functionalization which lead to facile optical property tuning. Dipolar push-pull ($D-\pi-A$) organic chromophores are constituted by a π -bridge (aromatic or heteroaromatic) substituted with strong electron donors D (e.g., NR_2 or OR groups) and strong electron acceptor A groups (e.g., CN, NO_2 , etc.). This $D-\pi-A$ arrangement guarantees efficient intramolecular charge transfer (ICT) between the donor and acceptor groups and generates a dipolar push-pull system featuring low-energy and intense CT absorption (Fig. 2a). The polarizability and the corresponding NLO properties, namely, the SHG of these systems,

depend mainly on their chemical structure, particularly the strength of the attached donors and acceptors groups as well as the electronic nature and length of the π -conjugated bridge. However, dipolar chromophores have the tendency for unfavorable aggregation at high concentrations, and it is rare that they undergo noncentrosymmetric crystallization. One way of circumventing these disadvantages is the use of octopolar molecules.

Octopolar NLO chromophores are not so commonly investigated compared to dipolar ones. Nevertheless, their advantage is that they show the same optical transparency as their linear analogues but their second-order response is higher. Due to the fact that they possess zero overall dipole moment, they can often be oriented in the bulk phase in a manner that their molecular polarizability is additive. On the other hand, not possessing a permanent dipole moment invalidates their use in some electro-optical applications. The usual way to design SHG octopolar molecules consists in the synthesis of substituted trigonal or tetrahedral π -conjugated systems that allows an efficient charge transfer from periphery to the center of the molecule (Fig. 2b).

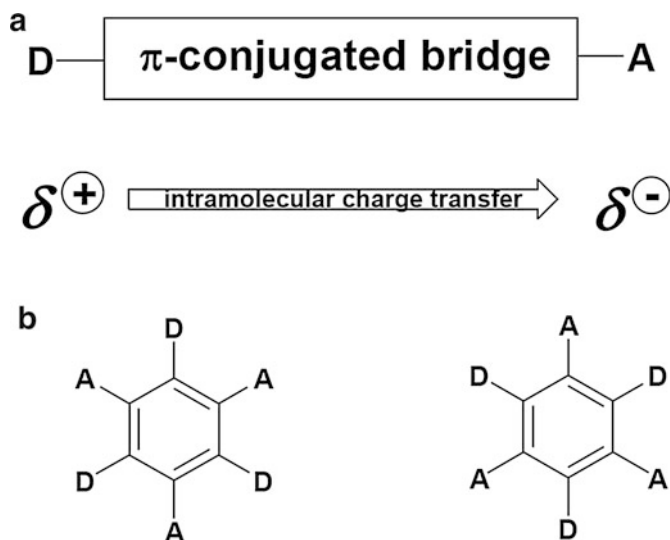
Several one-dimensional charge transfer systems with good SHG properties were developed during the 1980s. Typical examples of such molecules are *p*-nitroaniline (*p*NA) and *N,N*-dimethylaminostilbene (DANS) (Fig. 3); both are still being used as standards to evaluate SHG of other molecules.

Chromophores with other types of conjugated spacers have been also investigated (e.g., substituted benzenes, biphenyls, stilbenes, and azostilbenes). All these systems have a predominantly aromatic ground state and a corresponding charge transfer state that is quinoidal in nature. More recently, the investigation by several groups has led to the development of highly advanced SHG organic chromophores. The results of these studies suggest that optimal β values could be obtained through two different but complementary methodologies: the optimization of the π -bridge structure for particular donor and acceptor pairs and optimization of donor and acceptor groups for a given π -bridge. Novel highly efficient donor-acceptor pairs have been developed

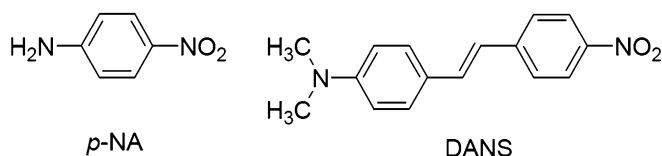


Colorant, Nonlinear Optical, Fig. 1 Structure of several SHG organometallic and coordination metal complexes [11]

Colorant, Nonlinear Optical, Fig. 2 Schematic representations of (a) a dipolar organic D- π -A system featuring ICT and (b) octopolar chromophores



Colorant, Nonlinear Optical, Fig. 3 Structure of *p*-nitroaniline (*p*-NA) and *N,N*-dimethylaminostilbene (DANS)

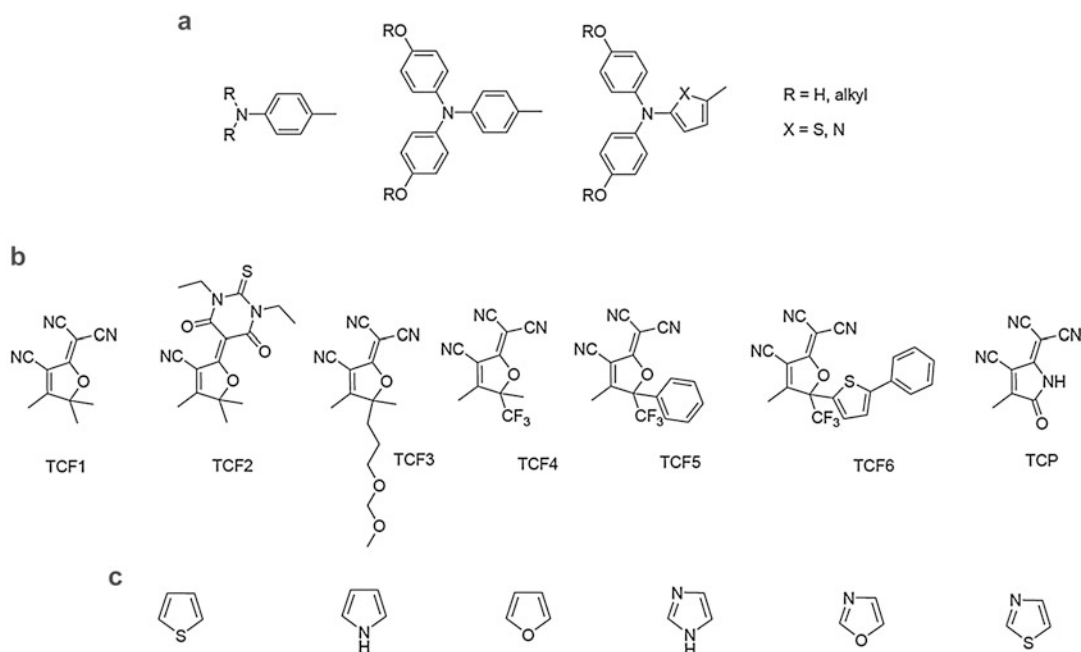


such as alkyl and arylamine electron donors (Fig. 4a) and tricyanofuran-based electron acceptors, for example, 2-dicyanomethylene-3-cyano-4,5,5-trimethyl-2,5-dihydrofuran (TCF1), 2-dicyanomethylene-3-cyano-4,5-dimethyl-5-trifluoromethyl-2,5-dihydrofuran (TCF4), and the tricyanopyrrole (TCP) electron acceptor (Fig. 4b). This arrangement of donor-acceptor pairs combines high chemical, thermal, and photostability [5, 8, 9]. In recent years the enhancement of β values by using ever stronger electron donor or acceptor groups has tended to reach its limit. A versatile methodology to overcome this problem was the optimization of the π -bridge. Therefore, several experimental and theoretical studies have confirmed that substitution of the benzene ring of a chromophore bridge with easily delocalizable heterocycles (e.g., thiophene, pyrrole, furan, and thiazole) (Fig. 4c) results in improved molecular hyperpolarizability of push-pull systems. Due to their electronic nature and low aromaticity, they can act efficiently as π -bridges as well as auxiliary donors (electron-rich heterocycles: thiophene, pyrrole, furan) or as

auxiliary acceptors (electron-deficient heterocycles: thiazole, oxazole, imidazole). In fact, the increase or decrease of the molecular nonlinear activity of these heteroaromatic systems depends not only on the electronic nature of the aromatic rings but also on the location of these heterocycles in the system. Additionally, these heterocyclic systems are also thermally and photochemically stable [8, 9, 17–20].

Concurrently, multidimensional charge transfer (e.g., X-shaped and higher-order symmetry) and twisted intramolecular charge transfer (TICT) chromophores (Fig. 5) have been explored as alternative approaches to improve hyperpolarizability [9].

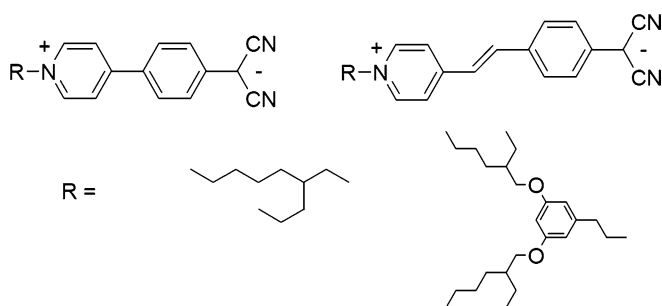
The easiest method to impose order on the molecules of a compound is to assemble them into a crystalline matrix. Crystals have several advantages such as their high optical quality and high laser damage thresholds. On the other hand, they have several serious drawbacks; a major one is the fact that most of the promising molecules have significant ground state dipole moments, which tend to make them crystallize



Colorant, Nonlinear Optical, Fig. 4 Structures of (a) aryl amines as examples of strong electron donor groups, (b) electron acceptor groups belonging to the general TCF

and TCP classes, and (c) heterocyclic compounds as π -bridges/auxiliary donor or auxiliary acceptor groups [8]

Colorant, Nonlinear Optical, Fig. 5 Structures of twisted intramolecular charge transfer molecular (TICT) chromophores [9]



centrosymmetrically. For example, 4-nitroaniline (pNA) packs centrosymmetrically and exhibits no appreciable SHG in crystalline form, while the analogous 2-methyl-4-nitroaniline (MNA) packs in an almost perfect head-to-tail arrangement and has a large χ_2 value. There are some examples of noncentrosymmetric small molecules such as derivatives of nitroanilines and nitropyridines or enantiomers of an optically active component. For example, MMONS (3-methyl-4-methoxy-4'-nitrostilbene) has a very high powder SHG value (1,250 times that of urea) and POM

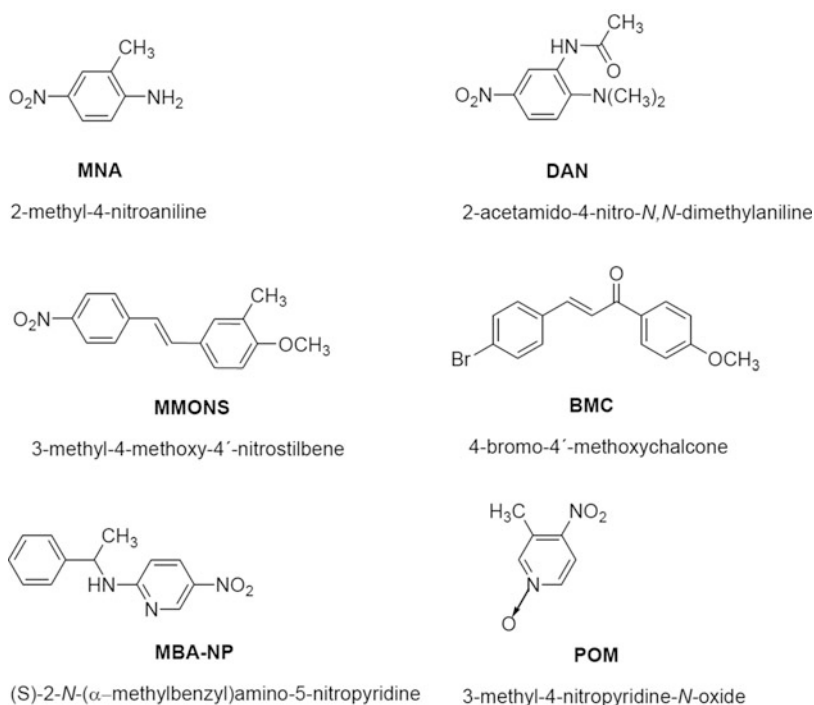
(3-methyl-4-nitropyridine-*N*-oxide) is the only commercially available organic crystal for SHG, other than urea (Fig. 6).

Polymeric

Oriented Guest-Host Polymers

Second-order NLO applications that require crystalline materials limit the scope of molecule types that can be employed to those that crystallize in acentric space groups. To achieve good device

Colorant, Nonlinear Optical, Fig. 6 Structure of selected organic NLO crystals [1, 4, 5, 9]

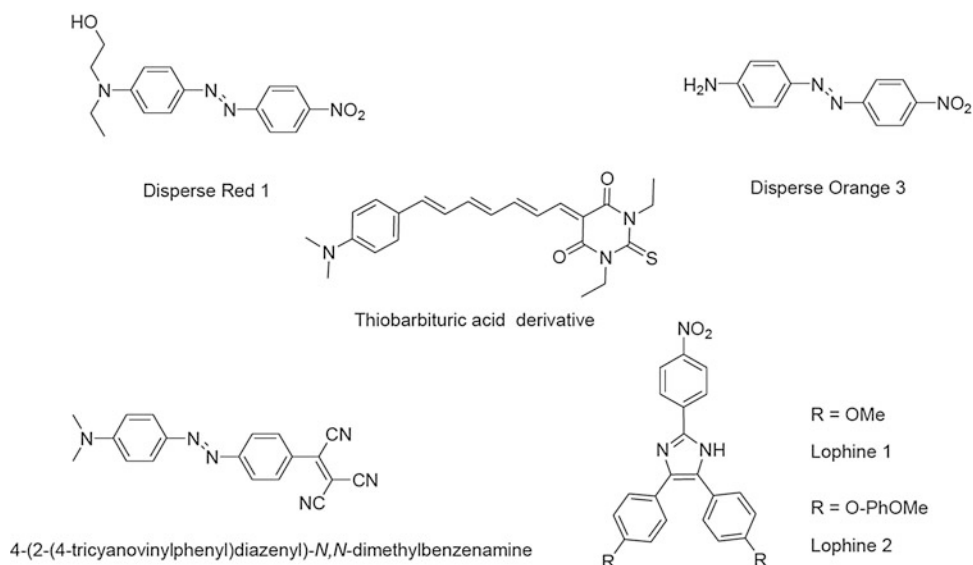


functionality, the NLO chromophore must simultaneously possess high microscopic molecular nonlinearity, good thermal stability, good photostability, low optical absorption, and weak intermolecular electrostatic interactions in a given host matrix. Polymeric materials are attractive because they are compatible with manufacturing methods practiced in industry and can provide durability, environmental protection, and packing advantages not provided by crystalline materials. Nonlinear optical chromophores can be incorporated into a macroscopic polymeric environment in a variety of ways. Probably the most important and most widely used is the incorporation of dipolar chromophores into a polymer host. In this approach, the active species (the guest) is dissolved in a polymeric host, which is processed to give a thin film. At this stage the molecular dipoles are randomly orientated with respect to each other. The polymer is heated above its glass transition temperature (T_g) allowing the guest molecules to become freely mobile. A strong external electric field is then applied, aligning the dipoles along one direction. With the field still applied, the polymer is cooled below its T_g

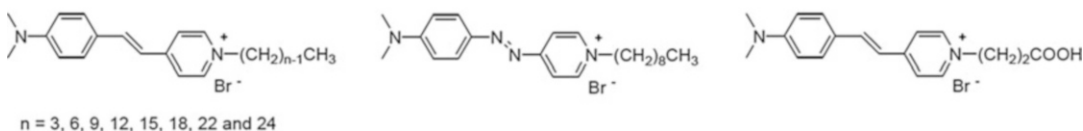
again, freezing the alignment of the NLO molecules. This approach is successful in obtaining highly oriented NLO materials showing large bulk susceptibilities. The main disadvantages are (i) gradual disordering of the dipoles, (ii) limited solubility of the active species in the host polymer which limits the attainable NLO activity, and (iii) dielectrically induced breakdown of the NLO species during poling. The advantages of the poled polymer approach are the relative ease of thin film making by spin coating and its compatibility with existing semiconductor technology. Figure 7 shows the structures of some examples of organic chromophores used as guests in guest-hosts polymeric systems.

Oriented Side-Chain and Main-Chain Polymers

Nonlinear optical chromophores can be also incorporated into a polymer by covalently attaching the chromophores to a polymeric backbone as part of the side chain (side-chain polymers) or by incorporation of the chromophores into the backbone of the polymer (main-chain polymers).



Colorant, Nonlinear Optical, Fig. 7 Structure of selected organic chromophores used in guest-host polymers [8, 9]



Colorant, Nonlinear Optical, Fig. 8 Structure of second harmonic generator compounds that have been incorporated into zeolite hosts [12]

Langmuir-Blodgett Films

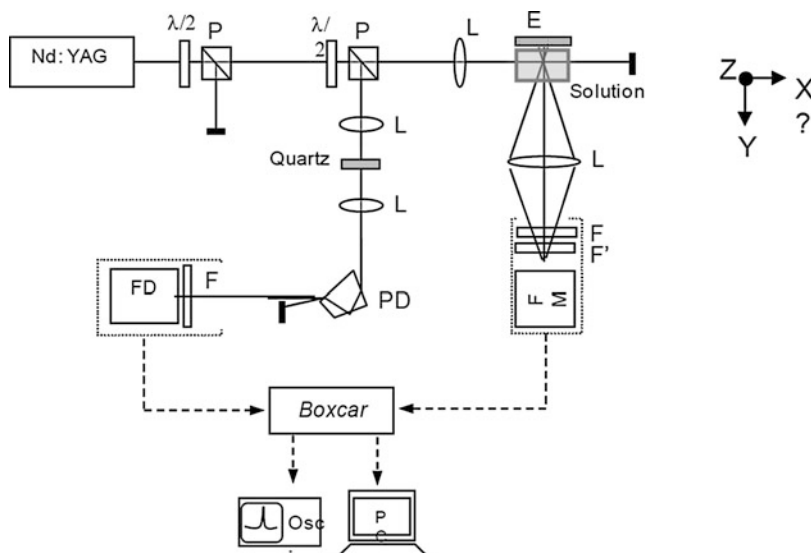
Another approach to organize molecules has been to incorporate organic NLO chromophores into noncentrosymmetric Langmuir-Blodgett films. The Langmuir-Blodgett film technique is used to build up ordered assemblies of molecules onto a substrate from a floating monolayer on a liquid, usually water. The molecule must be amphiphilic, that is, hydrophilic at one end and hydrophobic at the other, so that the molecules in the monolayer at the water's surface have uniform orientation. This approach offers the advantage of much greater chromophore alignment and chromophore density. However, these films often have poor optical quality and poor temporal stability and are often very fragile.

Zeolites

Zeolites and the related mesoporous materials have been also tested as the hosts for aligned inclusion of dipolar organic NLO dyes as a possible means to overcome the problems arising from the use of polymer matrices. The first dyes tested were *p*-nitroaniline (*p*NA), 2-methyl-4-nitroaniline (MNA), and 2-amino-4-nitropyridine which have low molecular second-order hyperpolarizability values, and the zeolites have been mostly limited to powders that bear limited practical applicability. More recently hemicyanine dyes (Fig. 8) exhibiting higher β values and a high degree of uniform orientation were introduced into transparent zeolite films with uniformly oriented channels. These second-order

Colorant, Nonlinear

Optical, Fig. 9 Scheme of assembly for measurements of diffusion hyper-Rayleigh. *P* polarizer, $\lambda/2$ half-wave plate, *L* lens, *E* mirror, *PD* disperser prism, *F'* band-pass filter, *F* low-pass filter, *FD* photodetector, *FM* photomultiplier (Nonlinear Optics Laboratory of the Physics Center at the University of Minho) [19, 20]



NLO dye-incorporating films have shown higher thermal and mechanical stability without any notable loss of activity with time. They have a strong potential to be practically applied in industry [12].

Nanofibers

Very recently, nanofibers of poly(L-lactic) acid (PLLA) produced by the electrospinning technique, in which donor-acceptor organic compounds such as 2-methyl-4-nitroaniline, urea, and β -glycine were imbedded, exhibit a permanent nonvanishing quadratic nonlinear susceptibility. The nonlinear optical properties displayed indicate that a noncentrosymmetric polar state was achieved and maintained a long time, allowing the use of otherwise centrosymmetric organic materials. Moreover, it was proved that the SHG efficiency of these fibers strongly depends on the diameter of the nanofibers and can be increased up to an order of magnitude by controlling the electrospinning processing parameters [13].

As most of the donor-acceptor organic molecules with large hyperpolarizabilities tend to crystallize in centrosymmetric structures which invalidate their use in quadratic nonlinear optical applications, the inclusion of these molecules in

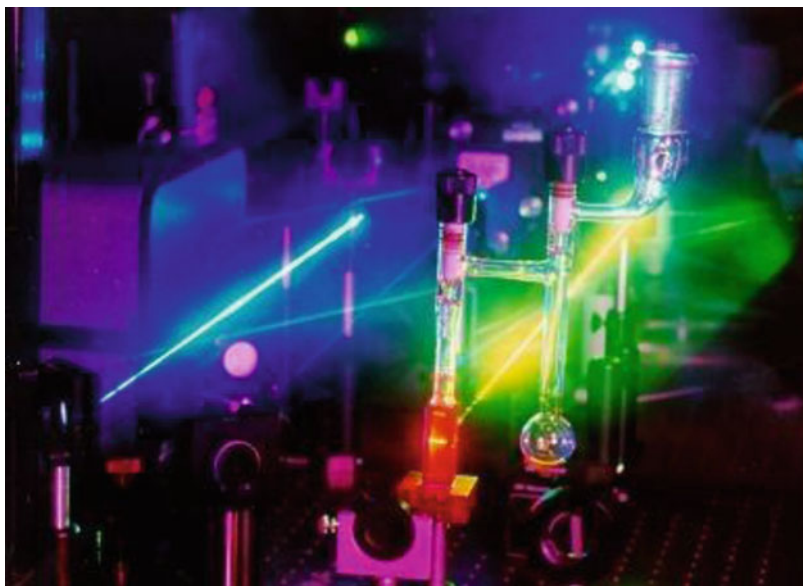
electrospun nanofibers may provide a means of overcoming this limitation for any donor-acceptor organic molecules with delocalized π electrons. The results of this work could have a direct impact on the design of novel nanodevices for a variety of nanophotonic applications (e.g., electro-optical transducers, pyroelectric sensors, optical frequency converters).

Experimental Methods for the Determination of Second-Order Nonlinear Effects

The second-order polarizability β and the second-order susceptibility $\chi^{(2)}$ are two parameters indicative of a second-order response [5, 9]. The first is a molecular parameter and is usually measured in solution, whereas the latter is typically measured by second harmonic generation from the solid state. Several experimental techniques can be used in order to study these parameters in solution or in solid state: solvatochromic method, Kurtz powder technique, Maker fringes, electric field-induced second harmonic (EFISH) technique, and the hyper-Rayleigh scattering (HRS) technique (Figs. 9 and 10). In solution, the two major techniques that are presently used are EFISH and HRS.

Colorant, Nonlinear Optical,

Fig. 10 Experimental setup for the determination of SHG through hyper-Rayleigh scattering technique (Nonlinear Optics Laboratory of the Physics Center at the University of Minho) [19, 20]



Future and Perspectives

From the proceeding description, it is clear that the molecular design of molecules with strong nonlinear optical responses has reached a high level of sophistication. Important advances are likely to come from breakthroughs in methods that are able to make use of these strong individual molecular responses by incorporating in restricted environments that break their tendency to aggregate in centrosymmetric forms. The future of organic nonlinear optical materials is undoubtedly a bright one.

Cross-References

- [Dye](#)
- [Dye, Functional](#)
- [Dye, Metal Complex](#)

References

1. Chemla, D.S., Zyss, J.: *Nonlinear Optical Properties of Organic Molecules and Crystals*, vol. 1 and 2. Academic Press, New York (1987)
2. Prasad, P.N., Williams, D.J.: *Introduction to Nonlinear Optical Effects in Molecules and Polymers*, pp. 132–174. Wiley, New York (1991)
3. Zyss, J.: *Molecular Nonlinear Optics: Materials, Physics and Devices*. Academic Press, Boston (1994)
4. Nalwa, H.S., Miyata, S. (eds.): *Nonlinear Optics of Organic Molecules and Polymers*. CRC Press, New York (1997)
5. Verbiest, T., Houbrechts, S., Kauranen, M., Clays, K., Persoons, A.: Second-order nonlinear optical materials: recent advances in chromophore design. *J. Mater. Chem.* **7**, 2175–2189 (1997)
6. Meyers, F., Marder, S.R., Perry, J.W.: Advanced polymeric materials - high performance polymers. In: Interrante, L.V., Hampden-Smith, M.J. (eds.) *Chemistry of Advanced Materials: An Overview*, pp. 207–269. Wiley-VCH, New York (1998)
7. He, G.S., Tan, L.-S., Zheng, Q., Prasad, P.N.: Multiphoton absorbing materials: molecular designs, characterizations, and applications. *Chem. Rev.* **108**, 1245–1330 (2008)
8. Cho, M.J., Choia, D.H., Sullivan, P.A., Akelaitis, A.J. P., Dalton, L.R.: Recent progress in second-order nonlinear optical polymers and dendrimers. *Prog. Polym. Sci.* **33**, 1013–1058 (2008)
9. Dalton, L.R., Sullivan, P.A., Bale, D.H.: Electric field poled organic electro-optic materials: state of the art and future prospects. *Chem. Rev.* **110**, 25–55 (2010)
10. New, G.H.C.: Nonlinear optics: the first 50 years. *Contemp. Phys.* **52**, 281–292 (2011)
11. Di Bella, Dragonetti, C., Pizzotti, M., Roberto, D., Tessore, F., Ugo, R.: Coordination and organometallic complexes as second-order nonlinear optical molecular materials. *Top. Organomet. Chem.* **28**, 1–55 (2010)
12. Kim, H.S., Cao, T., Pham, T.C.T., Yoon, K.B.: A novel class of nonlinear optical materials based on host–guest composites: zeolites as inorganic crystalline hosts. *Chem. Commun.* **48**, 4659–4673 (2012)

13. Isakov, D.V., de Matos, G., Belsley, M.S., Almeida, B., Cerca, N.: Strong enhancement of second harmonic generation in 2-methyl-4-nitroaniline nanofibers. *Nanoscale* **4**, 4978–4982 (2012), and references cited
14. Franken, P.A., Hill, A.E., Peters, C.W., Weinreich: Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119 (1961)
15. Reeve, J.E., Anderson, H.L., Clays, K.: Dyes for biological second harmonic generation imaging. *Phys. Chem. Chem. Phys.* **12**, 13848–13498 (2010)
16. Pawlicki, M., Collins, H.A., Denning, R.G., Anderson, H.L.: Two-photon absorption and the design of two-photon dyes. *Angew. Chem. Int. Ed.* **48**, 3244–3266 (2009)
17. Varanasi, P.R., Jen, A.K.-Y., Chandrasekhar, J., Namboothiri, I.N.N., Rathna, A.J.: The important role of heteroaromatics in the design of efficient second-order nonlinear optical molecules: theoretical investigation on push – pull heteroaromatic stilbenes. *J. Am. Chem. Soc.* **118**, 12443–12448 (1996)
18. Breitung, E.M., Shu, C.-F., McMahon, R.J.: Thiazole and thiophene analogues of donor-acceptor stilbenes: molecular hyperpolarizabilities and structure-property relationships. *J. Am. Chem. Soc.* **122**, 1154–1160 (2000)
19. Raposo, M.M.M., Sousa, A.M.R.C., Kirsch, G., Cardoso, P., Belsley, M., Matos Gomes, E., Fonseca, A.M.C.: Synthesis and characterization of dicyanovinyl-substituted thienylpyrroles as new NLO-chromophores. *Org. Lett.* **8**, 3681–3684 (2006), and references cited
20. Raposo, M.M.M., Fonseca, A.M.C., Castro, M.C.R., Belsley, M., Cardoso, M.F.S., Carvalho, L.M., Coelho, P.J.: Synthesis and characterization of novel diazenes bearing pyrrole, thiophene and thiazole heterocycles as efficient photochromic and nonlinear optical (NLO) materials. *Dyes Pigments* **91**, 62–73 (2011)

Colorant, Photochromic

Andrew Towns

Vivimed Labs Europe Ltd., Huddersfield, West Yorkshire, UK

Definition

The defining characteristic of photochromic colorants is that they change color reversibly in response to variations in the intensity of particular wavelengths of light to which they are exposed.

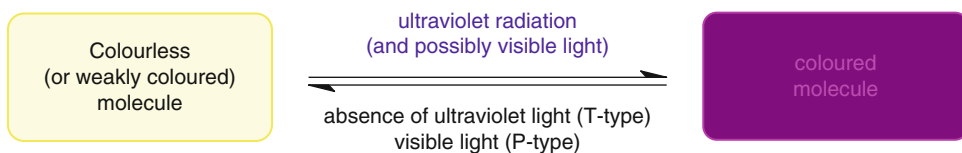
Light-responsive dyes worth millions of dollars are manufactured each year as a result of their successful exploitation over the last quarter century [1]. The bulk is consumed in the production of ophthalmic lenses that darken reversibly when exposed to strong sunshine. Photochromism continues to attract the interest of both industrial and academic researchers, who are looking to harness photochromic colorants in fields like optoelectronics and nanotechnology.

Photochromism

The widely accepted definition of photochromism is that of a reversible color change induced in a compound driven in one or both directions by the action of electromagnetic radiation [2, 3]. Photochromic systems are classified as either “P-type” or “T-type”. The former kind can be switched in each direction with different wavelengths of light. P-type systems change color when irradiated with a specific wavelength range then remain in this state after removal of the stimulus. It is only when they are subjected to light of a different set of wavelengths that they return to their original color. In contrast, T-type behavior is exhibited if light is able to drive the change in just one direction. T-type systems will fade back to their original state, through a thermal back reaction, when they are no longer exposed to the light source. Reversibility of response is a key aspect in both types of photochromism: light-sensitive materials that undergo changes of an irreversible nature cannot be described as photochromic.

Real-world colorants do not always match the strict definitions of the two kinds of behavior, as discussed further below. Nevertheless, most are readily categorized. Photochromic compounds of either type are available commercially. While T-type dyes are far more important industrially, there is much interest in P-type materials. The behavior of both is captured very generally in Fig. 1.

The first major application of photochromism was commercialized in the mid-1960s: glass ophthalmic lenses that relied on inorganic halide crystals. These systems have been superseded during



Colorant, Photochromic, Fig. 1 General behavior of most commercial T-type and P-type photochromic colorants

the past two decades by organic materials in the form of plastic lenses incorporating T-type dyes [4, 5]. Such colorants are colorless – so to describe them as dyes might initially seem odd! – but they become colored when irradiated with sunlight and fade back thermally to colorless in low levels of light. Developing an economic technology with commercially acceptable durability and performance did not prove easy because photochromic colorants are less robust than conventional dyes and orders of magnitude more expensive. Light-responsive organic compounds that can be switched from one color to another are well known, but because photochromic lens manufacture remains the dominant application, the most industrially important dyes are those which exhibit T-type behavior as depicted in Fig. 1 involving colorless to colored transitions. This kind of color change, in which light causes a shift in absorption to longer wavelengths, is known as “positive photochromism”. The term “negative photochromism” does not mean non-photochromic, but instead covers the converse phenomenon of a colored dye becoming colorless upon irradiation with light, only to return to its original colored state in the dark [2].

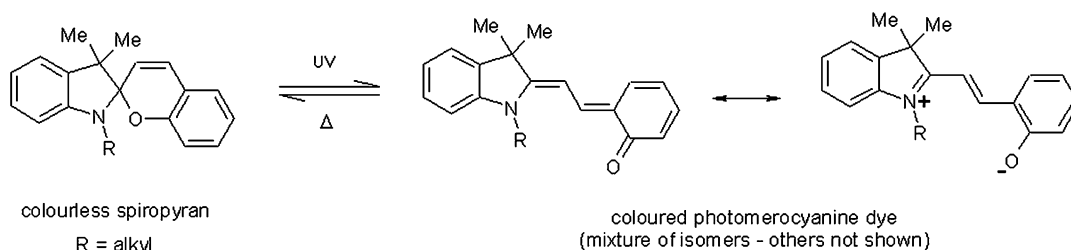
T-Type Photochromic Colorants

There are many examples of organic compounds that change color upon irradiation with light and revert to their original state following removal of the illuminant [2, 6, 7]. The photochromism is T-type because the back reaction is driven thermally, although for commercial photochromic classes, visible light may also contribute. The rate of thermal fading is often expressed as “half-life” which is the time taken for absorbance to halve once the activating light has been

removed. For ophthalmic utility, a short half-life is desirable to stop vision being impaired when there is a sudden drop in light intensity [4]. However, a commercially interesting colorant for lens production must satisfy numerous other requirements:

- Weak visible absorption when unactivated so residual color is low
- Quick response to an increase in illumination
- Have a strongly absorbing activated state, because even when irradiated with light of high intensity, only a relatively small proportion of colorant will exist in this form
- Show a good compromise between depth of activated color and rate of fade to ensure both are acceptable as the properties of high intensity and short half-life tend to be mutually exclusive
- Produce satisfactory performance over 2 years by having reasonably lightfast colored and colorless forms which respond well to photostabilizers
- Resist the tendency to “fatigue”, whereby during activation, a proportion of the dye is undesirably and irreversibly converted to non-photochromic molecules, leading to gradual weakening of color upon repeated activation
- Color up in a manner that is not greatly affected by the temperature of its environment
- Exhibit sufficient solubility in lens media to give solutions rather than dispersions because commercial T-type classes do not exhibit useful photochromism in solid form

All of the commercial T-type dye classes undergo the same kind of molecular transformation, photoisomerization, as illustrated by an example of the well-studied spiropyran family in



Colorant, Photochromic, Fig. 2 Photochromism of typical example of the spiropyran class

Fig. 2 [3, 6]. The geometries of commercial T-type colorants change substantially when switching between colorless and colored states, which means that the medium in which a dye is dissolved can markedly influence its photochromism. Non-polar solvents typically provide a favorable environment for photoisomerization, whereas, as discussed below, polymers can inhibit interconversion.

The wavelengths of light that effect the forward conversion shown in Fig. 2 are normally within the UVA region (315–400nm), but blue light can also play a role for some commercial spiro derivatives. Absorption causes rearrangement of the bonding between the atoms within a colorless or weakly colored molecule to create structures that confer intense color. The colorless form consists of a ring-closed structure, which is made up of two halves that are perpendicular to each other and joined by a spiro carbon atom. The π -systems of these moieties are small, hence the lack of absorption in the visible region. However, absorption of energy in the form of UV light can lead to ring-opening as a result of the bond between the spiro carbon atom and its adjacent oxygen atom breaking. Molecular twisting and bending via intermediates then ensues giving planar species with extended conjugated π -systems whose absorption moves into the visible, generating color. Low light levels result in ring closure back to the colorless form because thermal fading dominates.

Different forms of the dyes exist in a dynamic equilibrium: at a given moment, molecules are isomerizing between colored and colorless species, the concentrations of which are determined by the intensity of incident light. As the flux of

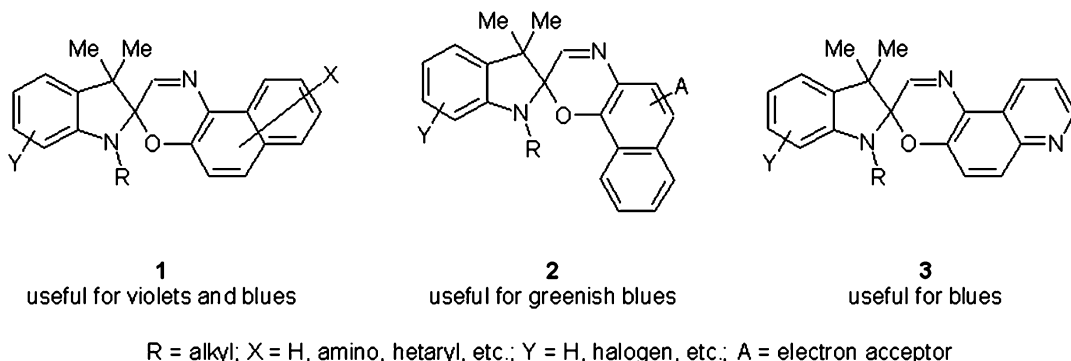
UV that the dye is exposed to increases, the proportion of dye that is in its colored state grows through promotion of ring opening relative to ring closure. Removal of the light source leads to the concentration of the colorless ring-closed form rising which is observed as fading. When the intensity of illumination is held constant, the isomer concentrations will settle down into what is known as a photostationary state, where depth of color does not change. These proportions are dependent on the dye, the nature of the illumination and the medium. Because the photostationary state is a dynamic equilibrium, dye molecules will continue to swap between colorless and colored isomeric forms even after it has been attained.

Since a significant proportion of sunlight is made up of radiation in the UV, it is capable of causing pronounced photochromism. In contrast, most commercial T-type dyes do not respond well to artificial light sources, such as tungsten filament bulbs, because the proportion of their UV output is low.

Several classes of T-type compounds have been extensively investigated (e.g., anils, perimidinespirocyclohexadienones, spirodihydroindolizines, etc.) [6, 7], but few have attained commercial significance. The three families of dye that have had the greatest industrial importance will be discussed next.

Spiropyrans

This class were intensively studied during the 1950s through to the 1970s because they are readily synthesized and photochrome to deep colors, typically violets and blues, that fade at useful rates [6]. However, because members of the spiropyran family generally have poorer photostability than their spirooxazine and naphthopyran



Colorant, Photochromic, Fig. 3 Some photochromic oxazines of commercial importance

counterparts, they are much less important commercially. Nevertheless, spiropyrans are still exploited in research where light stability is not a prerequisite, for example, in the fields of biochemistry and materials science.

Spirooxazines

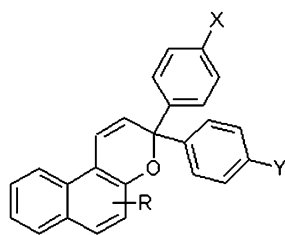
Although spirooxazines [6, 9] are similar in structure to spiropyrans, the former are much less prone to fatigue [3, 8]. As a result, they have become well established since they were employed in the production of the first commercial plastic photochromic lenses in the early 1980s [4].

Dyes derived from the spiroindolinonaphth[2,1-*b*][1,4]oxazine **1**, spiroindolinonaphth[1,2-*b*][1,4]oxazine **2**, and indolinospiripyridobenzoxazine **3** classes have all enjoyed extensive use in ophthalmic lens manufacture (see Fig. 3) [1]. The simplest examples give relatively fast-fading blue photocoloration but small adjustments to structure furnish dyes with useful intensities and half-lives. For example, placing bulky alkyl groups on the indoline nitrogen, i.e., substituent R in Fig. 3, slows down fading and increases strength. Manipulation of color is also possible with appropriate design making bluish-red through to turquoise dyes accessible commercially. Dyes **3** were successfully utilized as the blue components of a photochromic colorant mixture in the production of the first true gray photochromic lenses in the early 1990s.

Naphthopyrans

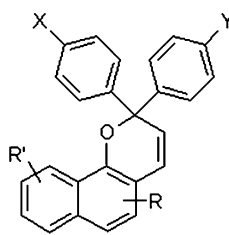
All the major manufacturers of plastic photochromic lenses make use of the naphthopyran class [6, 8, 10] in their formulations [1]. Its chemistry facilitates the economical production of a palette of dyes that not only spans the visible spectrum from yellows through to oranges, reds, purples, and blues but also features more neutral colors such as olive, brown, and gray. This family offers considerable scope for fine-tuning of photochromic properties because many convenient modifications to structure are possible. However, careful design is needed since substituent choice usually affects both kinetics and color. The stability of such dyes is generally as good as any other class, while their photochromism tends to be more independent of temperature than that of spirooxazines. The above advantages have led to naphthopyrans becoming the most commercially important type of photochromic colorant in the form of two subclasses: 3*H*-naphtho[2,1-*b*]pyrans **4** and 2*H*-naphtho[1,2-*b*]pyrans **5** (see Fig. 4).

Another reason for the significance of naphthopyrans is that they simplify the development of colorant recipes for lenses of the most commercially important shades, which are gray and brown [1, 8]. Incorporation of appropriate structural features into **5** produces relatively dull, neutral dyes, enabling use of fewer colorants in a mixture. Given that all of the components of the recipe must activate, fade, and fatigue at the same rate, this is of great advantage to the formulator.



4

useful for yellow, oranges and reds



5

useful for reds, purples and blues

X, Y = H, alkoxy, amino, etc.; R = H, alkyl, aryl, ester, amino, etc.; R' = H, aryl, amino, alkoxy, etc.

Colorant, Photochromic, Fig. 4 Some photochromic naphthopyrans of commercial importance

Applications of T-Type Dyes

While the properties of photochromic colorants have been put to purely aesthetic use in artwork, T-type dyes have also been exploited as functional materials, for example, in security printing where light-responsive marks are used as indicators of authenticity. The main outlet for photochromic dyes, lenses for spectacles, falls between these two extremes: variable transmittance is a key function in regulating light intensity reaching the eye, yet color is also an important consideration both for style and comfort. Although T-type colorants have been investigated for many applications, success in developing marketable products has been limited by the challenges that their usage presents. They are not as robust as conventional dyes and pigments, rendering certain methods of application unsuitable. Also, in order to get strong, durable photocoloration, care must be taken to provide T-type dyes with the right environment in which to operate. It must be conducive to the changes in molecular geometry associated with photochromism and not lead to rapid degradation of dye during and/or following application. These considerations will be illustrated for polymers and surface coatings, which are the two most common media for photochromic dyes.

Arresting photochromic effects can be produced by incorporation of dyes into thermoplastics. However, the chemical and physical nature of the polymer (as well as the dye) has a large influence on the kinetics and resilience of the photochromism. For example, spirooxazines and

naphthopyrans work well when used in mass coloration of polymers that have relatively low glass transition temperatures with flexible chains, such as polyethylene and polypropylene, giving striking color changes at inclusion levels of 0.3%w/w and less. However, rigid, crystalline materials restrict the necessary conformational changes for photoisomerization, severely inhibiting photochromism. Dyes can tolerate brief exposure to the high temperatures experienced in injection molding or extrusion, but certain polymers, such as polyamides, require processing at elevated temperatures that strongly degrade colorants, leading to discoloration and a lack of observable photochromism. Even when suitable polymers are employed, additives are often needed to enhance dye photostability in order to achieve an effect with a commercially acceptable lifetime. Loss of photochromism is related to cumulative amount of incident UV radiation rather than the number of times that the material is switched between colorless and colored states. Consequently, UV absorbers can usefully shield dyes from excessive radiation provided that they do not strongly absorb the UV wavelengths which activate the dyes. Other additives include hindered amine light stabilizers that scavenge light-generated free radicals, which would otherwise attack colorants, and triplet state quenchers that inhibit photochemical reactions other than the desired one of photoisomerization.

Application to polymers need not involve monomolecular dissolution in the polymer itself.

Photochromic dyes can also be used in microencapsulated solvent droplets of typically 1–10 μm in diameter. In this form, the dye solution is encased in polymer shells, producing a photochromic powder that can be dispersed like a pigment. The advantage of this material is that its photochromism is much more independent of the medium, i.e., the properties are that of the dye dissolved in the solvent and do not tend to be influenced by the polymer in which the microcapsules are dispersed. In this way photochromic effects can be produced in substrates using the microencapsulate that would not be possible using dye alone. Disadvantages of this technique are that relatively high loadings of pigment are required and the microcapsules can be physically damaged during application. An alternative approach to obtaining photochromism in inhospitable polymers (as well as potentially improving the robustness and responsiveness of the effect) is to attach oligomeric fragments to dyes. These “tails” are thought to provide a favorable microenvironment for the photoactive part of the colorant. All of these methodologies have been commercialized for coloration of polymers. Photochromic materials are found in products as diverse as toys, fashion accessories, fishing line, and motorcycle helmet visors [1].

T-type colorants have also been used for applications ranging from security printing to cosmetics in the form of photochromic inks and other surface coatings. Commercial dyes work well in nonpolar solvents such as toluene when dissolved at a suitable concentration: solubility and photochromic behavior tend to be good, enabling the formulation of solvent-based inks and varnishes. Aqueous media require an alternative approach since such dyes are not water soluble but must be in solution to exhibit photochromism. One way is to disperse microencapsulated dyes, which were described earlier, into water-based systems like a pigment using commercially available powders or aqueous slurries. Alternatively, micronized particles of an appropriate polymer into which photochromic dye has been incorporated can be employed. It is possible to formulate inks for a variety of printing techniques provided care is taken to ensure that

dye is not damaged during formulation or application. Additives may be needed to improve light fastness. Even following optimization, photostability can be problematic. This is also true of photochromic textiles: the most effective method of applying dye to fabric is screen printing microencapsulated colorant because typical polymers, such as polyester, inhibit photochromism, while conventional techniques, e.g. exhaustion dyeing, tend to damage typical commercial dyes. Photochromic detail can be added to garments through the use of polypropylene thread that has been melt spun with dye [1].

The general lack of robustness of photochromic dyes compared to conventional colorants has prevented their use in particularly demanding applications. One example is light-responsive glass for architectural windows. Lifetimes in excess of 10 years are required but such a demand cannot be met by existing dye technology.

T-type colorants are also not suited to potentially important uses which require controlled switching between one or more states (colored and/or colorless) on demand: for these applications, P-type dyes are needed.

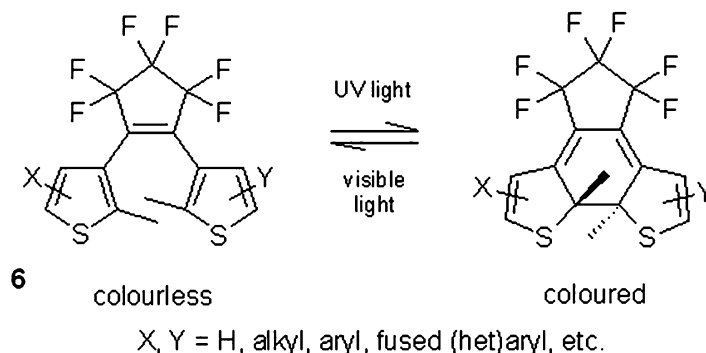
P-Type Photochromic Colorants

A significant amount of effort, both in academia and industry, has been invested in P-type colorants over the past three decades because of their potential as molecular switches [11]. Such compounds are converted from one state to another by irradiation with light, remaining so until switched back by other wavelengths. This behavior is of great interest in high technology sectors, but while much time and money has been spent on developing P-type applications, such work has yet to bear commercial fruit. The classes that have been studied most in this connection are discussed next.

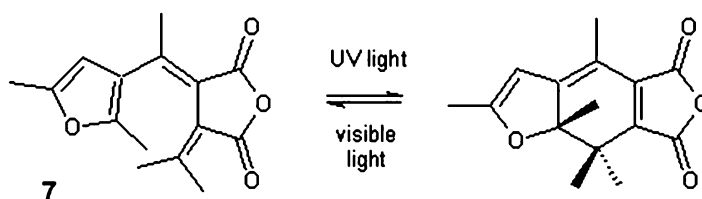
Diarylethenes

Of P-type families of colorants, this class [6, 8, 11] has arguably been subjected to the most intense scrutiny. In contrast to commercial T-type dyes, their photochromism relies upon UV light causing ring-closure rather than ring-opening (see Fig. 5).

Colorant, Photochromic,
Fig. 5 Photochromism of diarylethenes using the dithienylhexafluorocyclopentene class (**6**) as an example



Colorant, Photochromic,
Fig. 6 Photoisomerization of fulgides as exemplified by furylfulgide **7**



With appropriate structural design, the colored cyclized material is essentially thermally stable (owing to a negligibly small reaction rate for thermal reversion) so that the reverse reaction does not occur in the dark. Instead, ring-opening requires the absorption of longer wavelengths of visible light than those that cause activation. Consequently, unlike pyran- and oxazine-derived colorants, diarylethenes do not fade spontaneously once the illumination is removed. A further contrast is that, whereas commercial T-type compounds must be in solution to produce useful photochromism, diarylethenes give photochromic responses in solid form (follow this <http://www.rsc.org/suppdata/cc/b5/b505256d/b505256d.mpg> for a video). The derivatives that have attracted the greatest attention are members of the dithienylhexafluorocyclopentene subclass **6**. A few of them are available on the open market in small quantities. Activated colors can be varied widely by modification of their structure, covering most of the visible spectrum. By linking together different diarylethene skeletons, photochromic molecules have been created that can even be selectively switched between more than one different colored state.

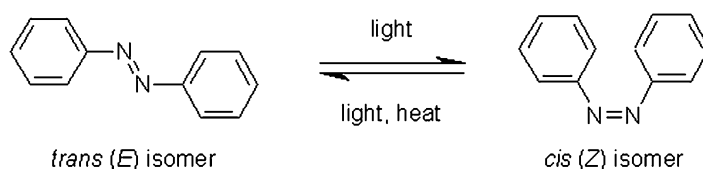
Fulgides

Derivatives drawn from this class [3, 6] have been observed to exhibit various kinds of chromic behavior of which P-type photochromism is one. Like diarylethenes, generation of color involves a ring-closure reaction upon exposure to UV radiation either in solution or as a solid. By judicious choice of structure, transitions from nearly colorless to yellow, to red, and even into the infrared region is possible. Fading is caused by absorption of certain wavelengths of visible light. An example of a dye **7** that has been commercially available is shown in Fig. 6; when activated it changes color from pale yellow to red.

Applications of P-Type Dyes

Fulgides have been used in conventional coloration areas such as textiles and printing inks [8]. However, the greatest interest in P-type photochromic systems stems from their potential use as functional colorants within the fields of optoelectronics, data storage, and nanotechnology [11, 12]. In these contexts, diarylethenes have been the most intensely studied family as they offer a wide scope for the design of molecules whose optical properties can be switched in a controlled manner

Colorant, Photochromic,
Fig. 7 Photochromism of
 azobenzene



between persistent states. Much effort has been expended on trying to exploit this behavior, for example, in developing components for all-optical circuitry, such as switches and logic gates. Optical equivalents to electrical components, perhaps reliant on P-type colorants, are needed if the technology is to replace slower, more power-hungry conventional electronics. Molecules that can be switched optically are also being studied in the field of information technology because in theory they could furnish memory systems with much greater densities than those of current commercial devices.

Another avenue of research for P-type colorants is nanotechnology because of their solid phase photochromism. For example, crystals of dihetarylethenes undergo changes in shape, as well as color, as a consequence of molecular geometry altering during photochromic transitions (follow this http://pubs.acs.org/doi/suppl/10.1021/ja105356w/suppl_file/ja105356w_si_002.avi for a video). Such alteration in particle dimensions could form the basis for light-driven actuators in nanomachinery.

One of the oldest known photochromic systems, azobenzene, has come under increased scrutiny as a means of creating materials with light-sensitive properties [13]. Azobenzene has an element of both T- and P-type character as shown in Fig. 7.

The rate of thermal back reaction can be slowed by modification to structure, creating materials with strong P-type behavior where *trans* to *cis* photoisomerization occurs with one range of wavelengths and *cis* to *trans* with another set [13]. These transformations have been explored for use in optical switching and data storage. The geometry change associated with photoisomerization has also been put to work in creating polymeric materials that undergo reversible photomechanical deformation. Films that

contract upon exposure to UV light and then expand when irradiated with visible light have been used to demonstrate the concept of light-driven motors (follow this http://onlinelibrary.wiley.com/store/10.1002/anie.200800760/asset/supinfo/anie_z800760_ikeda_movie1.mov?v=1 for a video).

Cross-References

► [Dye, Functional](#)

References

1. Corns, S.N., Partington, S.M., Towns, A.D.: Industrial organic photochromic dyes. *Color. Technol.* **125**, 249–261 (2009)
2. Dürr, H., Bouas-Laurent, H.: Organic photochromism. *Pure Appl. Chem.* **73**, 639–665 (2001)
3. Dürr, H., Bouas-Laurent, H.: *Photochromism: Molecules and Systems*. Elsevier, Amsterdam (2003)
4. McArdle, C.B. (ed.): *Applied Photochromic Polymer Systems*. Blackie, Glasgow (1992)
5. Crano, J.C., Flood, T., Knowles, D., Kumar, A., Van Gemert, B.: Photochromic compounds: chemistry and application in ophthalmic lenses. *Pure Appl. Chem.* **68**, 1395–1398 (1996)
6. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds Vol. 1: Main Photochromic Families*. Plenum, New York (1999)
7. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds Vol. 2: Physicochemical Studies, Biological Applications, and Thermochromism*. Plenum, New York (1999)
8. Bamfield, P., Hutchings, M.G.: *Chromic Phenomena: Technological Applications of Colour Chemistry*. The Royal Society of Chemistry, Cambridge (2010)
9. Lokshin, V., Samat, A., Metelitsa, A.V.: Spirooxazines synthesis, structure, spectral and photochromic properties. *Russ. Chem. Rev.* **71**, 893–916 (2002)
10. Hepworth, J.D., Heron, B.M.: Photochromic naphthopyrans. In: Kim, S.-H. (ed.) *Functional Dyes*, pp. 85–135. Elsevier, Amsterdam (2006)

11. Feringa, B.L., Browne, B.R. (eds.): *Molecular Switches*. Wiley-VCH, Weinheim (2011)
12. Irie, M. (ed.): *Photochromism: memories and switches* special issue. *Chem. Rev.* **100**, 1683–1890 (2000)
13. Zhao, Y., Ikeda, T. (eds.): *Smart Light-Responsive Materials: Azobenzene-Containing Polymers and Liquid Crystals*. Wiley, Hoboken (2009)

Colorant, Textile

Andrew Towns

Vivimed Labs Europe Ltd., Huddersfield, West Yorkshire, UK

Definition

Textile colorants impart color to a textile material, usually with a high degree of permanency, as a result of their chemical binding or physical entrapment within or around the textile material. The textile material may be in one of several forms such as fiber, yarn, fabric, garment, etc. Textile colorants are supplied in both solid and liquid forms, for example, as powders, granules, solutions, or dispersions. In certain instances, precursors are applied to textile materials to generate the colorant in situ within the textile.

Textile Dyes and Pigments

Both dyes and pigments are used in the coloration of textiles [1]. The former substances are present in solution at some point during their application, whereas the latter colorants remain insoluble within any vehicle in which they are applied as well as within the textile material itself. Pigments must therefore either be incorporated into textile fibers during their construction (e.g. mass coloration of a polymer followed by its melt extrusion into fibers) or be printed onto a fabric as part of a formulation that contains a binder so that the colorant is physically held to the surface of the textile in a coating. Textile dyes are usually applied from solution although certain types may

be present initially as a dispersion or applied from the vapor phase. The mechanism by which dyes remain within a textile depends on the particular colorant type. Retention may rely on intermolecular forces operating between dye and fiber following adsorption onto and/or dissolution within the polymer, formation of covalent bonds between the dye and the fiber, or entrapment of colorant particles within the textile by deposition of an insoluble form of the dye.

Natural Textile Colorants

Prior to the synthesis of picric acid in the eighteenth century as a yellow dye for silk, all textile colorants were obtained directly from natural sources, such as plants, insects, and shellfish [1, 2]. These natural colorants began to be superseded by synthetic dyes and pigments during the second half of the nineteenth century since the latter products offered a wider and brighter gamut of color as well as greater economy and fastness. While there has been a revival of interest recently in natural textile colorants driven by perceptions of renewable sourcing and low environmental impact, they are not suited to industrial use and offer only a limited color gamut and display moderate levels of fastness at best. In addition, natural dyes often require the use of a fixative, known as a *mordant*, to achieve satisfactory permanency; traditional metallic mordants are environmentally unfriendly. Textile dyes that were originally obtained from natural sources, such as Indigo, are more efficiently obtained by chemical synthesis [3].

Synthetic Textile Colorants

The vast majority of textile colorants are synthesized chemically on an industrial scale [3, 4]. Over a million tonnes are produced globally each year. Tens of thousands of colorants have been marketed since the first commercially successful synthetic textile dye, *Mauveine*, was manufactured in the late 1850s [1, 5]. The financial rewards from this particular colorant and its

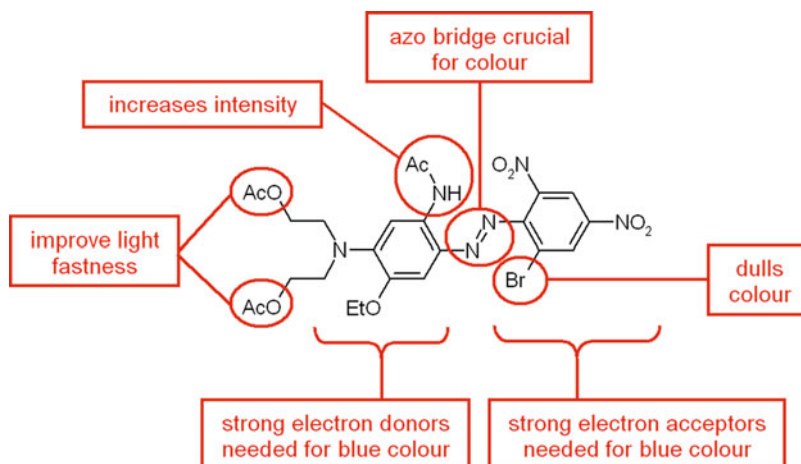
contemporaries spurred on much endeavor into making synthetic dyes. The genesis of the modern chemical industry, which dwarfs the colorant sector and now manufactures products as diverse as plastics, pharmaceuticals, and drugs, lies in the attempts of mid-nineteenth-century chemists to prepare textile dyes and other colorants. These efforts initially involved a trial-and-error approach since little was known about molecular structure, and dyes often reached the market as mixtures of different compounds with their discoverers having little idea of their composition. However, a more systematic approach to textile colorant research, principally led by the German dye industry, resulted in a better understanding of both chemistry and color–structure relationships. At the beginning of the twentieth century, Germany manufactured 85 % of the world's synthetic dyes with 10 % being produced in Britain, France, and Switzerland. A century later, manufacture is concentrated in Asia, particularly China and India, because of a lower cost base, although some of the biggest suppliers remain headquartered in Europe. The global market for textile colorants is worth several billion dollars [6]. The textile industry is the largest consumer of dyes and organic pigments. Some of the biggest-selling textile dyes are truly commodities, being produced in volumes of over 1,000 tonnes each year for supply at just a few dollars per kilogram or even less.

Nomenclature and Composition of Commercial Textile Colorants

Traditionally, the trade names which manufacturers give to colorants are typically comprised of three parts [3]. The first element is a brand name, often incorporating elements of the producer's name, the type of colorant, and/or its intended use. The next part indicates general color, occasionally with modifiers to highlight shade or application characteristics to which the manufacturer wants to draw attention. The last part of the name is made up of a code of letters and numbers for further differentiation of the colorant from others – these may be related to

color, application properties, as well as strength. For example, the now-defunct manufacturer, *Holliday Dyes and Chemicals*, marketed the dye: *Polycron Yellow C-5G 200%*. “Polycron” was the name shared by the company's disperse dyes developed solely for the coloration of polyester; “C” denoted the disperse dye application category into which this colorant fell; “5G” indicated that it had a relatively greenish hue; and the suffix “200%” signified that its tinctorial strength was double that of the market norm (“100%”), i.e., twice the amount of active dye was present per unit mass of colorant. Many other companies employ this system of nomenclature for their colorants. However, reliance on commercial names alone will not necessarily enable informed colorant selection (i.e., which colorant to use for a particular substrate and application technique) nor show which textile colorants in the market are equivalent. Fortunately, many manufacturers link their products to the *Color Index* (CI) generic naming system which assists users in making sense of the vast array of textile colorants that are commercially available. For instance, Polycron Yellow C-5G 200% has a CI Generic Name assigned to it, CI Disperse Yellow 119, thereby allowing users to identify equivalent products. While colorants with the same CI Generic Name ought to contain the same main colored compound, they may not be exactly equivalent. Variations in impurity content, whether inadvertent or through deliberate inclusion of shading components, may affect color or other application properties. Differences in physical form can affect performance. Members of several types of dye class typically include substantial quantities of additives, present as processing aids, such as dispersing agents or buffers to stabilize pH. Another source of variation is the presence of a diluent, for example salt or dextrin, which is added to standardize certain dye types to a desired tinctorial strength. A further potential complication is that some textile colorants do not have a CI Generic Name, either because they are mixtures of colored components as is often the case with navy and black dyes, or simply because one has not been disclosed or assigned.

Colorant, Textile,
Fig. 1 Structural features
 of CI Disperse Blue 79



Textile Colorant Structure

The constitution of commercial dyes and pigments used for textile coloration can be described as lean: each part of a colorant molecule has one or more purposes as will be illustrated in some of the following sections. These functions may be related to adjustment of color (e.g. hue, intensity, brightness), physical character (e.g. solubility, crystal structure, volatility), dyeing behavior (e.g. substantivity, leveling), fastness (to, e.g. light, heat, moisture), and so on, although there will be occasions when substituents are present merely for convenience or cost [5]. Often, dye design involves an element of compromise as certain properties will have a degree of mutual exclusivity, e.g. rapid dyeing and good leveling behavior at the expense of good wash fastness. The most general structural classification centers on the key molecular features of a colorant that gives rise to its color. Azo derivatives are the largest such class of textile colorants, although there are many others of significance.

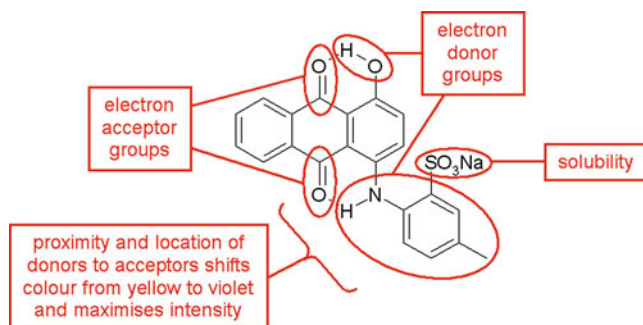
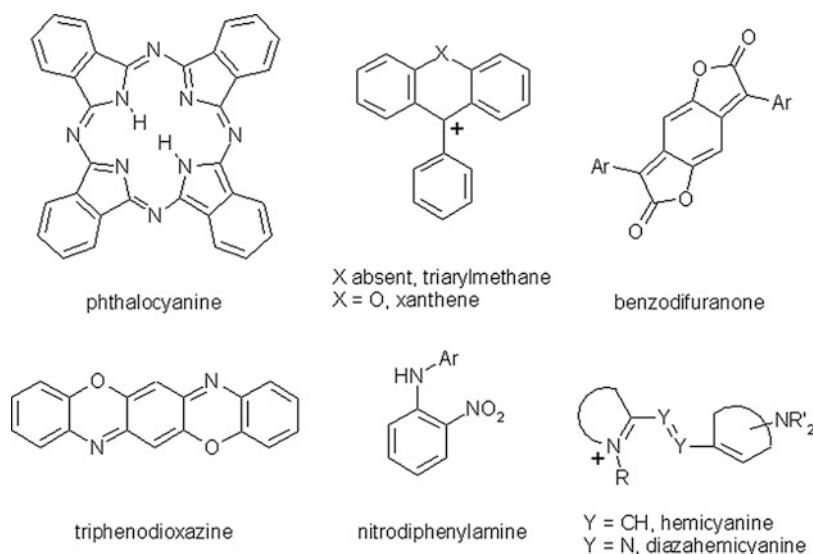
Azo Textile Colorants

This class is defined by the presence within the colorant molecule of an arylazo group of general structure ($Ar-N=N-X$) where X is most commonly another aryl ring [3–7]. In many industrial colorants, the azo function exists exclusively in a more energetically stable hydrazone form ($Ar-NH-N=X$) instead [1]. Coverage of the

whole visible spectrum is possible using commercial dyes containing just a single azo or hydrazone bridge, although many important textile dyes contain two or more such groups [3, 7]. Azo colorants make up around half the number of textile dyes and pigments available. This dominance lies in their economy, robustness, and versatility. Not only are they relatively inexpensive to manufacture, but generally azo derivatives also have good tinctorial strength and so tend to be economical compared to other colorant types. The class has also offered manufacturers much scope to adjust structure, enabling them to readily modify properties including shade, solubility, dyeing behavior, as well as fastness [5]. Figure 1 illustrates how structure can be broken down into fragments with different purposes using the highest-volume navy blue dye for polyester (CI Disperse Blue 79) as an example.

Carbonyl Textile Colorants

The color and application properties of this group of dyes and pigments are dependent on carbonyl functions ($>C=O$) [1, 3, 7]. A major subclass is based on 9,10-anthraquinone, because it is a source of red to blue, as well as green, dyes of high brightness and fastness [5]. However, economic considerations restrict their use – they are generally more expensive to manufacture than azo derivatives, while their relatively low tinctorial strength compared to other classes means that more dye has to be used to achieve a particular

Colorant, Textile,**Fig. 2** Structural features of CI Acid Violet 43**Colorant, Textile,****Fig. 3** Some important textile colorant structural types

depth. An example of a commercial anthraquinone dye is shown in Fig. 2.

This group also encompasses one of the oldest types of textile dyes, *indigoid*, and one of the newest, *benzodifuranone*. The latter is shown in Fig. 3 and is a commercially useful source of intense bright red dyes [3].

Other Textile Colorant Structural Types

There are numerous other structural classes of textile dyes [1, 3, 5, 7] and pigments [4, 5], each occupying its own application niches. A few examples are presented in Fig. 3. The *triarylmethane* class dates all the way back to Mauveine, although the importance of such purple, blue, and green dyes has diminished with the development of alternative chemical types of superior fastness [5]. The same is true of yellow

to red fluorescent *xanthene* dyes as textile colorants. However, the *phthalocyanine* class remains as much prized as ever since its commercialization in the first half of the twentieth century [1]. This colorant type furnishes robust and tinctorially strong bright blue to green dyes and pigments with diverse uses in the textile arena, especially when they are in the form of copper complexes [3, 4]. Other classes occupy more specific sectors in terms of color and/or application. For example, *triphenodioxazines* are exploited as intense bright blue dyes for natural fibers, while *nitrodiphenylamines* are cost-effective yellow colorants of good photostability for synthetics [3]. Polymethine dyes are a broad class of chromophore of which certain subclasses are employed in textile coloration: two such types, *hemicyanines* and *diazahemicyanines*, are shown

in Fig. 3. Both families are useful for intense bright reds, while the latter also supplies blue colorants [3].

Textile Colorant Application Classes

Although knowledge of the molecular structure of a textile colorant is essential for a manufacturer, the user is concerned more with its method of application and performance [3, 7]. Modes of dye and pigment use include mass coloration, exhaustion dyeing, thermofixation, and printing by screen, inkjet, or sublimation – each calls for a different mix of behaviors, while commercial demands will dictate economic and quality criteria [8]. A useful way of grouping textile colorants into sets with very broadly similar characteristics in terms of color, usage, fastness, and economy is by CI application class. The most important

application classes of textile colorant are listed in Table 1 and described in more detail below [1, 3, 5, 7–9].

Acid Dyes

This type of anionic colorant is commonly used to dye and print natural (wool, silk) and synthetic (nylon) polyamides [3, 7–9]. The class is named after the acidic (pH 2–6) dyebaths used for dye application. Lowering pH increases the concentration of cationic ammonium groups within these substrates, enhancing their attraction for anionic dyes. While anthraquinone and triarylmethane derivatives are significant for violet, blue, and green colorants, azo compounds are by far the most important structural class. A substantial proportion of acid dyes are metal complexes that comprise one or two dye ligands; of particular importance are 1:2 chromium/dye ligand complexes as these furnish dull, deep shades of high wet fastness and high photostability. Dyeing performance and fastness properties are readily modified by adjusting dye hydrophobicity (see Fig. 4).

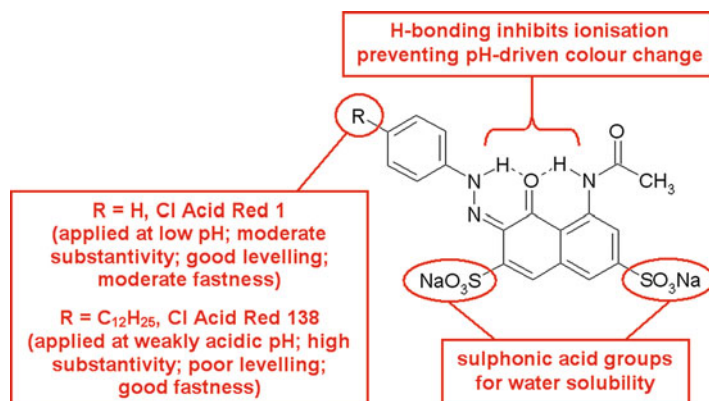
Colorant, Textile, Table 1 Major textile colorant application classes

Class	Principal textile substrate(s)
Acid	Wool, silk, nylon, modified polyacrylonitrile
Azoic	Cotton and other cellulose, acetates
Basic	Polyacrylonitrile, modified nylon, and polyester
Direct	Cotton and other cellulose, polyamide
Disperse	Polyester, acetates, nylon, polyacrylonitrile
Pigment	Cotton, polyester
Reactive	Cotton, wool, silk, nylon
Sulfur	Cotton and other cellulose
Vat	Cotton and other cellulose, polyamide

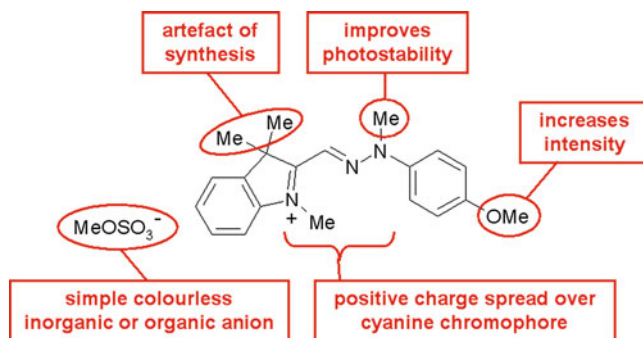
Azoic Colorants

These colorants are employed predominantly on cellulosic fibers [7, 8]. They are insoluble azo compounds that are synthesized within the textile substrate during the dyeing process from soluble precursors. The colorants generated are physically trapped as solid particles within fiber pores, so dyeings display good wash fastness and photostability. There are numerous variations in technique and materials, but the usual method is to

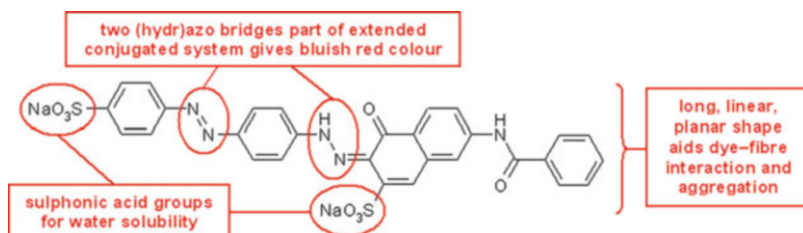
Colorant, Textile,
Fig. 4 Structural features
of CI Acid Red 1 and CI
Acid Red 138



Colorant, Textile,
Fig. 5 Structural features
 of CI Basic Yellow 28



Colorant, Textile,
Fig. 6 Structural features
 of CI Direct Red 81



apply a naphthol coupling component followed by a diazo component, which reacts with the coupler to produce a non-ionic colorant. The diazo component is an aromatic amine, which is diazotized as part of the application process or supplied as stabilized pre-diazotized materials, for example, aryldiazonium tetrafluoroborate salts (i.e., $\text{ArN}_2^+\text{BF}_4^-$). Azoic textile colorants have largely been sidelined industrially owing to the greater convenience and economy of other systems but are still of commercial interest in the red shade area.

Basic Dyes

Members of this class are used primarily for coloration of polyacrylonitrile, either by exhaustion dyeing or during fiber production [3, 8, 9]. They are often referred to as cationic dyes because their molecular structures feature a positive charge as shown, for example, in Fig. 5. This charge may be pendant (i.e. localized but isolated from the chromophore) or delocalized, forming part of the chromophore itself [7]. In either case, the dye binds to the substrate by electrostatic interaction. Triarylmethane colorants are significant in the blue and green sectors, while azo and cyanine-type dyes dominate the yellow to violet areas

[5]. All can furnish bright intense shades of high wet and light fastness.

Direct Dyes

This class is so named because its members can be applied to cotton and other cellulosic fibers without the need for mordants [3, 7, 8]. The crucial molecular features of direct dyes are (a) long, linear planar geometries and (b) multiple substitution with negatively charged sulfonic acid groups as depicted in Fig. 6. Sodium chloride or sulfate is added to the dyebath to enhance dye adsorption. Dye geometry enables close approach to the polymer chains of the substrate. Intermolecular forces that operate at short distances can therefore become significant, aided by the large surface area of the molecule. While diffusion rates of dye into fiber tend to be low because of their large molecular size and propensity to aggregate, diffusion rates out of the dyed cellulose during washing are also small so that wash fastness is moderate. The class is dominated by azo derivatives: disazo dyes for yellows to blues and polyazo colorants for blue, green, and neutral shades. Direct dyes are used for their economy where wash fastness is not critical [7, 8].

Disperse Dyes

These non-ionic colorants are hydrophobic like the synthetic textile substrates to which they are applied [8, 9]. Originally developed for cellulose acetate fibers, their principal use is the coloration of polyester: the importance of this textile material means that disperse dyes have become one of the two most important types of textile colorant [3, 7]. They have only sparing water solubility and so are applied as fine dispersions in water (apart from when they are used in transfer-printing ink films). Dye particle sizes are typically distributed in the range of around 0.1–1 μm diameter for supply either in solid or liquid dispersion form. These forms are achieved by milling in the presence of dispersing agent, usually anionic polymeric materials such as lignosulfonates or arylsulfonic acid condensates, to inhibit reaggregation and maintain dispersion stability. During exhaustion dyeing, a small proportion of colorant is in aqueous solution: it is from this phase that dye is adsorbed onto the textile and diffuses within it. Colorant lost from the aqueous phase is replenished by dissolution of dye remaining in suspension. As disperse dyes must have relatively compact structures to enable them to diffuse satisfactorily within hydrophobic textiles, commercial ranges therefore consist mainly of monoazo and anthraquinone derivatives: the latter are important in the bright red and blue sectors, but the former dominate the rest of the spectrum. Several other chemical dye classes are employed, but these tend to occupy niches, such as yellow nitro-diphenylamines. Brown and black disperse dyes are usually formulated using mixtures of azo dyes because of the technical and economic difficulties in creating single-component colorants of small molecular size with the desired hue.

Pigments

While used primarily for the coloration of media other than textiles, pigments can be printed onto fabrics. Examples of organic pigments used as textile colorants in this way include yellow acetanilide and red naphthol monoazo derivatives as well as blue copper phthalocyanines [4]. The insoluble colorant is incorporated into a paste, which is applied to the textile fabric by a printing

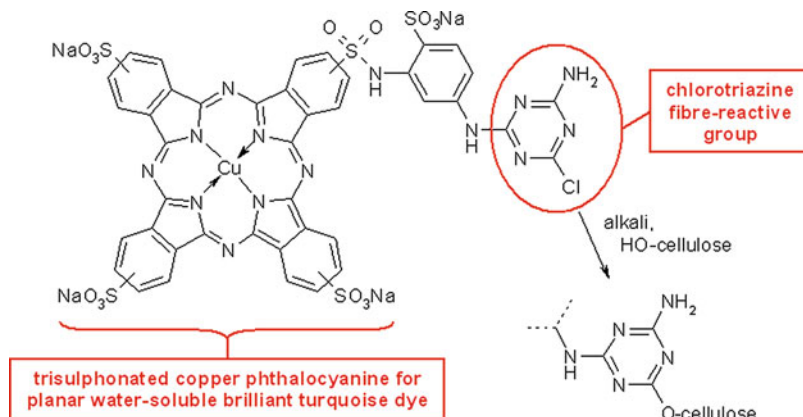
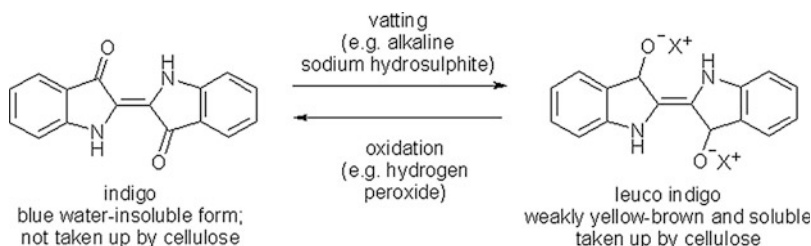
process (e.g., screen, roller, etc.). After curing of the paste, normally achieved by drying, the pigment is physically bound to the fabric. Pigments are also employed in the mass coloration of many types of textile fibers in which the colorant is dispersed in a solution or melt of the polymer prior to fiber formation so that the pigment particles become trapped within the fibers as they are produced. The pigment types mentioned above as well as numerous other chemical classes are employed in this capacity.

Reactive Dyes

These colorants are the only ones that form covalent bonds with the textile to which they are applied [3, 5]. While they can be applied to wool, nylon, and silk, reactive dyes for cotton have achieved the greatest commercial success, becoming one of the two most important application classes [7, 8]. After exhaustion or printing onto the fiber, the dye reacts with ionized cellulosic hydroxy groups, furnishing excellent wash fastness. Reactive colorants resemble direct dyes with one or more attached fiber-reactive groups. This geometry favors adsorption and reaction with cellulose, although the reliance on covalent bonding allows greater flexibility in choice of chromogen. The vast majority of reactive dyes are azo derivatives apart from the bright blue and green shade areas where anthraquinone-, phthalocyanine-, and triphenodioxazine-based colorants among others are important [5]. Commercially important reactive groups bond to the substrate either by addition, e.g. vinylsulfone anchors, or by substitution, such as through halogenotriazine functions like that illustrated in Fig. 7.

Sulfur Dyes

These colorants are prepared by heating aromatic compounds with sulfur or a sulfur compound to produce compounds of large molecular size that contain disulfide ($-\text{S}-\text{S}-$) linkages between the aromatic components [1, 3, 5]. Dyes of this kind are chemically complex and, in the majority of cases, their structure is unknown. Their application resembles that of vat dyes (see below) [7]. The initially water-insoluble dyes are reduced

Colorant, Textile,**Fig. 7** Key features of CI Reactive Blue 15**Colorant, Textile,****Fig. 8** Key features of CI Vat Blue 1 (Indigo)

in the presence of alkali to a water-soluble “leuco” form containing water-solubilizing thiolate ($-S^-$) groups which diffuse into the fiber in the presence of electrolyte. At the end of dyeing, these groups are then oxidized in situ to regenerate the insoluble disulfide which remains trapped in the fiber. Sulfur dyes are widely used for the production of dull deep shades of reasonable wet fastness because of their low cost [8].

Vat Dyes

Members of this class, which are applied mainly to cotton and other cellulosic fibers, generally provide the highest all-round fastness but are expensive, so they tend to be the class of choice when durability is a priority, e.g. production of upholstery [3, 7, 8]. Key aspects of vat dyes are water insolubility and the presence of one or more pairs of carbonyl groups. The class derives its name from the crucial operation during dyeing and printing called “vatting” which involves reducing the carbonyl functions of the finely dispersed colorant to produce a water-soluble “leuco” form (see Fig. 8). Once the leuco form of

the dye has been applied to the textile, oxidation is then undertaken to regenerate the colored water-insoluble form of the dye within the substrate where it becomes physically trapped. Fastness to washing and light are therefore excellent once loose colorant has been removed by a soaping step. The class is dominated by anthraquinonoid and more complex polycyclic derivatives although indigoid colorants are also important, especially the parent compound Indigo which is the highest-volume vat dye (see Fig. 8) [5].

Future Prospects

The enormous variation in both the chemical and physical characteristics of textiles has forced the emergence of a diverse array of coloration technologies. As a consequence, many different types of colorant have been commercialized for application to textiles. Pigments are used in the printing of textiles or in the mass coloration of polymer intended for fiber manufacture, while dyes are applied industrially in different ways ranging

from exhaustion dyeing to transfer printing. The wide spectrum of techniques available to a dyer or a printer has only been made possible through the development of colorants that possess chemistries and physical properties tailored for application to specific types of textiles. There is no such thing as a “universal” colorant that can be applied satisfactorily to all textiles for all intended outlets nor is one likely to materialize in the near future. It is widely accepted that the chances of new application classes or structural types being introduced are remote. Nevertheless research along these lines continues alongside incremental work to improve on the technical properties, economy, and environmental impact of existing textile colorants.

Cross-References

- [Coloration, Mordant Dyes](#)
- [Colorant, Natural](#)
- [Coloration, Textile](#)
- [Dye](#)
- [Dye, Metal Complex](#)

References

1. Zollinger, H.: *Color Chemistry: Synthesis, Properties, and Applications of Organic Dyes and Pigments*. VCH/VCH, Zürich/Weinheim (2003)
2. Hofenk De Graaff, J.H. (ed.): *The Colourful Past: Origins, Chemistry and Identification of Natural Dyestuffs*. Abegg-Stiftung and Archetype Publications, Riggisberg/London (2004)
3. Hunger, K. (ed.): *Industrial Dyes*. Wiley-VCH, Weinheim (2003)
4. Herbst, W., Hunger, K.: *Industrial Organic Pigments*. Wiley-VCH, Weinheim (1997)
5. Shore, J. (ed.): *Colorants and Auxiliaries*, vol. 1. SDC, Bradford (1990)
6. Bamfield, P., Hutchings, M.G.: *Chromic Phenomena: Technological Applications of Colour Chemistry*. The Royal Society of Chemistry, Cambridge (2010)
7. Waring, D., Hallas, G. (eds.): *The Chemistry and Application of Dyes*. Plenum, New York (1994)
8. Richards, P.R.: Dye types and application methods. In: Best, J. (ed.) *Colour Design: Theories and Applications*, pp. 471–496. Woodhead Publishing, Cambridge (2012)
9. Burkinshaw, S.M.: *Chemical Principles of Synthetic Fibre Dyeing*. Blackie, Glasgow (1995)

Colorant, Thermochromic

- [Colorant, Thermochromic](#)

Colorant, Thermochromic

Mary Anne White and Alex Bourque
Department of Chemistry, Dalhousie University,
Halifax, NS, Canada

Synonyms

[Colorant](#), [halochromic](#); [Colorant](#), [thermochromic](#); [Dye](#), [functional](#)

Definition

Thermochromism is the property of temperature dependence of the electronic absorption spectrum of a material, resulting in a color that depends on temperature. Formerly, the term thermochromism, also known as thermochromatism, was reserved for isolated compounds and their solutions. However, the advancement of the field has led to a broadening of this definition; the term thermochromism now can also be used to describe multicomponent mixtures that are able to change color in response to changes in temperature. For both isolated compounds and mixtures, reversibility of the color change generally is regarded as a necessary condition for thermochromic behavior.

Interestingly, thermal copy and receipt paper, which have been the most commercially important thermally responsive color-changing products for the past several decades, undergo an irreversible coloring reaction and therefore do not fall under the narrow technical definition of thermochromism. High-technology products such as thermal copy paper, which make use of multicomponent thermochromic *mixtures*,

become thermochromic due to the thermal initiation of a secondary color-changing reaction. *Halochromism*, the color change of a compound in response to changing pH, is the temperature-dependent auxiliary process occurring in thermal copy paper which causes the color change.

Introduction

Thermochromic compounds change color in response to temperature changes [1]. Most thermochromic compounds are organic in nature and undergo thermally activated chemical modifications which give rise to the color change. In many cases the chemical modification is a result of *tautomerism*. Tautomerism refers to reversible structural isomerism that consists of multiple steps usually involving bond cleavage, molecular reconfiguration, and subsequent bond reformation [2]. Thermochromic behavior can be observed in a wide variety of isolated compounds. Common organic thermochromic compounds include crowded ethenes, Schiff bases, spiro heterocycles (e.g., spiropyrans, spironaphthalenes, etc.), and macromolecular systems including liquid crystals and polymeric materials [3]. There are comparatively fewer examples of inorganic thermochromism. However, vanadium (IV) oxide (VO_2) has recently garnered much attention from researchers as a potential smart coating material due to its thermally tunable infrared and near-infrared absorption spectrum [4].

Today, significant research efforts in the academia and industry focus on the development of new technologies and devices based on multicomponent thermochromic mixtures including smart coatings, erasable printing media, and temperature sensors. This article provides a brief review of important examples of thermochromic compounds, followed by a description of the most recent advancements in the field with an emphasis on applications for new high-technology materials. Advanced *thermochromic materials* take advantage of the growing field of *functional dye* chemistry, and a few examples are presented.

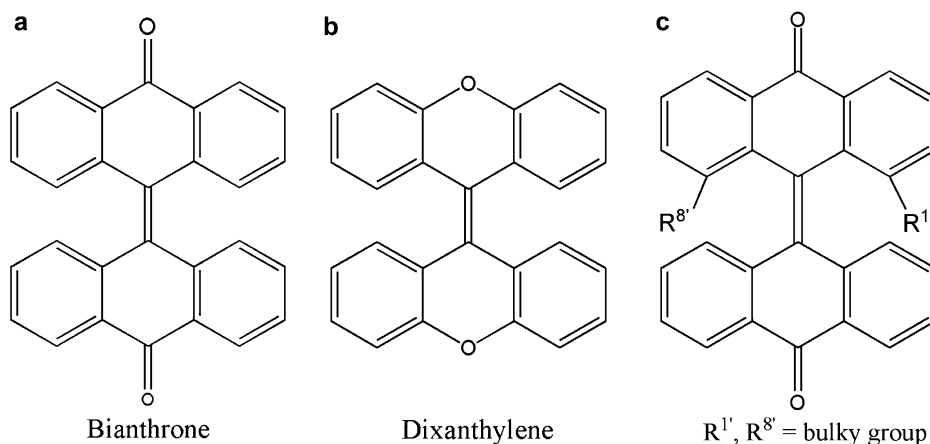
Organic Thermochromism

Thermochromic behavior in organic compounds is often caused by thermally activated chemical rearrangements, i.e., *thermal tautomerism*. Some of the more common examples of tautomerism include acid-base reactions, keto-enol rearrangement, and lactim-lactam equilibria. Tautomerism is usually influenced by changes in temperature and solvent properties such as composition, polarity, and pH, and thermally activated tautomerism can lead to thermochromism.

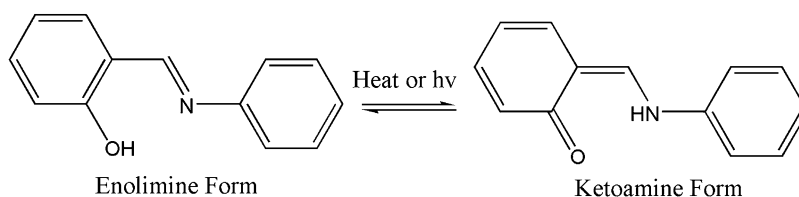
Bianthrone and Crowded Ethenes

One of the first examples of reversible thermochromism observed in organic compounds was that of bianthrone [1]. The structures of bianthrone and some of its analogues are shown in Fig. 1. Bianthrone's structure can be considered in terms of the planarity of each anthrone moiety (i.e., the upper and lower halves of the molecule, as shown in Fig. 1). At room temperature, the anthrone moieties are curved such that the aromatic rings within each anthrone moiety are not coplanar. When the temperature is increased, the central carbon-carbon double bond expands. This slight bond expansion allows the two anthrone moieties to rotate with respect to one another such that the dihedral angle across the double bond approaches 90° . When this occurs, the previously bent anthrone moieties each take a more planar structure as the steric repulsion created by the proximity of the other anthrone moiety is reduced by rotation about the central bond [5].

The increased planarity at higher temperature permits π -conjugation to extend more effectively across each anthrone moiety, decreasing the HOMO-LUMO gap and concomitant electronic absorption energy and thereby giving rise to a change in color of the compound with a change in temperature. Bianthrone is yellow in the solid (absorbing violet light at low temperature) and green (absorbing red light) in the melt. Substituents play a strong role in determining if these crowded ethenes will be thermochromic. Dixanthylene is colorless in the



Colorant, Thermochromic, Fig. 1 (a) The structure of bianthrone and (b) related compound dioxanthylene. (c) Bulky groups at the 1, 1', 8, and/or 8' positions would result in non-thermochromic compounds



Colorant, Thermochromic, Fig. 2 Tautomeric equilibrium between enol-imine and keto-enamine forms observed for thermochromic Schiff bases [7]

solid and green in the melt. Bulky groups at the 1' and 8' positions would cause excessive steric repulsions and prevent the formation of the bent anthrone state, which is required for the color-changing process [3].

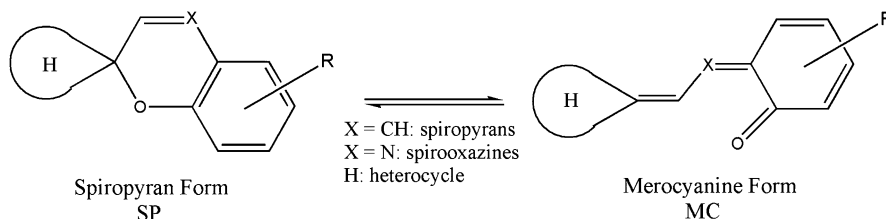
Schiff Bases, also known as Salicylidene-Anilines

Thermochromic Schiff bases, also known as imines or azomethines, are formed by the condensation of an aromatic amine with either an aldehyde or ketone. The simplest compound in this class, salicylidene-aniline, is generated by the reaction of salicylaldehyde with aniline. The thermochromic Schiff bases exist in a state of equilibrium between two forms, the *enol-imine* and *keto-enamine* forms. For salicylidene-aniline, the enol-imine form is more stable at room temperature (Fig. 2). Increased temperature allows an intramolecular tautomerization to occur: the

proton from the hydroxyl oxygen migrates to the imine nitrogen. Due to the rearrangement of the π -electrons, the keto-enamine form has more extended π -bonding than the parent enol-imine, which gives rise to a change in color [6].

Substituent effects play a very important role in this system and define which tautomeric form of the enol-imine-keto-enamine equilibrium dominates. To switch between forms, thermal energy sufficient to exceed the activation energy of the tautomeric reaction must be added to the system. A recent review of the Schiff bases by Minkin et al. demonstrates the vast variability in this family of compounds, where simple modifications and ring substitutions can push the enol-imine-keto-enamine equilibrium in either direction [7].

Note that photons (light) also can be used to provide sufficient energy to initiate tautomerism. In that case, the process is considered to be *photochromic*. Thermochromic and photochromic properties in this class of compounds were long thought to be mutually exclusive. However,



Colorant, Thermochromic, Fig. 3 Schematic view of the tautomeric equilibrium between a ring-closed spiropyran (SP) and a ring-open merocyanine (MC) form as observed for spiro compounds [3]

recent studies have indicated that salicylidene-anilines are almost always thermochromic in the solid state and are occasionally also photochromic [8].

Spiro Compounds, Including Spiropyrans

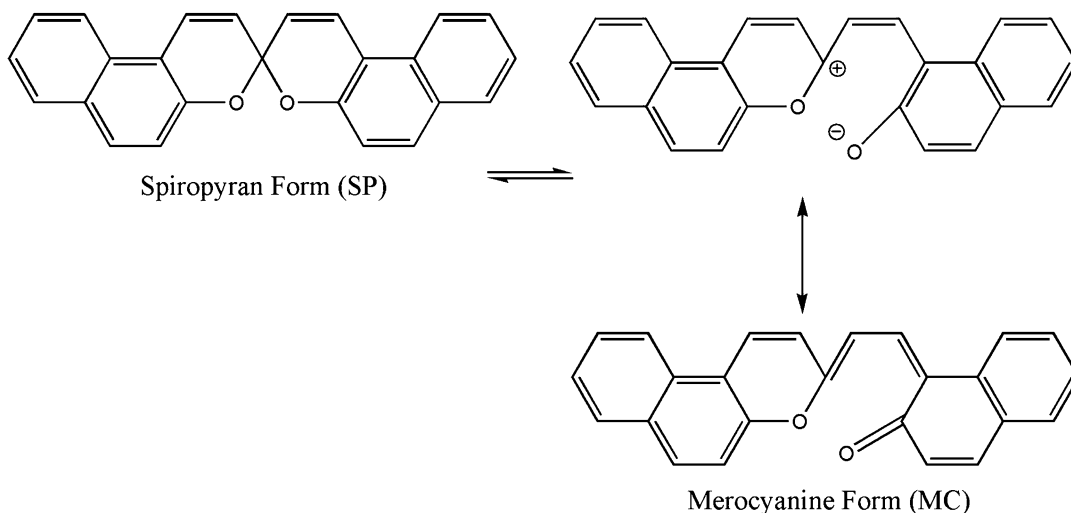
Spiro compounds are arguably the most important class of compound used in thermochromic applications; halochromic triarylmethane and fluoran dyes are widely used as colorants in multicomponent thermochromic mixtures (e.g., in thermal receipt paper and erasable printing media). This class includes spiropyrans, spironaphthalenes, spirooxazines, and fluoran and triarylmethane dyes. Spiro compounds, so named for the “spiro” central sp^3 tetrahedral carbon center that all members share, are subject to numerous tautomeric equilibria including lactim-lactam, acid-base, and the aforementioned enol-keto equilibria. These equilibria result in chemical modifications that have significant impact on the electronic structure and, subsequently, on the color of these compounds [9]. Many *functional dyes* belong to this category of compounds: some of the more important examples are the triarylmethane dyes (e.g., crystal violet lactone, CVL) and the fluoran dyes.

The spiropyrans form one of the longest-known and best characterized classes in this group. Figure 3 shows an equilibrium that is observed commonly in this class of compounds. The spiro center is located within the pyran ring which, upon excitation with thermal energy, can undergo a ring-opening reaction that alters the

electronic structure. As a result of the disruption of the spiropyran (SP) form, the molecule undergoes electronic rearrangement giving the merocyanine (MC) form, which is deeply colored (e.g., violet, red, blue) for nearly all members of this family [1].

An important structural aspect of this class of compounds is that in the uncolored spiropyran form, two or more aromatic moieties are segregated from each other by the central spiro carbon atom. After the tautomeric rearrangement to the more planar merocyanine form, the formerly segregated π -electronic domains are able to form resonance structures which extend across the entire molecule. This delocalization of the π -electronic structure lowers the HOMO-LUMO gap (which would be in the UV typically for the ring-closed structure) and gives rise to the formation of intense color in these compounds. The mechanism for this process is shown in Fig. 4 for di- β -naphthopyran, which was first studied by Dickinson [10]. An important observation is that ring opening leads to a zwitterion, which has significant implications concerning the solubility of the ring-opened merocyanine form.

Ring-opened spiropyrans usually adopt the quinoid structure as shown in the bottom right of Fig. 4. However, if the R-group on the pyran moiety (see Fig. 3) is a strong electron-withdrawing group, the negative charge on the oxygen in the zwitterionic form will be stabilized, allowing the molecule to have zwitterionic character. Substitutive modifications to either of the aromatic functionalities in spiropyrans can significantly modify the thermochromic properties by stabilizing the charges formed in the ring-opened configuration.



Colorant, Thermochromic, Fig. 4 Mechanism of the ring-opening reaction in the spiropyran di- β naphthopyran. Cleavage of the spiro carbon-oxygen bond yields a

structure with greater planarity and extended π -conjugation. The compound becomes colored due to this extended conjugation [3]

Colorant, Thermochromic, Table 1 Structure and thermochromic temperature ranges of some cholesterol esters [12]

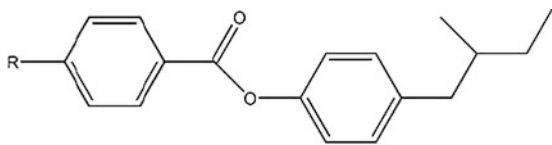
R	Temperature Range of Color Change
CH_3CO_2 to $\text{C}_8\text{H}_{13}\text{CO}_2$	94 – 113 °C
CH_3OCO_2 to $\text{C}_6\text{H}_{13}\text{OCO}_2$	94 – 108 °C
$\text{C}_7\text{H}_{15}\text{OCO}_2$ to $\text{C}_9\text{H}_{19}\text{OCO}_2$	77 – 81 °C

Liquid Crystals

A well-known example of thermochromic behavior which results from macromolecular interactions is the thermochromic liquid crystals. Thermochromic liquid crystals can be found in products as diverse as mood rings, “stress testers,” warning indicators, and thermometers. Thermochromic effects have been exploited in many different types of liquid crystals, but only

two basic series of liquid crystals have found widespread use in thermochromic applications. The esters of cholesterol (Table 1) were first studied by Reinitzer and led to the identification of the liquid crystalline phase [11]. The ester derivatives of (S)-4-(2-methylbutyl)phenol (Table 2) form the basis of the majority of today’s thermochromic liquid crystal devices.

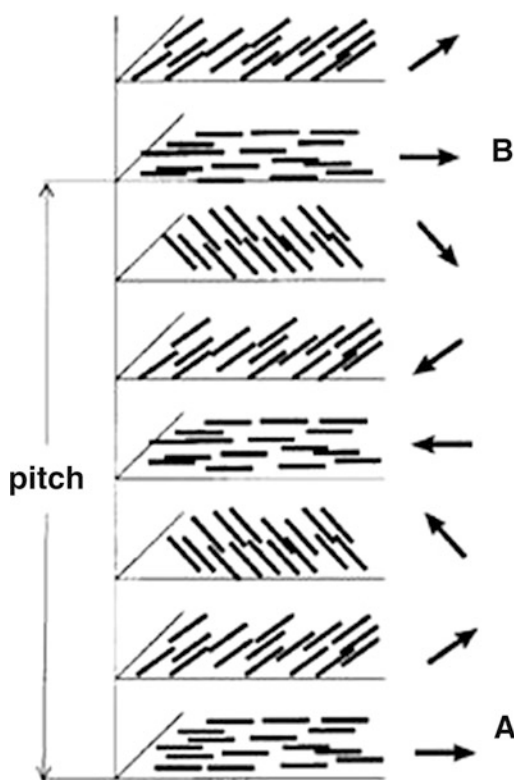
Thermochromic liquid crystals take advantage of the special optical properties of the chiral

Colorant, Thermochromic, Table 2 Structure and thermochromic temperature ranges of some 2- and 3-ring 2-methylbutyl phenol esters [12]


R	Temperature Range of Color Change
C ₅ H ₁₁ & C ₈ H ₁₇	-5 - 0 °C
C ₆ H ₁₃ O to C ₁₀ H ₂₁ O	35 - 50 °C
C ₅ H ₁₁ -Ph to C ₇ H ₁₅ -Ph	130 - 140 °C

nematic phase (abbreviated Ch and/or N*). The nematic phase is formed by calamitic (“rod-shaped”) liquid crystals and demonstrates only orientational order; this phase lacks long-range positional order [13]. The angular distribution of the long molecular axes of the calamitic molecules is central to the special optical properties of chiral nematic phase liquid crystals. The most probable direction of the long molecular axes defines the *director* which coincides with the principal optical axis of the uniaxial phase. In the chiral nematic phase, the director spirals about a helical axis. The pitch length, p , corresponds to the distance along the helical axis required for the director to make a full rotation about the helical axis. The pitch length can be on the order of a few hundred nanometers, i.e., the wavelength of visible light [12]. This effect is shown schematically in Fig. 5 [14].

Light reflected by the layer in the chiral nematic phase at location A (see Fig. 5) can constructively interfere with light reflected from the layer at position B (see Fig. 5) if the extra distance traveled (layer A compared with layer B) is an integer number of wavelengths of light. This phenomenon is analogous to Bragg reflection in layered crystalline solids. In such a way, chiral nematic phase liquid crystals act as a diffraction grating, or, more precisely, a monochromator. Temperature variations in the sample can cause the pitch length to change via thermal expansion, giving rise to variations in the wavelength of light that is constructively reflected (aka selective reflection). An important practical consideration

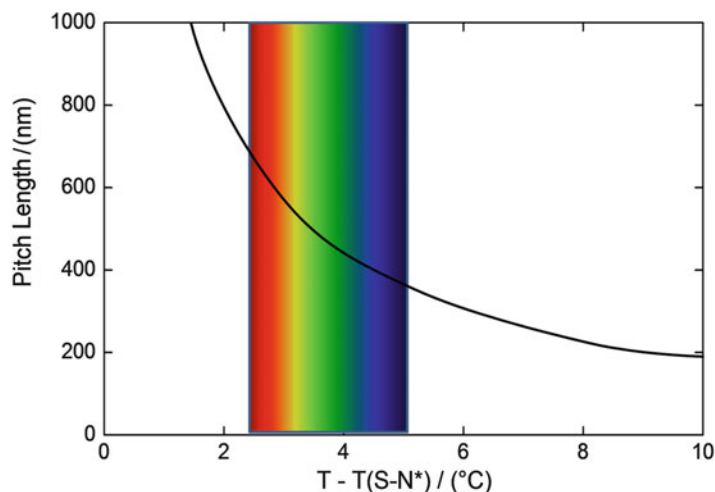


Colorant, Thermochromic, Fig. 5 Pitch length corresponds to the distance required for the director to make a complete rotation about the central helical axis, from layer A to layer B (Reproduced with permission from *Journal of Chemical Education*, 1999, 76(9), 1201-120. Copyright (1999) American Chemical Society)

arises from this selective reflection; light that is not reflected by the liquid crystal must be transmitted or absorbed. If the backing material is lightly colored, any transmitted light can be

Colorant,**Thermochromic,**

Fig. 6 An illustrative representation of the change in pitch length and corresponding color of a chiral nematic phase liquid crystal above the S-N* transition [12]



reflected back through the liquid crystal, interfering with the single selected wavelength of reflected light, changing the color. Therefore, thermochromic liquid crystal devices are almost always printed on black backings to absorb the light of wavelengths other than the one selected for reflection [12].

To obtain thermochromic behavior in chiral nematic liquid crystals, the pitch must vary rapidly with temperature. Phase transitions are usually exploited to control variation in pitch. The most common transition exploited is the S-N* (smectic to chiral nematic) phase transition. As the material cools within the N* phase, the pitch elongates (i.e., the liquid crystal becomes more locally structured), and the reflected color changes from blue (at higher temperatures) to red (nearer to the transition temperature). This effect is shown graphically in Fig. 6. Note that further heating of the liquid crystal brings about a transition to the isotropic liquid phase concomitant with a loss of selective reflection and color.

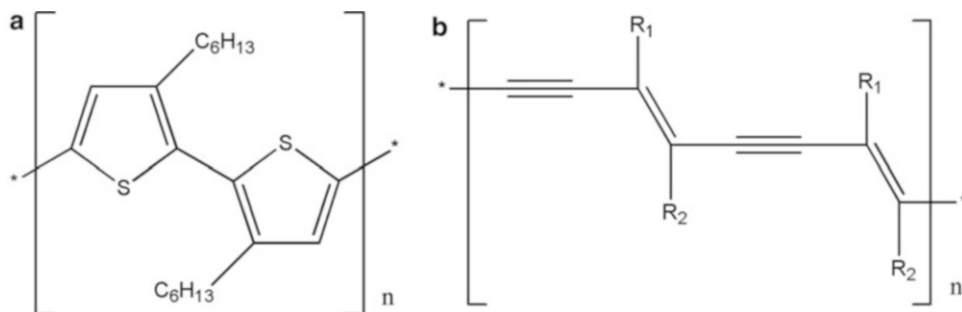
In general, the two important categories of thermochromic liquid crystals behave in much the same way. The major difference can be found in the applicable temperature range for each of the materials. Cholesteric liquid crystals generally have much higher transition temperatures and tend to find applications in thermometers on pasteurization equipment, ovens, and warning indicators on hot surfaces. The (S)-4-(2-methylbutyl)phenol derivatives have

transitions at much lower temperatures, including physiological temperatures and find use in thermometers, in mood rings, and in refrigerator and food spoilage warning labels. Thermochromic liquid crystal devices can be engineered to behave in both reversible and irreversible manners, with the latter being particularly important if the thermal history of a product (e.g., perishable food products) is of particular importance.

Polymers

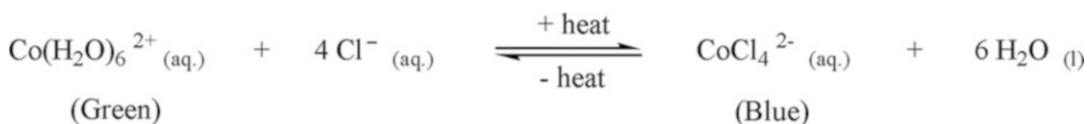
Highly conjugated organic polymers can demonstrate interesting electrical properties resulting from very long conjugation lengths and a high degree of delocalized electron density. As with conjugated small molecules, conjugated oligomers and polymers are susceptible to thermally induced structural modifications and can exhibit thermochromism. Some common thermochromic polymers include polythiophenes (Fig. 7a), polydiacetylenes (Fig. 7b), and α -conjugated polysilanes [15].

Thermochromism in conjugated polymers arises when sufficient thermal energy causes an order-disorder transition involving the bulky side chains of the polymer. The side chains in polymers generally keep the polymer backbone organized in some fashion; the backbone bonds tend to be in either all-*trans* or helical conformations. Above a certain temperature, the side-chain



Colorant, Thermochromic, Fig. 7 (a) The polymer 3-hexyl polythiophene, an important material for organic electronics, has coplanar thiophene rings at low

temperature. (b) Polydiacetylenes also can display thermochromic behavior depending on the choice of side-chain groups



Colorant, Thermochromic, Fig. 8 The equilibrium form for Co²⁺ in aqueous solution and in the presence of Cl[−] changes from green to blue upon heating from 0 °C to room temperature, resulting in thermochromism

groups become dynamically disordered and can no longer keep the backbone chain in its original conformation. The most common transformation is from all-*trans* to *gauche* conformation. Thermochromism results from the change in the HOMO-LUMO gap.

Polythiophenes are widely employed in organic electronics as a conducting layer. They are highly conjugated when the thiophene rings are in a *trans*-planar configuration. Regioregular poly-3-alkylthiophenes (Fig. 7) undergo a reversible color change from red-violet to yellow when heated under vacuum. This change arises from weakening side-chain interactions that are no longer able to maintain the coplanarity of the thiophene rings, resulting in twisting along the chain, a decrease in conjugation, and a change in the wavelength of light absorbed [16].

Inorganic Thermochromism

Thermochromism in inorganic materials can have many different origins: changes in ligand geometry, changes in metal coordination, changes in solvation, changes in bandgap energy, changes

in reflectance properties, changes in distribution of defects in the material, and phase transitions [17].

An example of thermochromism arising from a phase transition is the compound Ag₂HgI₄. At room temperature, the compound adopts a tetragonal crystal structure and is yellow. Upon heating to 50 °C, Ag₂HgI₄ undergoes a first-order phase transition from tetragonal to a cubic phase concomitant with a color change to orange. Upon further heating, it undergoes a gradual (second-order) order-disorder transition to a phase in which the silver ions become mobile in the lattice and the color of the compound changes to black. Therefore, across a temperature range from 25 °C to 75 °C, the material changes from yellow to orange to black [18].

Reversible thermochromism also can be observed for inorganic compounds in solution. In solution, the color changes are often associated with modification of the solvation sphere, changes in coordination number, or ligand exchange with the solvent. A commonly cited example is CoCl₂ in water which is blue at room temperature and green at 0 °C, as shown in Fig. 8 [19]. As the system is heated, the Co(II) coordination changes

from octahedral to tetrahedral and is coupled with ligand exchange. This structural change alters the electronic field experienced by the central Co^{2+} as a function of temperature, changing the wavelength of light absorbed and giving rise to thermochromism.

Perhaps the most interesting inorganic thermochromic compound is vanadium (IV) dioxide, VO_2 , which undergoes a semiconductor to metal transition at 68 °C. The transition modifies the absorption spectrum in the infrared and near-infrared regions. Vanadium dioxide is infrared transmissive below ~ 68 °C and infrared reflecting at higher temperatures [4]. Vanadium dioxide is being considered for use in smart coatings which would allow visible sunlight to pass through a thin film coating but block infrared radiation, thus reducing building cooling requirements. Inorganic thermochromic compounds are of great interest for building coatings owing to their stability to light, which is substantially better than organic thermochromic compounds which are notoriously susceptible to decomposition under extended light exposure [4].

Thermochromic Materials

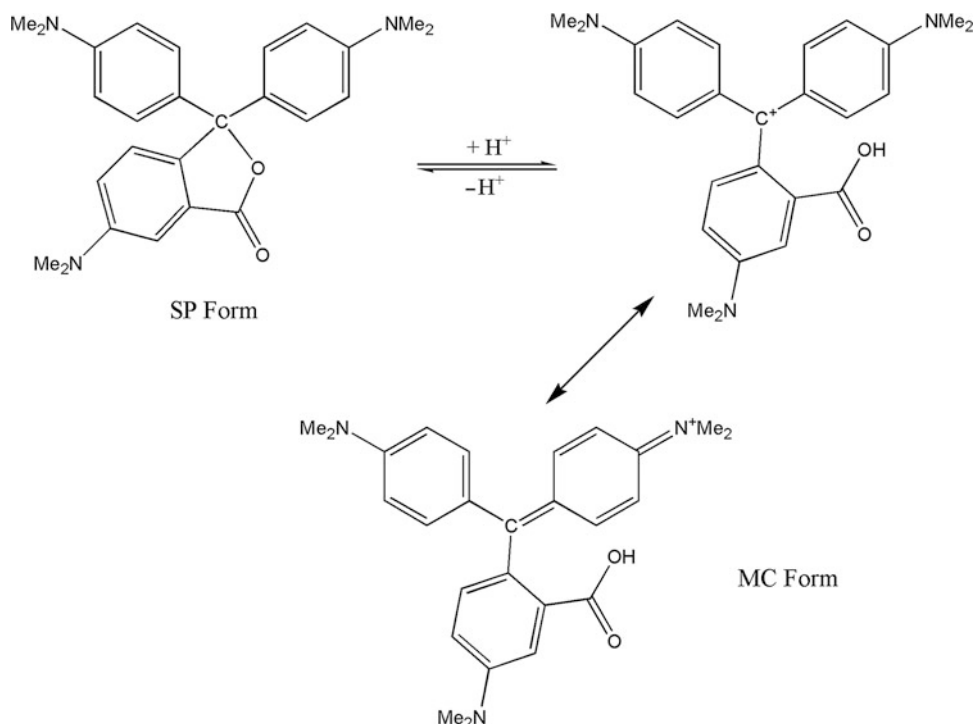
The term *thermochromic materials* refers to multicomponent mixtures of chemicals which, although not necessarily thermochromic individually, create a thermochromic system when mixed in the appropriate proportions. Two popular examples of commercial products incorporating this type of thermochromic material are the Pilot FriXion erasable pen and the Coors Light beer bottle label. The color of ink from the FriXion pen can be erased thermally by the friction created by rubbing the eraser head on the page [follow this “<http://myweb.dal.ca/mawhite/Video/Frixion%20Pen%20Erasing.MOV>” for a video]. The label on the Coors Light beer bottle contains a color-changing dye system that reversibly changes from colorless to blue upon cooling the container to below ~ 6 °C [follow this “<http://myweb.dal.ca/mawhite/Video/Coors%20Light%20Bottle.MOV>” for a video showing warming].

In such mixtures, a color-forming agent, the *chromophore*, reacts with a color-developing agent, the *developer*, to initiate the color-changing reaction. The color-change reaction also is controlled by another component of the mixture, usually referred to as the *cosolvent*, which forms the bulk of the mixture. The cosolvent melts and its melting point determines the color-change temperature. The cosolvent's interactions with the other components also determine if the colored form of the mixture occurs at high or low temperature.

Significant effort within industry has been aimed toward the development of thermally erasable printing inks and toners for large-scale printing in an office setting. National Cash Register Co. developed an irreversible “thermochromic” receipt paper in the 1950s using the spiroactone dye crystal violet lactone (CVL) as the chromophore [20]. The closed ring SP form of CVL is colorless; upon interaction with an acidic compound (or an electron acceptor), the lactone ring opens forming the intensely blue, ring-opened MC form (Fig. 9). Attapulugus clay was originally used to develop the color although more recently phenolic compounds such as bisphenol A (BPA) have been the developers of choice in the United States.

Most of the receipt paper used today in commercial enterprises employs this type of technology, although the chemicals employed are changing. Fluoran dyes are generally used to produce the black color of modern receipt paper, and bisphenol A is being replaced by other, less harmful, phenolic compounds. Today, the chromophores and developers are separated via microencapsulation of the chromophore. Heating the receipt paper causes the microcapsules to rupture, releasing the contents and initiating the coloring reaction. Although this process is technically not thermochromic due to the lack of reversibility, the widespread use of thermal receipt paper warrants its inclusion in this section [20].

An interesting commercial development from Japan is Toshiba's e-Blue erasable laser jet toner. The toner is composed of a blue-colored spiroactone dye and phenolic developer embedded in a polymer matrix. When printed, the toner



Colorant, Thermochromic, Fig. 9 Ring-opening reaction of crystal violet lactone (CVL) in the presence of an acidic developer. The ring-opened charged form has multiple resonance forms, only two of which are shown here

is blue. Heating a printed page will cause a decolorization reaction in which the developer is segregated from the dye, returning the initial uncolored state of the dye and erasing the printed image [21]. The potential benefits of reducing the amount of paper that enters the recycling waste stream are substantial, although poor resistance to color fade, and low image quality thus far have precluded wide usage of such rewritable printing media. These examples of thermochromic materials also fall under the umbrella of the broad *functional dye* field and are discussed further in another chapter in this text.

Thermochromic Colorants

Thermochromic leuco dyes and liquid crystal systems are used in the textile industry for both functional and artistic purposes. The breadth of variation in leuco dye structure permits the

formulation of thermochromic products demonstrating an amazing array of colors. The choice of cosolvent allows for precise control of activation temperatures (i.e., the color-changing temperature). Companies supplying thermochromic products can design products to suit the needs of the textile manufacturer. Virtually any color imaginable can be produced by precise mixing of primary colors (e.g., red, green, blue, etc.), while clever selection of activation temperatures can result in interesting and aesthetically pleasing color-play effects [22].

Thermochromic colorants need to be isolated from their surroundings prior to use in textile applications in order to preserve the intended coloring behavior of the colorant system. To this end, microencapsulation is used to isolate the thermochromic system, with the *coacervation* method being the most common. The microcapsules are dried to form a powder, after which they are usually referred to as *thermochromic pigments*. The pigments can then be made into

slurries, emulsions, or pellets, dissolved into inks and paints, or applied directly to a fabric.

One of the major problems concerning the use of thermochromic pigments is dilution of the dye throughout the processing steps; final dye concentrations can range from 3 %–5 % for pellets to 15–30 % for inks and paints [22]. Other problems include poor stability against UV radiation (e.g., photobleaching), poor resistance to the effects of water and detergents, the cost of the thermochromic material, and, for some, toxicity of the components. In the case of liquid crystals, the fiber onto which the thermochromic pigment is printed must be black to prevent unwanted reflection effects [23]. Additionally, the microencapsulation process can disrupt carefully engineered interactions in multicomponent thermochromic mixtures (i.e., dye, developer, cosolvent mixtures) such that the final microencapsulated product does not behave in the same way as the isolated system.

The use of thermochromic colorants in the textile industry has been mainly limited to novelty applications (e.g., hypercolor T-shirts). More recent textile applications have been focused on using the thermochromic effect for artistic purposes [23]. Many of the thermochromic artistic works reviewed by Christie et al. [23] employed fabric-bound, microencapsulated thermochromic dyes coupled with heat-producing microelectronic devices to initiate the color-changing behavior. Smart materials including electronics-coupled textiles [24], multisensory interactive wallpapers [25], surface coatings containing heat-storing *phase change materials* (PCMs) [26], and soft-woven thermochromic fabrics [27] have been reported. These artistic applications of thermochromic colorants demonstrate the important link between scientific and technological developments and the creativity of the artistic world.

Cross-References

- [Colorant, Photochromic](#)
- [Dye, Functional](#)
- [Dye, Liquid Crystals](#)

References

1. Day, J.H.: Thermochromism. *Chem. Rev.* **63**, 65–80 (1963)
2. Christie, R.M.: *Colour Chemistry*. Royal Society of Chemistry, Cambridge (2001)
3. Crano, J.C., Gugliemetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds*, vol. 2. Kluwer Academic/Plenum Publishers, New York (1999)
4. Granqvist, C.G., Lansaker, P.C., Mlyuka, N.R., Niklasson, G.A., Avendano, E.: Progress in chromogenics: new results for electrochromic and thermochromic materials and devices. *Sol. Energy Mater. Sol. Cells* **93**, 2032–2039 (2009)
5. Sidky, M.M.: Thermochromism. *Chem. Stosow.* **27**, 165–182 (1983)
6. Vitry, J.: Photochromie et Thermochromie. *Chim. et Indus. – Genie Chim* **102**, 1333–1349 (1969)
7. Minkin, V.I., Tsukanov, A.V., Dubonosov, A.D., Bren, V.A.: Tautomeric schiff bases: iono-, solvato-, thermo- and photochromism. *J. Mol. Struct.* **998**, 179–191 (2011)
8. Fujiwara, T., Harada, J., Ogawa, K.: Solid-state thermochromism studied by variable-temperature diffuse reflectance spectroscopy. A new perspective on the chromism of salicylideneanilins. *J. Phys. Chem. B* **108**, 4035–4038 (2004)
9. Muthyala, R., Lan, X.: The chemistry of leuco triarylmethanes. In: Muthyala, R. (ed.) *Chemistry and Applications of Leuco Dyes*, pp. 125–157. Plenum Press, New York (1997). Chapter 5
10. Dickinson, R.E., Heibron, I.M.: Styrylpyrylium salts. Part IX. Colour phenomena associated with Benzonaphtha- and Dinaphtha-spiropyranes. *J. Chem. Soc.* 14–20 (1927)
11. Reinitzer, E., quoted by Lehmann, O. Z.: *Physik. Chemie.* **4**, 462–472 (1889)
12. Sage, I.: Thermochromic liquid crystals. *Liq. Cryst.* **38** (11–12), 1551–1561 (2011)
13. White, M.A.: *Physical Properties of Materials*. CRC Press, Boca Raton (2012)
14. White, M.A., Leblanc, M.: Thermochromism in commercial products. *J. Chem. Ed.* **76**(9), 1201–1205 (1999)
15. Song, J., Cheng, Q., Stevens, R.C.: Modulating artificial membrane morphology: pH-induced chromatic transition and nanostructural transformation of a bolaamphiphilic conjugated polymer from blue helical ribbons to Red nanofibers. *J. Am. Chem. Soc.* **123**, 3205–3213 (2001)
16. Yang, C., Orfino, F.P., Holdcroft, S.: A phenomenological model for predicting thermochromism of regioregular and nonregioregular poly(3-alkylthiophenes). *Macromolecular* **29**, 6510–6517 (1996)
17. Day, J.H.: Thermochromism of inorganic compounds. *Chem. Rev.* **68**, 649–657 (1968)
18. Schwierz, J., Geist, A., Eppler, M.: Thermally switchable dispersions of thermochromic Ag₂HgI₄ nanoparticles. *Dalton Trans.* **38**, 2921–2925 (2009)

19. Marinovic, M., Nikolic, R., Savovic, J., Gadzuric, S., Zsigari, I.: Thermochromic complex compounds in phase change materials: possible application in an agricultural greenhouse. *Sol. Energy Mater. Sol. Cells* **51**, 401–411 (1998)
20. White, M.A., Leblanc, M.: Thermochromism in commercial products. *J. Chem. Ed.* **76**, 1201–1205 (1999)
21. Sekiguchi, Y., Takayama, S., Gotanda, T., Sano, K.: Molecular structures of the coloring species of a leuco dye with phenolic color developers. *Chem. Lett.* **36**, 1010–1011 (2007)
22. <http://www.hwsands.com/>
23. Christie, R.M., Robertson, S., Taylor, S.: Design concepts for a temperature-sensitive environment using thermochromic colour change. *Colour: Des. Creat.* **1**(1), 1–11 (2007). 5
24. Worbin, L.: *Nordic Text. J.* 51–69 (2005)
25. Berzina, Z.: Text: for the study of textile, art, design and history. 33 5 (2005/06)
26. Bremner, F.: Intelligent Tiles. http://www.infotile.com.au/services/techpapers/TT50_3.pdf
27. Xslabs, <http://www.xslabs.net/work-pages/krakow.html>

Coloration, Fastness

Martin Bide
Department of Textiles, University of Rhode
Island, Fashion Merchandising and Design,
Kingston, RI, USA

Definition

Humans use a range of technologies to produce color: solid surfaces are painted, paper and packaging is printed, color images are viewed on screens, and they color the textiles and leather used in their clothing and around homes. Screen color is transient, and few paper products are expected to last long. More durable color is found on painted surfaces and textiles, and the color is provided to them by colorants. That durability is expressed as “fastness.”

Colorants, with the minor exception of interference pigments, have chemical structures that efficiently absorb light and are thus strongly colored. Present in small amounts, they provide their color to an item. That color may change for a number of reasons. Some transient changes can

occur, such as when an item becomes wet, or if the colorant undergoes reversible structural changes under the influence of heat (thermochromism) or light (photochromism). Beyond that, if a colored item maintains its color, it is said to possess fastness. Otherwise, the color of the item may change when either:

The colorant remains in place but is permanently changed or destroyed.

The colorant is unaltered, and maintains its color, but is removed, for example, by washing.

Such a change or loss of color is described as a *lack* of fastness. If the colorant is removed, for example, during laundering, it can stain other items with which it comes into contact. Even when the color of the original item is not appreciably altered, this is also considered a lack of fastness.

Color Fastness May Thus Be Defined as the Resistance to Change or Removal of the Color of an Item

Colorants are divided into dyes and pigments. The difference between them affects both their application and the likelihood of their permanence (► [Colorant, Natural](#)).

A pigment consists of molecular aggregates, usually in the micron range, with minimal solubility in any solvent with which it is likely to come into contact with (► [Pigment, Inorganic](#)). Its removal by dissolution is therefore unlikely. If a few of the pigment atoms or molecules on the surface of an aggregate particle are removed or changed, the overall color of the particle (and hence of the substrate it is coloring) is little affected. Low solubility and particulate nature thus mean that fastness in coloration by pigments is readily achieved.

In contrast, a dye is characterized by its solubility, most often in water, in its application (► [Dye](#)). In many cases its resistance to removal derives from the forces that drive its absorption by a substrate, and it remains soluble when the colored item is in use. The details are discussed in greater detail below, but the solubility may limit the resistance of the colorant to removal, and the

monomolecular dispersion makes the effect of any removal or destruction more apparent.

Dyes are little used for non-textile items. Pigments have some use on textiles, but dyes and textiles are largely mutually inclusive. Solid surfaces typically do little flexing or moving, whereas textiles face a greater range of challenges than most other colored objects: as clothing they must have sufficient flexibility to move with the wearer and be soft enough not to irritate. They rub against each other. Curtains hang in sunny windows. Textiles are regularly cleaned and must withstand hot water and detergent or the solvents used in dry-cleaning. For these reasons, the subject of color fastness is most widely studied for textiles. As discussed below, truly permanent color is an unrealistic expectation, and the extent to which color change is tolerable varies widely and is affected by the cost of the item, the cost of replacing it (or its color), and the possible effect of the color loss on other items. Problems of fastness are therefore strongly associated with dyed textiles, and fastness in other colored materials is often tested using methods and principles drawn from textile tests.

Overview

As outlined in the definition, the color of an item may change when either the colorant remains in place but is permanently changed or destroyed, or when the colorant is unaltered and maintains its color, but is removed. These two cases are considered below.

Color Change: Colorant Destruction

The colorant on an item may undergo a chemical reaction that changes or destroys its color. Such a change is perceived as a lack of fastness. The reaction may take place slowly and only become apparent after a long period of time or may be quite rapid.

The former relates most often to the chemical reactions that occur as a result of exposure to light. Absorption of light raises a colorant molecule to an excited state in which condition it is susceptible to decomposition, or reaction with oxygen, water,

or other available reactants. The reactions involved are complex. They have been the subject of extensive study but much remains to be understood: the factors that affect the reaction are many and varied (light intensity, spectral power distribution, temperature, humidity, etc.); it is also clear that color change is not simply a property of the chemical identity of the colorant alone, but is affected by colorant-substrate interactions, by colorant aggregation, and (in mixtures) by colorant-colorant interactions [1, 2]. The same dye can have different fastness on different fibers. Pigments are less involved in these external interactions, but their physical form (particle size, crystalline form, etc.) may affect their fastness.

Whether dye or pigment, for a constant exposure, the destruction of colorant occurs at a fairly constant rate. The effect on the color, however, is not constant. A darker color, resulting from a greater amount of colorant, will be affected less by a given exposure than a light one, where the same amount of colorant destroyed represents a greater proportion of the colorant present. In other words, a pale shade of a given colorant will have lower fastness to colorant destruction. This is in contrast to the lack of fastness represented by staining as a result of colorant removal discussed below.

The more rapid destruction of dye is usually the result of exposure to a chemical agent: most obviously this might be sodium hypochlorite or hydrogen peroxide used as a bleach, but other chemicals such as benzoyl peroxide used in skin medication will effectively decolorize many dyes. Again, pigments are less prone to such reactions, given their particulate nature and insolubility.

Color Change and Staining: Colorant Removal

Color may be removed from an item, particularly a textile, for a number of reasons. Unlike the case of colorant destruction, the color remains, and the removed colorant may color an item with which it comes into contact: such staining is a component of (a lack of) fastness. For reasons that will become clear, once again, this is best examined from the point of view of dyes.

Dyes are applied to textiles in a number of ways: batch dyeing, continuous dyeing, and

textile printing (► [Coloration, Textile](#)). The principles of the interaction are common to all, but best illustrated in the case of batch dyeing [3, 4]. In a batch dyeing process, dissolved dye molecules are preferentially sorbed from the external dye solution at the fiber surface and penetrate the interior of the fiber as a result of the intermolecular forces that comprise the “substantivity” of the dye. The relative motion of the dyebath and the substrate replenishes the dye-depleted solution at the fiber surface, and the rate-determining step of the overall process is typically the rate of sorption into the fiber. If dyeing conditions are maintained long enough, an equilibrium is established between dye in fiber and dye in solution, and all fibers are fully and equally penetrated. In practice this situation is rarely achieved, and microscopic examination of a dyed textile will reveal fibers and yarns with only partial penetration of dye. Such microscopic unlevelness is tolerable.

An essential requirement of a colored textile is that the color be level at the macroscopic scale. Assuming a clean and absorbent substrate from earlier preparation processes, and homogeneously dyeable fibers, levelness results from a balance of two possible routes. Either the dye is absorbed in a level manner, or an initial unlevel sorption is made level by the migration of dye [5].

The substantivity of the dye is based on the molecular size and shape of the dye and functional groups present. Together, these allow the formation of a range of non-covalent bonding that makes up the attraction of dye for fiber and which enables dyeing to take place [4]. If nothing else happens, these same forces will provide the fastness the dye has in use, in competition with the solubility of the dye in the particular medium (e.g., laundry liquor).

For fastness (as resistance to removal), a high substantivity (strong dye-fiber bonding) is required. But dyes with high substantivity do not migrate readily, and if initial application is unlevel, it is difficult to achieve levelness via migration. To use a high-substantivity dye, the dyer must control the initial uptake (“strike”) of the dye to be as level as possible. For a given system (machine, dye, fiber), dye uptake can be controlled by control of temperature (rate of rise

and ultimate temperature), time, and dyebath conditions of ionic strength (salt) pH, and auxiliaries that moderate the dye-fiber interactions (often surfactant based). However, the demands of levelness may require that lower-substantivity (and thus less fast) dyes be used.

As discussed earlier with colorant destruction, the fastness of a particular dyeing will vary with the depth of shade. For the case of colorant removal, a dark shade will exhibit lower fastness than a pale one: the same percentage of colorant removed from a dark shade will have the propensity to stain more heavily than from a pale shade.

For some dye-fiber systems, the substantivity of the dye that drives dyeing is all that provides fastness in later use. Most notably, acid and metal-complex dyes on protein and polyamide fibers (wool, silk, and nylon) direct dyes on cellulosic fibers and disperse dyes on acetate. In essence, a subsequent aqueous challenge represents the equivalent of a dyebath and the (beginnings of) reestablishment of an equilibrium in which some proportion of the dye is lost from the substrate.

Several dye-fiber systems have fastness attributable to reasons beyond those provided simply by the dye-fiber interactions. Disperse dyes on polyester have fastness greater than the same dyes on acetate. The high glass transition temperature of polyester means that practical dyeing takes place only at temperatures considerably higher than that of boiling water: these are found in pressure dyeing machinery. After dyeing (and in use), the fiber is cooled below T_g , and the dye is essentially trapped in the fiber: the inaccessibility of the dye to outside agencies is reflected in the use of “reduction clearing” to remove any surface deposits of the low-solubility dye. Similar arguments apply to the fastness of basic dyes applied to acrylic fibers.

While a dye must be soluble in application, greater fastness can be achieved if the dye is later converted (or reverts) to an insoluble or less soluble form. This form of the dye may also comprise molecular aggregates and/or larger molecules for which physical entrapment may contribute to the fastness. Vat dyes for cellulosic fibers are produced and sold in insoluble form and must be chemically reduced to a soluble and

substantive form for application. Once on the fiber, they are oxidized to their original insoluble form and form aggregates inside the fiber: essentially they are now pigments. Several dye types are formed inside the fiber as low-solubility moieties as part of their application. Sulfur dyes for cellulosic fibers are supplied as a reduced solution; once applied, they undergo a similar oxidation to vat dyes to form a low-solubility dye. Azoic colorants are generated by reaction within the fiber of a soluble coupling component and a soluble diazo component. The resulting azo dye is insoluble and, like a vat dye, aggregates into pigment form. Mordant (“chrome”) dyes on wool rely on a reaction of dye and mordant in situ within the fiber to form a low-solubility complex of good fastness, and its fastness is enhanced by the entrapment of a physically large molecule.

A further reason for additional fastness may arise from the presence in the dye of functional groups capable of reaction with the substrate to form covalent bonds. So-called reactive dyes achieve their fastness in this way and since their introduction in 1956 have become the dominant dye type for cellulosic fibers, with additional usefulness for wool dyeing where they can replace the mordant dyes and their associated use of heavy metals [6]. It should be noted that the reaction efficiency is not 100 %, and at the end of the dyeing process, any unfixed dye that is not removed by washing will be held only by the forces of substantivity and may be readily removed in later use. In such a case the dyeing as a whole will be considered to have poor fastness.

Achieving Fastness

Colorfastness, whether as resistance to destruction or resistance to removal, is largely achieved by choice of colorant. Dye manufacturers will supply customers with “pattern cards” of materials colored with individual dyes, together with the results of key fastness tests (see below) conducted on those materials. Colorists may thus select colorants that are likely to meet fastness requirements, although as mentioned earlier, the

depth of shade will affect the fastness achieved. Resistance to removal may additionally depend on the ability of the process to eliminate any dye that is not fixed or that is loosely held on the fiber surface.

Pigments are used for textiles in one of two very different ways. A pigment can be added to the melt or solution used to form manufactured fibers in the same way that a pigment is added to a melt used to mold plastic items. Subsequent solidification traps the pigment within the polymeric matrix. Polymers colored this way are among the most colorfast of all. However, for textiles, the time between fiber manufacture and ultimate sale may be a year or more, and the holding of multiple color stocks of fiber and yarn means that this coloration method is of limited use in this context.

Pigments can also be bound to a textile material by use of a polymeric binder: this is akin to the application of paint to a surface. The fastness of such colored materials is related to both the colorant and the binder and the strength of the pigment-binder-fiber interactions [7]. Once again, textiles tend to face greater challenges as they flex and move.

Fastness Testing

The permanence of color in an item is desirable, but the challenges it might face in use are many and varied. It is unrealistic and expensive to demand the best fastness in every case, so compromises are inevitably made. The questions then become, what are the likely challenges this item will face? How well *should* it be able to resist them? How well *does* it resist them? The answer to the first derives directly from the intended end use, while the answers to the second and third require some way of presenting the challenge and assessing the resistance. The most realistic way to do this is to conduct a trial with the item in its intended use to see when and how failure occurs. Such real-life trials are occasionally performed, but they are very expensive and take a long time to produce results. Standardized laboratory fastness testing provides an economical and useful alternative. Standard test methods are

developed that are designed to approximate to a single real-life challenge and to predict how an item will respond. Such tests typically provide results rapidly, either because the challenge itself is a rapid one (e.g., color being rubbed off) or because the test accelerates the challenge (e.g., using a very intense light instead of real daylight).

A fastness test is developed to provide the challenge and a way of assessing the result. The result is then compared to a requirement that forms part of an overall product specification. The tests may be widely applicable or represent a challenge that is more specific. Tests are developed by standard setting organizations. Much of the original work to develop these tests for textiles was carried out by the Society of Dyers and Colourists (SDC) [8] in the UK and the American Association of Textile Chemists and Colorists (AATCC) in the USA [9]. With the rise in global trade, fastness tests have become more international, and many countries now rely on tests published under the auspices of the International Organization for Standardization (ISO) [10].

A good test should be valid: the property it measures should have some real-life relevance, and the test should predict in-service suitability. Additionally, a good test should be simple in terms of how it is performed and how easy the instructions are to understand. It should also be reproducible giving the same results from operator to operator and lab to lab. In some cases the control of test conditions may be difficult, and standard fabrics of known susceptibility are used as control materials.

Issues of environmental, health, and sustainability have come to the fore in recent years. Several certification schemes are intended to reassure the consumer that an item has been produced in an environmentally friendly manner and that it does not contain (or will release) substances of concern. Colorants are among such substances, and the measure of colorant release is often based on standard color fastness testing. Thus, for example, a range of color fastness tests is included in the OEKO-TEX 100 certification [13] and in the Global Organic Textile Standard [14].

Testing for Fastness: Challenges

As discussed earlier, fastness tests are designed to reproduce the challenges that a colored item would face in real life. Again, the subject tends to emphasize textiles. Some tests relate to challenges faced in the sequence of manufacturing steps that a textile item undergoes before it gets to the consumer. Most relate to the challenges encountered after the item is sold: these can be divided into those met when the item is being used and those involved in its cleaning or refurbishment. Full details of the tests are sold by from the relevant standard setting organizations. The titles and scope are available in texts and on the organizations' web sites [11, 12, 15].

Fastness to (Textile) Production Processes

Textiles woven from multicolored yarns may be scoured and may include a white portion that requires bleaching, and thus a test for fastness to bleaching may be needed. Mercerizing can alter color. Dry heat can cause color changes. Wool undergoes a variety of wet treatments: tests determine the color changes caused when wool is carbonized, boiled in water, bleached, set with steam, or milled. The fastness challenges of the last are reflected in the naming of certain acid dyes that will withstand this as "milling acid dyes."

After makeup, garments may be pleated or hot-pressed: these can cause thermochromic or sublimation-based color changes. Silk fabrics may be degummed in hot alkaline soap solution, and a fastness test can predict any changes to colored materials that undergo this process.

Fastness in Use

In use, garments may be rubbed, and their color transferred to another item: fastness to rubbing (also known as "crocking") is one of the most common tests.

Spots of liquid can cause a color change, either by moving dye within the textile or removing the

dye and transferring it to an adjacent material. Thus there are tests for spotting by water, dry-cleaning solvent, seawater, perspiration, acids, and alkalis. Swimming pool water with chlorine represents a destructive challenge.

Atmospheric contaminants, notably ozone and oxides of nitrogen (“burnt gas fumes”) derived from combustion and sunlight, can destroy colorants.

Light represents a particular challenge. Real-life exposure is both slow to produce change and is highly variable (but such real-life exposure is included among standard tests). More usually, an accelerated test uses a xenon arc lamp and controls the temperature and humidity of exposure. The test may be extended by the inclusion of water sprays to cover “weathering.”

Fastness to Refurbishment

Textile items get dirty and require periodic cleaning. The processes to be used are given in a care label, and the suggested cleaning method should have been shown by testing to be appropriate to restore the item to an acceptable level while not causing any damage. However, such refurbishment represents another set of challenges to the color.

Laundering is conducted in aqueous surfactant solutions with agitation. The details can vary considerably: the detergent (type and amount used), the temperature, the amount of water and the extent of agitation (based on the type of machine used), the nature of the materials used to make up the bulk of the load (“ballast”), and the presence or absence of bleach. The tests to assess fastness reflect this variation in practice. Full-scale laundering, repeated three or five times, may be used, but since each item should be tested individually to avoid possible confusion in results, an accelerated test is used. A small sample, usually with a standard multifiber adjacent fabric, is agitated in standard detergent solution in a small cylinder for 45 min or so. The effect on color will approximate to that obtained in five real cycles of laundering. Similar considerations apply to dry-cleaning.

Ironing can affect color: tests examine the effects of dry, moist, and wet heat.

Assessing the Results of Fastness Tests

The tests produce color changes and thus assessment of the results is an assessment of color difference [16]. This might be the difference between the original color and the challenged color or the difference between an unstained fabric and one stained by color removed from the test sample.

Color difference assessment has been widely studied for judging colors for acceptable closeness to a standard color (“a match”). That use concentrates on small color differences and is as much concerned with the quality of difference as the quantity of difference. However, the assessment of color differences from fastness testing covers a much wider (quantity) range of color differences and is rarely concerned with the quality of difference. These two aspects of color difference measurement have thus tended to go their separate ways.

The results from a fastness test can be assessed subjectively, by reference to physical color standards (“gray scales”) or objectively, based on spectrophotometric or digital camera measurements. Physical gray scales have been so extensively used that even when an instrument is used, its output is converted to gray scale ratings.

Gray Scales

The use of physical standard color difference pairs has long been standard practice in fastness testing. These physical standards include a range of difference pairs, from small to quite large. There are two gray scales. One is used to assess a change of color, and the second is used to assess the extent to which a white material is stained. The gray scale for color change (Fig. 1) consists of pairs of neutral gray chips, one of the pair being a constant gray color, with the second ranging from the same to very different arranged in order of increasing difference. The pair with no difference is labeled “5,” and the increasingly different pairs are 4.5,

4, 3.5, 3, 2.5, 2, 1.5, and 1. The gray scale for staining is similar, but the constant color is white, and the increasing difference is provided by successively darker grays (Fig. 2).



Coloration, Fastness, Fig. 1 Coloration fastness: AATCC gray scale for color change (Copyright AATCC)

Coloration, Fastness, Fig. 2 Coloration fastness: AATCC gray scale for staining (Copyright AATCC)



In addition to the two basic gray scales, for color change and for staining, AATCC has produced a “chromatic transference scale” (Fig. 3). This is used mainly in the assessment of the staining produced in rubbing fastness tests.

The correct use of each of these scales is described in the relevant ISO and AATCC documents. The angle of viewing, the quality and intensity of the illumination, and the masking of the test specimen/gray scale pair are specified to eliminate variables and increase interlaboratory agreement.

Instrumental Color Measurement of Fastness Test Data

A numerical color difference equation might reproducibly express the color change or staining resulting from a fastness test. The need for this to correspond to human visual data over a wide color range and to make it agree with the results generated visually with gray scales has made this somewhat challenging. Nonetheless, while some deficiencies are recognized, mathematical formulae for gray scale ratings derived from spectrophotometric measurements or digital camera data are

Coloration, Fastness,

Fig. 3 Coloration fastness:
AATCC chromatic
transference scale
(Copyright AATCC)



published and used. Work to develop better formulae continues.

Conclusions

Color is an important property of many items, and its durability in items that are expected to last over a period of time is a matter of concern. Textile colorants have the greatest potential to produce unwanted changes in the textiles they color and to stain adjacent materials, and textiles are subject to the greatest range of potentially damaging agencies. Thus color fastness testing is most widely studied and performed on textiles. The tests reflect real-life challenges met in textile processing, in use, and in care, and many of them are accelerated. ISO and AATCC are the most active organizations in the development of such tests.

The tests require a measurement of the changes produced by the test challenge and the stains caused by removed color. For many years, measurement has been based on visual comparison with standard gray scales. Compared to other situations involving color assessment, the adoption of instrumental methods to assess fastness results has been slow. Standard methods for such instrumental assessments are in place, if yet to be widely adopted.

Cross-References

- Colorant, Natural
- Coloration, Textile
- Dye
- Pigment, Inorganic

References

1. Haillant, O.: The photofading of colored materials. *Sunspots* **39**(84), 1–11 (2009). www.atlas-mts.com/fileadmin/downloads/SunSpots/vol39issue84.pdf. Accessed 20 Jan 2013
2. Zollinger, H.: *Color Chemistry: Syntheses, Properties and Applications of Organic Dyes and Pigments*, pp. 245–280. VCH, Weinheim (1987)
3. Broadbent, A.: An introduction to dyes and dyeing, and dyeing theory. In: Broadbent, A. (ed.) *Basic Principles of Textile Coloration*, pp. 174–214. Society of Dyers and Colourists, Bradford (2001)
4. Johnson, A. (ed.): *The Theory of Coloration of Textiles*. Society of Dyers and Colourists, Bradford (1989)
5. Etters, J.N.: Linear exhaustion versus migration: which is the better road to Rome? *J. Soc. Dy. Col.* **112**(3), 75–80 (1996)
6. Renfrew, H.: *Reactive Dyes for Textile Fibers*. Society of Dyers and Colourists, Bradford (1999)
7. Anon: Factors Affecting the Drycleaning Fastness of Pigment Prints. In: *Pigment printing handbook*, pp. 131–142. American Association of Textile Chemists and Colorists, Research Triangle Park (1995)
8. The Society of Dyers and Colourists: <http://www.sdc.org.uk/>. Accessed 5 Feb 2013

9. The American Association of Textile Chemists and Colorists: www.aatcc.org. Accessed 5 Feb 2013
10. Smith, P.: Colour fastness testing methods and equipment. *Rev. Progress. Color.* **24**, 31–40 (1994)
11. ISO Standards Catalog: TC 38/SC1 – tests for colored textiles and colorants. www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm?commid=48172. Accessed 29 Jan 2013
12. AATCC Test Methods & Evaluation Procedures: www.aatcc.org/testing/methods/index.htm. Accessed 21 Jan 2013
13. Oekotex standard 100: Limit values and fastness. www.oeko-tex.com/en/manufacturers/test_criteria/limit_values/limit_values.html. Accessed 21 Jan 2013
14. Global Organic Textile Standard: Section 2.4.14, Technical quality parameters. www.global-standard.org/the-standard/latestversion.html gots-version3_01march2011.pdf p18. Accessed 21 Jan 2013
15. Bide, M.: Testing textile durability. In: Annis, P. (ed.) *Understanding and Improving the Durability of Textiles*, pp. 126–142. Woodhead, Cambridge (2012)
16. Bide, M.: Colour measurement and fastness assessment. In: Gulrajani, M. (ed.) *Colour measurement: Principles, Advances and Industrial Applications*, pp. 196–217. Woodhead, Cambridge (2010)

Coloration, Mordant Dyes

Martin Bide
Department of Textiles, University of Rhode
Island, Fashion Merchandising and Design,
Kingston, RI, USA

Definition

Probably the one word that those otherwise unfamiliar with dyeing know is “mordant” and its derivation from the Latin “mordere” (to bite) with the implication that such a compound forms a link between dye and fiber. However, given the complexities of the dyeing process, it is difficult to write a simple definition for “mordant.” The concept is rooted in history, when only natural fibers and natural dyes were available, and scientific understanding was insufficient to explain fully the functions of the various materials used to provide dyeings that were fast to washing, sunlight, etc. As that understanding has developed,

the need for mordants has declined to the point where they are commercial oddities. For the purposes of this entry, the definition is as follows:

A mordant is a substance applied to a textile substrate in parallel with the dyeing process that modifies the interaction of dye and fiber (and remains present in the subsequent dyed material) to provide better uptake, better fastness, and/or a wider range of colors than would otherwise be achieved.

Nevertheless, there are substances that fall into this definition which are not called mordants, and some substances described as mordants that do not fit the definition. O'Neill in the 1860s [1] suggested that a mordant should “exert an affinity for the fibrous material to which it is applied, and an attraction for coloring matters,” and this is largely in agreement with the definition given here, but implies that the attractions are simultaneous which is rarely the case. Shore [2] gives a definition based on the twentieth-century commercial practice but recognizes the implicit ambiguities.

From the definition it is clear that mordants compensate for some lack of substantivity or fastness of a dye. The watershed between a dyeing world that relied heavily on mordants and one that was becoming relatively mordant-free coincides with the move from natural to synthetic dyes. Mordants were the means of getting good results from natural dyes that were not “designed” for dyeing textiles. Mordants were used for centuries to achieve remarkable results. That strong historical aspect complicates the definition: more recent understanding of the chemistry behind the dyeing process has revealed that substances formerly included in the broad scheme of mordants do not truly behave as such.

For several decades prior to the full development of synthetic dyes, chemical knowledge was increasing, and as new elements were identified, and new compounds prepared, a wide range of substances was tested as mordants to expand the coloration technology of natural dyes. Even after the first synthetic dyes were developed, their usefulness was broadened by the use of mordanting materials. A notable development occurred in the 1880s, when a new group of synthetic dyes was named “substantive dyes” (later “direct dyes”),

because they were substantive toward and dyed cotton “directly,” *without* the use of a mordant. Since then, the skill of the chemist has obviated the need for mordants by creating dyes that not only have color but which also have the molecular configuration to provide dye-fiber substantivity, ultimate insolubility, or functional groups to react covalently with the fiber. The result is a much-simplified (and more efficient) dyeing process that readily provides fast colors. The (synthetic) dyes logically form groups of suitability for certain fibers and of similar dyeing behavior. These groups were codified in the second edition of the Colour Index in 1956 [3]. One such group was “mordant dyes.” These dyes were applied to wool in conjunction with a chrome mordant: no other synthetic dyes routinely employed a mordant, and few mordants other than chromium salts were being used. The use of these mordant dyes has lessened as environmental concerns over the use of heavy metals, such as chromium, increase. Together with the very limited use of natural dyes, it is not surprising that mordants are a rarity in (commercial) dyeing today. Current use of mordants is largely confined to home and craft dyeing, where the purported advantages of natural dyes are emphasized, and the inefficiencies of the processes and environmental limitations of the required mordants are not. In an interesting paradox, the scientific ignorance that accompanied the historic use of mordants is becoming apparent once more. A proportion of the community of interest in applying natural dyes is unaware of the temporary function of dyebath additives (described below), and it is not unusual to read of *any* chemical adjunct to the dyeing process being inappropriately referred to as a mordant.

Synthetic dyes were well established before the development of synthetic fibers, and thus when such fibers were introduced, dyers had mordant-free processes with which to dye them. With one or two exceptions, mordants have been almost exclusively associated with the dyeing of natural fibers. In some cases when materials have been used to improve the interaction of dyes and synthetic fibers, they have not been referred to as mordants, even though they fulfilled essentially the same function.

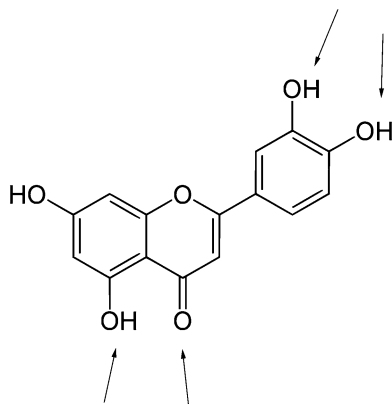
Overview

In a dyeing process, dyes are soluble (or sparingly soluble) and a (textile) substrate (fiber, yarn, fabric) absorbs the dye through forces of attraction from an external solution [4] (► [Coloration, Textile](#)). The sum of the attracting forces is referred to as substantivity. The concept of “substantivity,” relating to dyes that impart color without the use of a mordant, is attributed to Bancroft [5] and significantly predates the development of synthetic dyes, even though inherent substantivity is rare in natural dyes. Substantivity also contributes to the subsequent resistance of the dye to removal in use (► [Coloration, Fastness](#)).

A dyer seeks to achieve a steady and even uptake of dye by the substrate. For a given dye/substrate/machine, this is achieved with control of temperature, agitation, and liquid volume, augmented with dyebath additives. Such adjuncts to the dyeing process provide the correct ionic strength (as electrolyte), pH (acids and bases), and moderation of the dye-fiber-water interactions (often surfactants). While it may not be obvious, similar interactions take place in textile printing with dyes, albeit in the confines of a small volume of print paste, which will also include substances (“thickeners”) to control its rheology and balance the ready transfer of paste onto the fabric with the need to maintain a sharp printed mark. These various dyeing and printing adjuncts, however, do not fall under the definition of mordants. They remain within the exhausted dyebath or print paste at the end of the process and are lost from the colored fiber in any subsequent rinsing.

Natural Fibers

As mentioned earlier, the use of mordants was and is largely confined to the use of natural fibers. These fall into two distinct groups: protein (wool and silk) and cellulosic (chiefly cotton and linen). Mordants behave somewhat differently for each of these two groups of fibers. Protein fibers comprise many different amino acids with a range of functional groups that provide binding opportunities for colorants of all kinds [6]. Natural dyers overwhelmingly prefer to dye wool for this reason. Mordants are/were nevertheless widely used



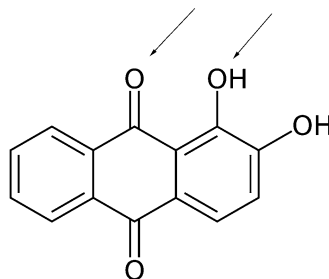
Coloration, Mordant Dyes, Fig. 1 Luteolin, the main coloring matter of weld (“dyer’s rocket”). Potential binding sites for mordants are *arrowed*

to provide better fastness and to broaden the range of colors obtainable from a limited number of natural dyes. Cellulose lacks this natural affinity for colorants and, with the notable exceptions of indigo (which does have substantivity, albeit low), turmeric, archil, saffron, and annatto (which have fastness limitations that make them of limited value), natural dyes do not readily dye cellulose. Mordants provide the main means by which cellulose may be satisfactorily dyed with natural dyes.

Natural Dyes

Many natural organic substances are colored, and a few of those colors can be employed as dyes. Chemically, many are flavonoid derivatives (Fig. 1); anthraquinones (Fig. 2), carotenoids, and indigoids also provide useful dyes [7, 8]. Most provide yellow-orange-brown colors: bright reds are less common, and blues are rare. Fastness varies, especially to light.

Over time, expert European dyers learned how to apply these dyes with the appropriate mordants to give bright and fast colors, versus those that were less satisfactory. Thus Venetian dyers were either of greater or lesser arts. French dyers were “au grand teint” and “au petit teint” [9]. In each case the former were the more expert. Global exploration would eventually provide better natural dyes. Woad was the original blue in Europe, replaced with indigo (the same essential colorant)



Coloration, Mordant Dyes, Fig. 2 Alizarin, the main coloring matter of madder. Potential binding sites for mordants are *arrowed*

when that became available. Reds and purples were from madder and kermes, the latter replaced when cochineal from Central America was discovered. Yellows were from weld or Persian berries, later from quercitron found in North America. Dye woods (sappan, brazil, cutch, fustic, logwood) were also important. Other than woad or indigo, these were used in conjunction with a mordant. The inventive chemistry that allowed these natural dyes to give good results when they were all that was available today allows craft dyers and researchers to reexamine a wide range of plants as sources for natural dyes in the supposition that these dyes are somehow more “sustainable.”

Mordants for Cotton and Cellulosic Fibers

Since few natural dyes display substantivity toward cellulosic fibers, the use of mordants is essential to achieving satisfactory dyeings and prints with those dyes. While mordants were being widely used, the understanding of the chemistry behind their use was limited: as knowledge increased, the use of mordants was declining. Two sources [10, 11] give a valuable survey of the use of mordants on cellulosic substrates when these were being widely used. The most successful mordants were metallic salts that could be applied as a soluble salt, rendered less soluble, and subsequently formed a complex with the dye. For dyeing a solid color, the fabric would be padded with this metal salt. For printed designs, the metal salt would be incorporated into a print paste. Such a metal salt, if it remained soluble, would redissolve in the subsequent dyebath: the dye-mordant

complex formed in the bath would tend to deposit on the textile surface and later be prone to rub off. Thus some means of reducing the solubility of the metal ion and keeping it in place on the substrate as the dye-mordant reaction took place in the dyebath was required, especially in a print where reserve of the unprinted (white) areas was essential [1].

Insolubilization could be accomplished in a number of ways. Using aluminum as an example:

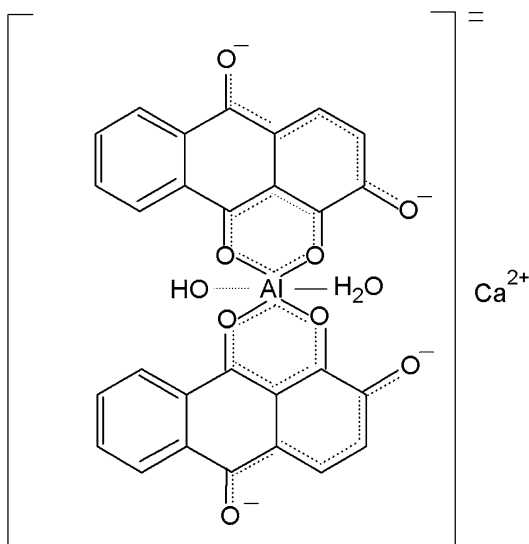
1. A soluble aluminum salt such as aluminum sulfate or aluminum potassium sulfate (a.k.a. potassium alum) could be neutralized with carbonate or bicarbonate to produce a “basic aluminum sulfate” or “basic alum,” respectively. Padding with this, followed by treatment with ammonia, would precipitate aluminum hydroxide on the fiber. Alternatively, the mordant metal could be precipitated as the phosphate, arsenate, or silicate [10].
2. Co-application of a complex anionic substance would provide insoluble complex salts. Tannins from a range of sources were ubiquitous in this regard, from galls, myrbolam, cutch, or (later) purified tannic acid, resulting in a deposit of aluminum tannate [10]. The effectiveness of “mordants” based on modified vegetable oils (usually castor or olive oil) was probably due to a similar mode of action. (Note: cotton will absorb tannic acid from solution, and cotton so treated does dye more readily, so tannic acid does have a moderate effect as a mordant by itself.)
3. Salts such as aluminum acetate could be “aged” in moist conditions under which acetic acid would be lost, generating an insoluble hydroxide. A “dunging” (originally with cow dung, later with synthetic alternatives such as sodium phosphate or arsenate) process would both complete the fixation of the mordant and remove unfixed mordant. This process was especially important in printing [11].

Given the above it is not surprising that nineteenth-century references to mordants usually include a similar application of a metal salt that would subsequently be reacted with a second

metal salt to generate a colored inorganic compound. For example, a “lead mordant” (lead acetate) applied by padding and drying would be passed through a bath of sodium or potassium dichromate to produce insoluble yellow lead chromate [12]. This precipitation reaction is sufficiently rapid that the insolubilization of the original “mordant” was not required. This use of the word mordant is not current.

Excluding, therefore, the use of the word mordant as a component of a precipitated mineral colorant, aluminum and iron were used most extensively as mordants for natural dyes on cotton in both dyeing and printing. Stannic salts were occasionally used. Both iron and aluminum could be produced as the acetate. Iron could be directly dissolved in pyroligneous (acetic) acid, while a reaction between aluminum or ferrous sulfate and lead acetate or calcium acetate would precipitate the lead or calcium sulfate and leave aluminum or ferrous acetate in the supernatant liquid. These acetate solutions were referred to as purple liquor (for iron) and red liquor (for aluminum), based on the colors they would produce in a madder dyebath [13]. Tin salts were common adjuncts to aluminum salts (where they prevented dulling by the presence of small amounts of iron salts). Chromium salts were less easy to precipitate in the fiber and were not extensively used, other than to provide a black color with logwood in which processed dye and mordant were often applied together (compare the dyeing of wool by the metachrome method discussed below). Texts of the time also include a range of other metal salts among the discussions of mordants for cotton, but indicate that their function is often that of an oxidizing (Cu, Cr) or reducing (Sn) agent.

The immersion of a mordanted fabric in a dyebath would allow the formation of a dye-metal complex. (This is discussed in greater detail under mordant dyeing of wool, below.) In a print, the unmordanted dye of the dyebath would inevitably stain the unmordanted portions of the cloth, and the fastness of the dye-mordant complex in the colored areas is indicated by its ability to survive the subsequent “clearing” of dye from white parts of the print by extensive washing processes.



Coloration, Mordant Dyes, Fig. 3 “Turkey red” (alizarin-aluminum complex). Compare Fig. 2, alizarin

As outlined above, the process is reasonably straightforward, although requiring care and attention to achieve successfully. Some idea of the added complexities and obfuscation generated by ignorance can be gleaned from any reading of the history of Turkey-red dyeings. Modern science has identified the final product on the fabric as a 2:1 alizarin: aluminum complex anion with an associated calcium cation [14] (Fig. 3). But for years the process was shrouded in mystery, and much money was paid for supposed recipes for this color [13]. As originally described, the process involved many steps carried out over a period of weeks or months. The deposit of fatty acids (from olive or castor oil) was a part of the process, leading to references to “oil mordants.” Advances during the nineteenth century such as the introduction of synthetic alizarin, “Turkey-red oil,” and chemical bleaching methods allowed for the dyeing to be carried out in days instead of months [15].

Given the difficulty in dyeing cotton and other cellulosic fibers with natural dyes, it is not surprising that as scientific knowledge increased in the early nineteenth century many attempts were made to modify those fibers to improve their dyeability. Dyers of the nineteenth century attempted such modifications to cotton. The

substances have been described as “mordants” [11], but if these modifications are carried out in bulk, separately from the dyeing process, they stretch the definition somewhat. The modifications were presumably prompted by the greater substantivity of dyes for protein fibers and mostly consisted of depositing protein material onto cotton. Albumen from eggs or blood, gelatin, and casein and lactarine from milk were used in these attempts. The efforts were classified as “animalizing” cotton. Such modifications continue to this day: cotton can be pretreated to have cationic groups and readily takes up anionic dyes [16]. Cotton so treated is not referred to as being mordanted, however.

A new category of mordants for cotton was required when the first synthetic dyes were introduced following Perkin’s synthesis of mauveine in 1856. For almost 30 years, these new dyes comprised what are now classed as acid and basic dyes, which have no substantivity toward cotton. Dyers were obviously eager to make use of the bright colors they offered, and the precipitation of tannic acid with salts of antimony (in the form of antimony potassium tartrate, or “tartar emetic”) on the fiber provided an anionic substrate with which cationic basic dyes would interact sufficiently well to provide substantivity and fastness [15].

A subplot in the saga of cotton and mordants occurs in efforts in the twentieth century to improve the fastness of the direct dyes. These efforts comprised aftertreatments to a dyed material. Mordants typically being applied before the application of dye, these aftertreatments are not usually regarded as mordants, but the interactions have factors in common with those undergone by mordants. Anionic direct dyes could thus be aftertreated with cationic compounds to form a complex with reduced solubility and thus better fastness to washing. Direct dyes with functional groups capable of chelating metal ions might be aftertreated with copper or chromium salts: a dulling of shade would be compensated by improved fastness to light and (less so) to washing. This technique is technically indistinguishable from the application of mordant dyes to wool by afterchroming but is

not regarded as an example of the use of mordants [3].

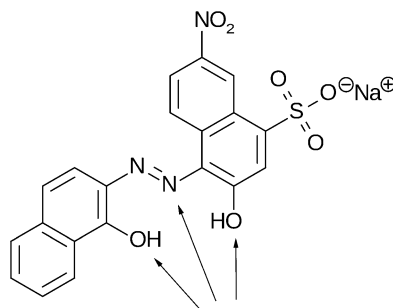
Mordant Dyeing of Wool

As discussed earlier, the use of a mordant for applying natural dyes to wool has less to do with achieving substantivity than it was to provide better fastness than the unmordanted dye and to extend the range of colors available. Mordants for wool were (and are) overwhelmingly metallic compounds: no parallel with the animalizing reactions carried out on cotton was developed. The non-mordant cationization of cotton is perhaps echoed in comparatively recent research to render wool dyeable with disperse dyes [17].

The same (natural) dye will provide distinctly different colors depending on which metallic salt it interacts with. The functional groups on the polypeptide chains that form the keratin of which wool is composed include groups, notably carboxylic acid, with which metal ions can form bonds, holding them in place during subsequent interaction with the dye. Thus the insolubilization required for cotton mordants is not needed. For those reasons, the number of metal salts useful as mordants is greater with wool than with cotton and includes tin, copper, and chrome. Wool additionally has sulfur-containing amino acids that will reduce (or maintain in a reduced form) polyvalent metals: thus chrome is applied as dichromate (Cr^{VI}) and interacts with dye as Cr^{III} . Tin is applied and interacts as Sn^{II} . Dye recipes often include mild reducing agents (tartrate, oxalate) to assist this.

Natural dye structures (Figs. 1 and 2) have quinonoid, hydroxyl, and carboxy groups capable of chelating metal ions. Synthetic mordant dyes have the same groups plus nitrogen-containing (amino and azo) groups (Fig. 4). Dyes are usually bi- or tridentate ligands, and the metals they interact with usually have coordination numbers of 4 or 6.

The chelation reactions on the fiber between chrome and synthetic mordant dyes have been most widely studied, but the principles involved are the same as those between other metals and natural dyes. Dyemakers simplified the mordant dyeing process by carrying out the dye-metal



Coloration, Mordant Dyes, Fig. 4 CI mordant black 11 (14645). Sites for coordination with chrome mordant arrowed

binding in dye manufacture and selling “premetallized” dyes (now known as metal-complex dyes), and the chemistry of this complexation is more easily studied in the absence of the fiber [2, 18, 19]. (While these are overwhelmingly complexes involving chrome, cobalt is also used, iron and aluminum have been examined as less polluting alternatives, and copper complex dyes are used on cotton.) 1:1 dye-metal complexes were introduced in the early twentieth century, 2:1 complexes in mid-century. Chromium has a coordination number of six: the dye molecule in 1:1 chrome complexes provides three ligands: the remaining three ligand sites are occupied typically by water. It has been suggested that these sites can interact with groups on the wool fiber, although evidence is lacking. If they do so, however, this would represent one of the few instances where the common mental model of mordant as a bridge linking dye and fiber is relevant. The metal atom in a 2:1 complex has no such spare binding capacity and achieves fastness via large molecular size and low solubility. In theory, either 1:1 or 2:1 complexes might be formed by a mordant dyeing process, but evidence suggests that the 2:1 complexes predominate. Whether produced on the fiber or as a metal-complex dye, the resulting complex is more stable to light and has limited solubility and thus good fastness to wet treatments.

The traditional method of mordant dyeing of (natural) dyes on wool involved pretreating the wool with a solution of the metal salt and subsequently introducing the mordanted wool into a

dyebath. This sequence is still used for dyeing natural dyes on wool today. When synthetic (mordant) dyes were developed, their varied chemistry could generate the range of shades required, and the need for multiple metal mordants declined: chrome (as dichromate) provided the best fastness and this became the main choice for mordant dyeing using the “chrome mordant method” and mordant dyes were not referred to as such, but as “chrome dyes” [20, 21]. Interest in making the processes more efficient took advantage of the independent substantivity of both dye and mordant. They could be applied together (the “metachrome method”) in a manner reminiscent of the chrome-logwood black dyeing of cotton. Eventually the process in which the dye was applied first and subsequently treated in a bath of chrome became the most widely used as the “afterchrome” or “topchrome” method. As discussed above, this is technically indistinguishable from the aftertreatment of a direct dyeing on cotton with a copper salt.

Synthetic Fibers

Given the similar dyeability of nylon to wool, and the search for fastness in nylon dyeings, it is not surprising that chrome dyes have been applied to nylon. The lack of reducing groups in nylon means that a reducing agent is included in the process to convert Cr^{VI} to Cr^{III} [22]. Chrome dyes are rarely applied to nylon today. In an interesting twist, the tannic acid/antimony mordant for basic dyes on cotton has been applied as an aftertreatment (“back-tanning”) to improve the fastness of acid dyes on nylon, although no interaction with the dye is suggested.

Acrylic fibers, introduced in the 1950s, were initially difficult to dye in dark shades with acceptable fastness. Before the development of modified basic dyes, an early solution was the so-called cuprous ion method in which the fiber was treated with a solution of copper ions: they are absorbed, and the fiber is then dyeable to dark shades with acid and direct dyes. A number of variations were published, including the application of dye and copper ions simultaneously and

the use of reducing agents, but the process was superseded before becoming a standard dyeing method. This is clearly a mordant dyeing technique, although none of the descriptions of the time refer to it as such [23].

Conclusion

Mordants have a long history in enhancing the coloration of textile fibers. They are essential in the dyeing of cellulosic fibers with natural and basic dyes and improve the fastness and color range obtainable on wool with natural dyes. With the advent of synthetic dyes, their use became restricted to the chrome dyeing of wool, and they have found little value in any dyeing of synthetic fibers. They are rarely used in large-scale commercial dyeing today, but are widely used among home and craft dyers for whom the use of natural dyes remains popular.

While the details may sometimes be hard to unravel from historic or unscientific descriptions/interpretations, the majority of mordants are metal ions (most notably of Cr, Al, Fe, Sn, Cu) that react with dyes to form coordination complexes of low solubility and better fastness. Other useful mordants are those that interact ionically with dyes, again forming a lower solubility complex.

Cross-References

- [Coloration, Fastness](#)
- [Coloration, Textile](#)

References

1. O'Neill, C.: *A Dictionary of Dyeing and Calico Printing*. Baird, Philadelphia (1869)
2. Shore, J.: *Colorants and Auxiliaries*, 2nd edn. Society of Dyers and Colourists, Bradford (2002)
3. Society of Dyers and Colourists, American Association of Textile Chemists and Colorists: *Colour Index*. Society of Dyers and Colourists, Bradford (1956); American Association of Textile Chemists and Colorists, Lowell

4. Johnson, A. (ed.): *The Theory of Coloration of Textiles*. Society of Dyers and Colourists, Bradford (1989)
5. Bancroft, E.: *Experimental Researches Concerning the Philosophy of Permanent Colours and the Best Means of Producing them*. Cadell and Davies, London (1813)
6. Pailthorpe, M.: The theoretical basis for wool dyeing. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 52–87. Society of Dyers and Colourists, Bradford (1992)
7. Hofenk-de Graaff, J.H.: *The Colorful Past. Origins, Chemistry and Identification of Natural Dyestuffs*. Archetype, London (2004)
8. Lee, D.: *Nature's Palette: The Science of Plant Color*. University of Chicago, Chicago (2007)
9. Brunello, F.: *The Art of Dyeing in the History of Mankind*. NeriPozza, Venice (1973)
10. Knecht, E., Fothergill, J.B.: Mordants, etc. In: Knecht, E., Fothergill, J.B. (eds.) *The Principles and Practice of Textile Printing*, 4th edn, pp. 203–269. Griffin, London (1952)
11. Hummel, J.J.: Mordants. In: Hummel, J.J. (ed.) *The Dyeing of Textile Fabrics*, pp. 156–247. Cassel, London (1885)
12. Sansone, A.: *The Printing of Cotton Fabrics Comprising Calico Bleaching*, pp. 129–130. Printing and Dyeing, Heywood (1887)
13. Bide, M.J.: Secrets of the printer's palette. In: Welters, L.M., Ordenez, M.T. (eds.) *Down by the Old Mill Stream: Quilts in Rhode Island*, pp. 83–121. Kent State University, Kent (2000)
14. Gordon, P.F., Gregory, P.: *Organic Chemistry in Colour*. Springer-Verlag, Berlin (1983)
15. Matthews, J.M.: *Application of Dyestuffs*. Wiley, New York (1920)
16. Lewis, D.M., McIlroy, K.A.: The chemical modification of cellulosic fibers to enhance dyeability. *Rev. Prog. Col.* **27**, 5–17 (1997)
17. Lewis, D.M., Pailthorpe, M.T.: The modification of wool with reactive hydrophobes. *J. Soc. Dy. Col.* **100**, 56–63 (1984)
18. Burkinshaw, S.M.: Dyeing wool with metal-complex dyes. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 196–221. Society of Dyers and Colourists, Bradford (1992)
19. Zollinger, H.: *Color Chemistry: Syntheses, Properties and Applications of Organic Dyes and Pigments*. VCH, New York (1987)
20. Bird, C.L.: *The Theory and Practice of Wool Dyeing*, 3rd edn. Society of Dyers and Colourists, Bradford (1963)
21. Duffield, P.: Dyeing wool with acid and chrome dyes. In: Lewis, D. (ed.) *Wool Dyeing*, pp. 176–195. Society of Dyers and Colourists, Bradford (1992)
22. Burkinshaw, S.M.: *Chemical Principles of Synthetic Fibre Dyeing*. Blackie, London (1995)
23. Wilcock, C.C., Ashworth, J.L.: *Whittaker's Dyeing with Coal Tar Dyestuffs*, 6th edn. Balliere, Tindall and Cox, London (1964)

Coloration, Textile

Renzo Shamey

Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA

Synonyms

Colouring, Coloring, Dyeing, Mass pigmenting, Printing

Definition

Dyeing can be described as the uniform application of colorant(s) to a coloring medium. The coloring of textiles may involve mass pigmenting (involving compounding), dyeing, and printing processes. The coloring medium in textile dyeing may take different physical forms (such as loose fiber, yarn, tow, top, woven, nonwoven and knitted substrates in open width or rope form), whereas in printing colorants are added to selected regions of the medium which is usually in a fabric form. Dyeing of homogenous fibers should result in a uniform solid color. Multicolored effects may be obtained by dyeing a blend of different fibers, bi- or multicomponent synthetic fibers, or via multiple colorant or illuminated discharge printing. In a technique known as space dyeing, colorant(s) can be applied from nozzles that inject different dyes to yarns and force steam in the vessel, thus generating a multicolored pattern.

There are difficulties encountered in controlling the physicochemical changes that occur during dyeing when attempting to maximize color yield, levelness of dyeing, color fastness, etc. Therefore, a vast sub-technology of specialty chemical auxiliaries is used in preparation for dyeing and in the dyeing process itself, such as levelling agents, dispersing agents, antifoams, etc., as well as auxiliaries specifically designed

for aftertreatments. Moreover, from the practical standpoint, methods of applying colorants to the coloring medium can be broadly divided into batch, semicontinuous, and continuous processes. These are discussed in more detail in the following sections.

Dyeing Process

The science of coloration entails several domains including chemistry, physics, physical chemistry, mechanics, fluid mechanics, thermodynamics, and others. Devising the most efficient dyeing process typically involves several concerns including machinery design, preselection of dyes of compatible properties, use of pH versus time profiles, selection of liquor ratio, flow rate and flow direction reversal times, and design of temperature versus time profiles. It is thus clear that dyeing is a complicated process with many different phenomena occurring simultaneously that require a suitable level of control.

When a textile fiber comes in contact with a medium containing a suitable dye under suitable conditions, the fiber becomes colored, the color of the medium decreases and that of the fiber increases, thus resulting in the dyeing of the substrate. True dyeing occurs when the dye is absorbed with a decrease in the concentration of dye in the dyebath and when the resulting dyed material possesses some resistance to the removal of dye by washing [1], rubbing, light, and other agencies.

Coloration of textiles is not limited to the simple impregnation of the textile fiber with the dye that occurs during the initial phase of the dyeing process. A dye is taken up by a fiber as a result of the chemical or physical interactions between the colorant and the substrate. Many dyes for textiles are water soluble, and their molecules are dissociated into positively and negatively charged ions in aqueous environments. The uptake of the dye by the fiber will depend not only on the nature of the dye and its chemical constitution but also on the structure and morphology of the fiber.

Textile Fibers

A fiber is characterized by its high ratio of length to thickness and by its strength and flexibility. Fibers may be of natural origin or formed from natural or synthetic polymers. They are available in a variety of forms. Staple fibers are short, with length-to-thickness ratios around 10^3 – 10^4 , whereas this ratio for continuous filaments is at least several millions [2]. The form and properties of a natural fiber such as cotton are fixed, but for man-made fibers, a wide choice of properties is available by design. The many variations include staple fibers of any length, single continuous filaments (monofilaments), or yarns comprised of many filaments (multifilaments). The fibers or filaments may be lustrous, dull or semi-dull, coarse, fine or ultra-fine, circular or of any other cross section, straight or crimped, regular or chemically modified, or solid or hollow.

From a processing standpoint, natural fibers have a number of inherent disadvantages. They exhibit large variations in staple length, fineness, shape, crimp, and other physical properties, depending upon the location and conditions of growth. Animal and vegetable fibers also contain considerable and variable amounts of impurities which must be removed prior to commencing dyeing. Man-made fibers are much more uniform in their physical characteristics. Their only contaminants are small amounts of slightly soluble low M_w polymer and some surface lubricants and other chemicals added to facilitate processing. These are relatively easy to remove compared to the difficulty of purifying natural fibers.

Water absorption is one of the key properties of a textile fiber that influences their coloration. Protein or cellulosic fibers are hydrophilic and absorb large amounts of water, which causes radial swelling. Hydrophobic synthetic fibers, such as polyester, however, absorb almost no water and do not swell. The hydrophilic or hydrophobic character of a fiber influences the types of dyes that it will absorb. The ability to be dyed to a wide range of hues and depths is a key requirement for almost all textile materials.

Another important property of a textile fiber is its moisture regain, which is the mass of water

absorbed per unit mass of completely dry fiber, when it is in equilibrium with the surrounding air, at a given temperature and relative humidity. The regain increases with increase in the relative humidity but diminishes with increase in the air temperature. Water absorption by a fiber liberates heat (exothermic) and will therefore be less favorable at higher temperatures. The heat released is often a consequence of the formation of hydrogen bonds between water molecules and appropriate groups in the fiber. When the final regain is approached by drying wet swollen fibers, rather than by water absorption by dry fibers, the regain is higher. For hydrophilic fibers such as wool, cotton, and viscose, the relatively high regain values significantly influence the gross mass of a given amount of fiber. This is significant in dyeing. Amounts of dyes used are usually expressed as a percentage of the mass of material to be colored. Thus, a 1.0 % dyeing corresponds to 1.0 g of dye for every 100 g of fiber, usually weighed under ambient conditions. For hydrophilic fibers, the variation of fiber mass with varying atmospheric conditions is therefore an important factor influencing color reproducibility in repeat dyeings.

Stages of Coloration Process

Most textile dyeing processes initially involve transfer of the colored compound, or its precursor, from the aqueous solution onto the fiber surface, a process called *adsorption*. From there, the dye may slowly *diffuse* into the fiber. This occurs through pores or between fiber's polymeric chains, depending on the internal structure of the fiber. The overall process of adsorption and penetration of the dye into the fiber is called *absorption*. Absorption is a reversible process. The dye can therefore return to the aqueous medium from the dyed material during washing, a process called *desorption*. Besides direct absorption, coloration of a fiber may also involve precipitation of a dye inside the fiber, or its chemical reaction with the fiber. These two types of processes result in better fastness to washing, because they are essentially irreversible.

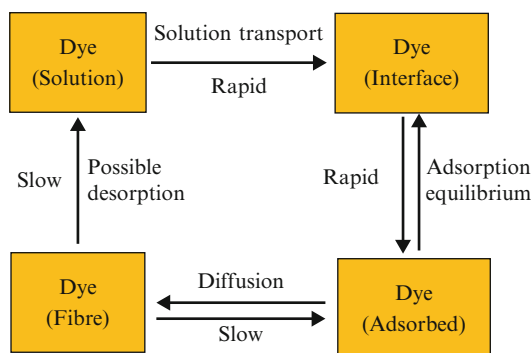
Dye Transport from the Bulk Solution to the Fiber Surface

Dyeing is a process which takes time. The transfer of a dye molecule from the dye solution into a fiber is usually considered to involve the initial mass transfer from the bulk dye solution to the fiber surface, adsorption of the dye on the fiber surface, followed by diffusion of the dye into the fiber as depicted in Fig. 1.

When a fibrous assembly is immersed in a dye solution, the rate at which the dye is taken up is generally dependent upon the extent to which the liquor is agitated and tends to approach a maximum value when the stirring is vigorous.

The transfer of dye from the bulk solution to fiber surface is fast, and the rate generally increases with increasing the flow rate. The adsorption equilibrium is also rapid, so it is usually assumed that the overall rate of dyeing depends on the rate of diffusion of the dye into the fiber. Inadequate control of the rate of dye adsorption will result in unlevel dyeings unless the dye can subsequently migrate from deeply dyed to lightly dyed regions of the substrate [1]. Therefore, the control of the first two stages of the process, namely, the initial mass transfer from the bulk solution to the fiber surface and adsorption of the dye on the surface, is important for a level dye distribution throughout the substrate to be achieved.

A fundamental issue in dyeing is to ensure that the dye liquor penetrates all parts of every fiber



Coloration, Textile, Fig. 1 Dye transfer from bulk solution into a fiber [3]

and is distributed within the substrate as evenly as possible. In practice, however, there may be differences in fibers, to a greater or lesser extent, because of natural or process-related parameters. These variations are of major importance, since, in wet processes, changes in regional density of the substrate result in variations in the degree of dye penetration and differences in flow rate which can lead to shade differences in dyeing.

The influence of agitation in increasing the rate of dye uptake is dependent in large measure on the hydrodynamic complexity of the system. Unfortunately, flow through the substrate cannot be described in any simple fashion with fibrous assemblies, due to the extreme complexity of defining the flow of liquor through a mat of fibers or through yarns or cloth.

In spite of the complications of real systems, some of the principles governing the process may be elucidated by consideration of simple ones, e.g., a plane sheet or film of material immersed in dye liquor whose direction of flow is parallel to the sheet. On making contact with the dye liquor, dye is adsorbed by the film so that the neighboring liquor becomes deficient in dye; dye is transported to the surface by dispersion from the bulk, but the quantity transferred is modified by the speed at which the liquor passes through the film.

Calculations of the rate at which the fibers can take up dye require knowledge of the flow pattern of the liquor before analysis of diffusion may be attempted. Experimental investigations of flow of dye liquor through masses of fiber lead to the conclusion that, at the common rates used in dyeing, the flow is streamline rather than turbulent, so attention may be confined to streamline conditions.

Intermolecular Forces Operating Between Colorants and Textile Material

The strongest dye-fiber attachment is that of a covalent bond. Another important interaction between dyes and fibers includes electrostatic attraction, which occurs when the dye ion and the fiber have opposite charges. Hydrogen bonds may also be formed between a specific range of

colorants and textile fibers. In addition, in nearly all dyeing processes, van der Waals forces and hydrophobic interactions are involved. The combined strength of the molecular interactions is referred to as the *affinity* of the dye for the substrate. The substantivity of the dye is a less specific term and is often used to indicate the level of exhaustion (which is described in the following sections) [4]. Thus, substantivity is the attraction between the substrate and a dye under precise conditions where the dye is selectively extracted from the medium by the substrate. Different types of textile fibers require different kinds of dyes, and in general, dyes which are suitable for one type of fiber may not dye other types effectively.

Levelness

Levelness is the uniformity of dye distribution (and hence color) on textiles. Two fundamental mechanisms contribute to a level dyeing. One is the initial sorption of dye during the dyeing; the other is the migration of dye after initial sorption on the fiber. An initial level sorption will lead to a level dyeing. An unlevel sorption may be corrected if sufficient migration takes place. These mechanisms are affected by dyes and chemicals, by textile substrate, and by controlling the parameters of the dyeing process such as dyebath pH, liquor ratio, flow rate, and temperature. Some dyes are more likely to level out, especially if they are small and do not have a high degree of affinity towards the substrate; other dyes on the other hand tend to be much less likely to migrate. These are often dyes of large size or with strong affinity towards the substrate.

Dyeing of Various Textile Materials

Textile fibers can be dyed at various stages of production such as loose fiber, top, tow, yarn, fabric, or garment. The levelness requirements for dyeing loose fibers are less strict than for dyeing at the yarn stage or at later stages, since further processing of the loose fiber results in

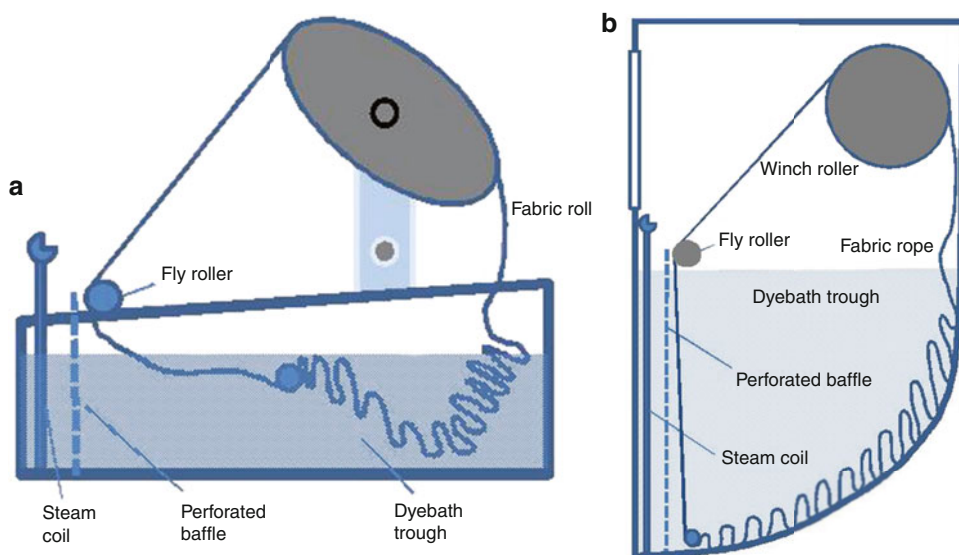
some mixing of the dyed fibers, e.g., during carding or gilling, which improves the levelness of the resultant color. For yarn dyeing levelness is more critical because while the yarn will be made into fabric, either by knitting or weaving, or used in the construction of carpets, any unevenness in the dyeing of the yarn will show in the finished goods. In the case of fabric and especially garment dyeing, the process will be less forgiving and variations will have to be remedied in different ways, either by overdyeing the substrate to a darker shade or by stripping the color and reapplying the colorant. Dyeing textiles at earlier stages of production such as loose fiber also reduces the degree of flexibility in response to market needs. Garment-dyed material can reach the market quickly and meet the rapidly changing demands of the market, whereas in the case of fibers or yarns, typically several additional processes have to be completed before the dyed material reaches the consumer. In the case of substrates with striped patterns yarn dyeing is quite common. When blending dyed yarn with undyed yarn to produce a woven pattern, it must be considered that subsequent bleaching and scouring may be needed. Thus, dyes of suitable fastness properties, to bleach and various agencies, have to be employed to ensure no staining of adjacent white regions occurs. Dyeing of fabrics and garments with variations in fabric density or stress can also result in systematic unlevelness. A common type of unlevelness in dyeing of nylon fabrics is known as Barré which appears as stripes across the substrate. A specific visual appearance test method to rate the degree of Barriness of dyed substrate has been introduced over the years.

Batch, Semicontinuous, and Continuous Coloring Processes

There are three main types of processes for the dyeing of textile materials: batch, continuous, and semicontinuous. Batch processes are the most common method used to dye textile materials and often depend on the type of equipment available and the weights or lengths of the material to

be dyed. Batch dyeing is often called exhaust dyeing because the dye is gradually transferred from a relatively large volume dyebath to the material being dyed over a long period of time. Batch dyeing, therefore, involves applying a dye from a solution or a suspension at a specific ratio of liquor to textile substrates where the depth of the color obtained is mainly determined by the amount of colorant present in relation to the quantity of fiber (known as L:R or liquor to goods ratio), although other factors can also influence the overall dye uptake. Batch processes are thus designed for specific quantities of substrate from few grams to several hundred kilograms. Batch dyeing of most natural fibers is carried out under atmospheric pressure conditions. Operations may also be carried out under elevated pressures. For instance, in the batch dyeing of polyester substrates, high temperatures (HT), around 125–130 °C, and correspondingly elevated pressures are required if the use of carriers is to be avoided. High-pressure beams, jets, and jigs can be used. Although such equipments tend to be more expensive than conventional atmospheric pressure equipments, their cost is more than offset by the omission of carriers. At high temperatures the diffusion rate of the dyes is high enough to produce satisfactory shades in a dyeing time of about one hour. Batch dyeing of polyester, however, can give rise to some bruises and pillings due to hydraulic and mechanical impacts during high temperature dyeing process. Generally, flexibility in color selection in batch dyeing is high, but the cost of dyeing is lower the closer the dye application is to the end of the manufacturing process for a textile product. Two examples of a common type of batch dyeing machine known as beck or winch (shallow-draft and deep-draft winches) are given in Fig. 2.

Semicontinuous dyeing is carried out in a continuous range where the substrate is fed into the dyeing range from one end and collected at the other. In semicontinuous processes typically fixation and washing steps are carried out discontinuously. Pad rolls transfer the colorant that is picked up from a trough into the substrate, and the process is based on impregnation of the substrate followed by fixation of the colorant. The pressure



Coloration, Textile, Fig. 2 (a) Shallow-draft winch dyeing machine. (b) Deep-draft winch dyeing machine

applied by rolls to impregnate the substrate can be adjusted to squeeze the excess liquor out of the substrate and obtain the required wet pickup percentage. Since the initial uniformity of dye deposition on the substrate is critical for a level dyeing, padding rolls responsible for the transfer of the colorant onto the substrate must have a uniform surface with no indentations. The pressure across the rolls should also be adjusted and controlled regularly to ensure the uniformity of color transfer onto the substrate. The dyeing range may include cans or alternative forms of dryers to prevent migration of dye across the substrate prior to fixation. At the end of the range, the substrate may be rolled onto a beam, covered with plastic sheets, and kept overnight, to enable fixation of the dye, or transported to other sections for subsequent fixation, washing, and aftertreatment operations. Pad-batch dyeing is a specific type of semicontinuous coloration process which is common in the application of reactive dyes to cellulosic substrates. In this process the batch is left overnight to enable the colorant to react with the substrate. Other forms of pad-fix processes may include pad-bake and pad-steam.

Continuous dyeing refers to operations at constant composition involving several application

and wash boxes (troughs), where a long length of textile fabric is pulled through each stage of the dyeing process including fixation and aftertreatment. Continuous dyeing operations are common when dyeing large quantities of substrate. This is often carried out on cotton and its blends with synthetic fibers such as polyester. In continuous dyeing processes, fixation and subsequent wash and rinse operations are combined with the coloration process to enable rapid throughputs. Thermofixation (commercially known as Thermosol) which is commonly carried out on polyester and polyester blends is particularly suited to continuous dyeing processes, since it involves padding the dyestuff from dispersion onto the fabric, drying, and then heating the padded substrate to a temperature of 180–220 °C. A specific category of disperse dyes capable of sublimation is employed for this purpose. At such temperatures, diffusion rates of sublimated dyes are so high that a few seconds suffice for adequate penetration of dye molecules into the substrate. Several variables, however, affect the final shade of the dyed fabric during the thermofixation process. Some of these variables are Thermosol period, temperature, type of disperse dyestuff, and pad bath auxiliaries.

In continuous printing processes, colorants are applied to specific sections of the cloth using a number of techniques that may include roller, flatbed, and rotary screen printing systems to obtain a preset design. Dye fixation is carried out by steaming or baking the printed material followed by washing to remove surplus dye, thickeners and any other auxiliaries.

The details of the dyeing process can vary considerably between different types of textile materials employing different types of dyeing equipment. For example, the maximum permissible rate at which the temperature of the bath can be raised may be determined by the relationship between the rate of circulation of the dye liquor and the rate of transfer of the dye from bath to fiber such as in the dyeing of yarn in hank dyeing machines.

The criteria for choosing a dyeing process vary and may include the following:

- Shade range
- Fastness requirements
- Quality requirements and control
- Cost
- Equipment availability
- Dye selection

The aim of a successful dyeing process is to achieve the desired shade, at the right price, with sufficient levelness, whether dyeing loose fiber, yarn, or piece goods, with sufficient color fastness to withstand both processing and consumer demands, but without adversely affecting the fiber quality. Of these, an acceptable level of uniform dye uptake at all parts of the substrate may be the most important criterion.

A typical dyeing process may be divided into several steps as follows:

- Establishment of equilibrium between associated molecular dye and single molecules of dye in solution
- Diffusion of monomolecular dye to the diffusional boundary layer at the fiber surface
- Diffusion of dye through the boundary layer at the fiber surface
- Adsorption of dye at the fiber surface

- Diffusion of the dye into the fiber interior
- Desorption and readsorption of dyes (migration)

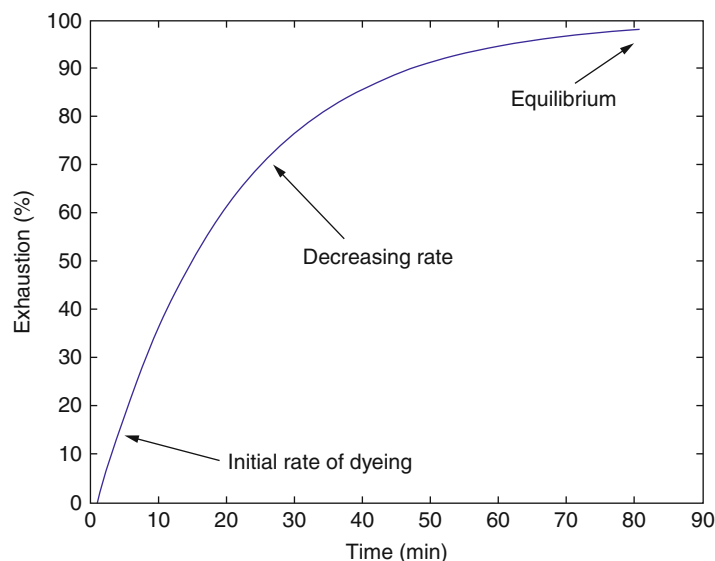
These steps form a reversible equilibrium system. Each of the six steps can influence the levelness of dyeings. A complete quantitative analysis of the effects of many factors which influence the levelness of dyeing would require the development of a mathematical model involving a significant number of parameters. This is, however, a very difficult task. For practical applications, one may initially try and identify a few variables, which are thought to have a larger impact on levelness and restrict the development of the model to the effects of these few most important variables.

The Donnan equation involves nine factors: the concentration of dye in fiber, the concentration of dye applied, the liquor ratio, the distribution of ions between solution and fiber, the ionic charge on the dyestuff molecule, the internal volume of the fiber, the affinity of the dye, the gas constant, and the dyeing temperature. All together, these terms describe the dyebath conditions.

To obtain reproducible dyeings, whether this is on a laboratory, pilot plant, or bulk scale, the following factors in the dyeing process must be controlled or measured:

- Quality of water supply
- Preparation of substrate
- Dyeability of substrate
- Weight of substrate
- Moisture content of substrate at weighing
- Selection of dyes
- Standardization of dyes
- Weighing of dyes and chemicals
- Dispensing method for dyes and chemicals
- Moisture content of dyes
- Liquor to goods ratio
- Dyebath additives
- pH of dyebath
- Machine flow and reversal sequence
- Time/temperature profile

Various workers have placed different degrees of emphasis on each of the above factors and also

Coloration, Textile,**Fig. 3** Dyebath exhaustion as a function of time [5]

on the factors which are not listed above. There may also be disagreement on the mechanism by which any one of these factors operates. Furthermore, it should be noted that the parameters in this list are not quite independent, and in some cases the effect of two of these factors may be considered as one effect.

Exhaustion

Dyeing is carried out either as a batch exhaustion process or as a continuous impregnation and fixation process. In exhaust dyeing, all the material is in contact with all the dye liquor from where the fibers absorb the dyes. The dye concentration in the bath therefore gradually decreases. The degree of dyebath exhaustion as a function of time describes the rate and extent of the dyeing process. For a single dye, the exhaustion is defined as the mass of dye taken up by the material, divided by the total initial mass of dye in the bath. For a bath of constant volume, this can be expressed by Eq. 1:

$$\% \text{Exhaustion} = (C_0 - C_t) / C_0 \quad (1)$$

where C_0 and C_t denote the concentrations of dye in the dyebath initially and at some time, t , during the process, respectively.

Exhaustion curves, such as that shown in Fig. 3, may be determined at a constant dyeing temperature, or under conditions where the temperature and other dyeing variables are changing. For many dyeings, a gradual increase of the dyeing temperature controls the rate of exhaustion, aided possibly by the addition of chemicals such as acids or salts. In cases where the dyes in the deeply dyed fibers are not able to desorb into the bath and then be redistributed onto paler fibers, such control is essential to ensure that the final color is as uniform as possible. Such redistribution of dyes is called migration.

The slope of a dyeing exhaustion curve (Fig. 3) defines the rate of dyeing at any instant during the process. The rate of dyeing gradually decreases until, if dyeing is continued long enough, an equilibrium is reached where no more dye is taken up by the fibers. There is now a balance between the rates of absorption and desorption of the dye. The equilibrium exhaustion is the maximum possible exhaustion under the given conditions. The lack of any further increase in exhaustion does not necessarily mean that a true equilibrium exists. It is possible for the dye in solution to be in equilibrium with the dye located on the outer surfaces of the fibers. True equilibrium only exists when the dye in solution is in equilibrium with the dye that has fully penetrated into the center of the fibers. Dyeings rarely continue to this point since it may

take a relatively long time to attain. In fact, many commercial dyeings barely reach the point of constant exhaustion.

There are two basic methods of achieving a level exhaustion dyeing in any dye/fiber system; the first is by dye migration, and the second is by controlled dye exhaustion. The first method involves exhausting all of the dye onto the fiber and then allowing it to migrate between the fibers in order to “level out” the dyeing. These are dyes which are able to migrate from the fiber back into the liquor and then transfer back to the fiber. This redistribution of dye improves the levelness of the dyeing and normally takes place when the dye liquor is at the boil. In this method the dye is not completely exhausted onto the substrate, and this can lead to poor reproducibility of color, and hence, additions of dye to correct the final shade are often necessary.

The second method is to ensure that the dye is exhausted in a level manner from the start of the dyeing. In this method, the dyeing rate is controlled by changing the parameters of the dye bath at a controlled rate so that the dye is deposited on the yarn in a uniform manner throughout the substrate. Careful control of these parameters, such as dyeing temperature, pH, or amount of electrolyte and flow rate, is often necessary to obtain level, well-penetrated dyeings. This is essential if the dye initially absorbed is unable to migrate from heavily dyed to poorly dyed areas during the process.

Exhaustion Profiles

Variation of the concentration of dye in the dyebath during the dyeing is referred to as the exhaustion profile, and the shape of this profile has been believed by many researchers to be the most determining factor in levelness of dyeing.

Exhaustion control has been developed theoretically and in the laboratory by several workers. These workers used knowledge of the dyeing kinetics to devise a time/temperature profile to give a particular exhaustion profile; others have attempted a direct control of the exhaustion rate.

Studies of the theoretical basis of the relationship between levelness of dyeing and the rate of dye uptake by textile substrates were initiated in the 1950s and the 1960s. Since then there have been many investigations into the methods of controlling the exhaustion of dye bath in order to improve levelness. This has been an area of much disagreement among researchers.

Linear Exhaustion Profiles

Carbonell et al. [6] developed a mathematical representation of various exhaustion profiles and went on to calculate practical time/temperature profiles that would result in linear exhaustion profiles. Later work aimed at establishing detailed kinetic relationships in order to carry out “isoreactive” dyeings, in which the dyeings have a linear exhaustion profile.

Cegarra et al. [7] later modified this approach to apply it to dyeings that used continuous addition (or integration) of dye into the dyebath. These dyeings were carried out at constant temperature, using a predetermined dye addition profile to achieve linear exhaustion. This method was defined as Integration Dyeing, which can be used to control the dye absorption during the integration, so as to avoid the possibility of initially fast and anomalous absorption, which may cause unlevel dyeings. In practice, this method is often used to improve the levelness, when all the dyes are added at the beginning of the process.

Several authors have stated that linear exhaustion is most likely to give a level result and developed a control strategy for automation of a dyeing machine such that the percentage exhaustion per circulation never rises above the critical value for levelness.

Other Exhaustion Profiles

The use of linear exhaustion profiles for the control of dyeing process is by no means generally accepted. A number of researchers have stated that a rapid uptake at the start of the process, with a gradual slowing of the exhaustion

thereafter, should give a better result than a linear profile. The argument has been that the critical part of the exhaustion phase is the final phase, where the amount removed from the bath is large compared to that remaining, leading to a greater risk of unlevelness. Two profiles of this type, i.e., exponential and one with the exhaustion proportional to square root time, have been suggested.

Experimental work suggests that both exponential and square root profiles give a clear improvement over both a linear profile and a standard constant temperature ramp dyeing. It has been recommended to devise a reliable concentration monitoring system to control the exhaustion process. Despite the suggested strategies this is still an area requiring further research and development.

Practical Difficulties Involved in the Dyeing Process

There are some practical difficulties in achieving a quality dyeing when considering dyeing process control. These are briefly described in the following sections.

Dyeing Rate

Dyeing rates are of greater practical significance than the exhaustion at equilibrium because continuation of dyeing to equilibrium is not economical. Long dyeing times increase the risk of fiber damage and dye decomposition, particularly at higher dyeing temperatures. On the other hand, very rapid dyeings will usually result in the color being unlevel. This implies that dyeing processes should be neither too slow nor too fast. In order that dyes are used economically and as little as possible is wasted in the dyehouse effluent, the dyer prefers a high degree of exhaustion in a relatively short dyeing time. However, dyeing must be controlled so it is not so rapid that it is difficult to produce a level dyeing. If there is a strong affinity between dyes and fibers and the conditions are not controlled, a rapid strike of dyestuff will occur which will often result in unlevel dyeings. To control the rate of dye uptake, a number of dyeing parameters, including temperature, pH, electrolyte concentration, and agitation

among other parameters may have to be controlled.

The slope of the exhaustion curve gives information on the rate of dyeing. The determination of these curves, however, requires much work, and they are dependent on the dyeing conditions and the nature of the goods. The dyeing rate is influenced by the temperature and by chemicals such as salts and acids, all of which also influence the final exhaustion. A clear distinction of the effects of process variables on the dyeing rate and on the final exhaustion at equilibrium is essential in the successful control of the dyeing process.

Initial Stage of Dyeing

The initial rate of dyeing (the initial slope of exhaustion versus time) is called the strike. Rapid strike by a dye often results in initial unlevelness and must be avoided for those dyes that cannot subsequently migrate from heavily to lightly dyed areas of the fabric. For dyes of rapid strike, the dyeing conditions must limit the initial rate of exhaustion and therefore improve the levelness of the dyeing. The strike depends on the dyeing temperature, the dyeing pH, and the presence of auxiliaries.

Even for dyes of moderate and low strikes, the objective of uniform dyeing of the fiber mass is rarely achieved during the initial stages of the operation. This is because of irregularities in the material's construction, in the fiber packing, and in the distribution of residual impurities, as well as differences in temperature and flow rate of the solution in contact with the fibers.

Dye/Fiber Types and Dye Migration

Different types of textile fiber require different kinds of dyes, and in general dyes which are suitable for one type of fiber will not dye other types effectively.

The degree of exhaustion of a dye at equilibrium is higher the greater the substantivity of the dye for the fiber being dyed. Often, a very substantive dye will give a high initial rate of absorption, or strike. Substantivity is the "attraction" between dye and fiber whereby the dye is selectively absorbed by the fiber and the bath becomes less concentrated.

The ability of a dye to migrate and produce a level color, under the given dyeing conditions, is obviously an important characteristic. It can overcome any initial unlevelness resulting from a rapid strike. Migration of the dye demonstrates that the dye can be desorbed from more heavily dyed fibers and reabsorbed on more lightly dyed ones. This is especially important in package dyeing where uniform color of the yarn throughout the package is essential.

While migration is important for level dyeing, it has two major drawbacks. Firstly, dyes with greater ability to desorb from dyed fibers during migration will usually have lower fastness to washing. Dyes of very high washing fastness are essentially nonmigrating dyes for which level dyeing depends upon very careful control of the rate of dye uptake by the material. The second problem with migrating dyes is that good migration may result in lower exhaustion, again because of their ability to desorb from the fibers.

Factors Affecting Levelness of Dyeing

As has been indicated dyeing is a very complicated process with different phenomena occurring simultaneously. Unlevelness can arise in many forms such as the unlevelness between sides, at ends of a fabric, or on the layers of a yarn package. The causes are as numerous as the effects. A quantitative analysis of the effects of many factors which influence the levelness of dyeing is a very difficult task since it would require the development of a mathematical model involving a significant number of parameters. Also, to investigate the transfer of dye through the substrate will involve the solution of nonlinear partial differential equations. Little work has been done on this aspect, according to the literature.

Flow Rate

During the dyeing process, the supply of dye through the solution to the surface of the fibers/yarns can occur in two ways, either by aqueous diffusion of dye through the liquor or by convective movements of the fluid which replace the depleted solution by fresh solution. Diffusion is a much slower process than the convective transport of dye, except at very low velocities of liquor flow.

However, an exact solution to the problem of convective diffusion to a solid surface requires first the solution of the hydrodynamic equations of motion of the fluid (the Navier-Stokes equations) for boundary conditions appropriate to the mainstream velocity of flow and the geometry of the system. This solution specifies the velocity of the fluid at any point and at any time in both tube and yarn assemblies. It is then necessary to substitute the appropriate values for the local fluid velocities in the convective diffusion equation, which must be solved for boundary conditions related to the shape of the package, the mainstream concentration of dye, and the adsorptions at the solid surface. This is a very difficult procedure even for steady flow through a yarn package of simple shape.

Measurement of Levelness

Although objective measurements have been proposed, the measurement of levelness and its causes is difficult. The levelling ability of dyes is routinely tested under a fixed set of circumstances, but the effect of changing circumstances is less often reported. Similarly, a distinction between levelness of strike and levelness from migration is not usually studied. A relatively simple means of determining levelness is via colorimetric measurement of dyed object in several locations representative of the entire substrate. A color difference metric may be used to determine the degree of variability across the substrate, and a tolerance volume may be established for this purpose. Visual assessments are also common, although these would be subjective and open to interpretation. A common procedure is to place side center side of a full width of dyed fabric to determine variations from center to sides due to variations in processing conditions. Such variability is called tailing. Another common type of variability in continuous operations is due to variability in the composition of the trough during the dyeing process that may generate shade variation between the beginning and the end of a roll, commonly denoted ending. Some online measurement systems have been proposed and used in some cases, in which digital imaging or colorimetric systems are installed in a continuous processing

line which can alert the operator to “larger than expected” variations in color during the operation.

Dyes

A dye is a substance capable of imparting its color to a given substrate, such as a textile fiber. A dye must be soluble in the application medium, usually water, at some point during the coloration process. It must also exhibit some substantivity for the material being dyed and be absorbed from the aqueous solution.

For diffusion into a fiber, dyes must be present in the water in the form of individual molecules. These are often colored anions, for example, sodium salts of sulfonic acids. They may also be cations such as mauveine or neutral molecules with slight solubility in water, such as disperse dyes. The dye must have some attraction for the fiber under the dyeing conditions so that the solution gradually becomes depleted. In dyeing terminology, the dye has substantivity for the fiber and the dyebath becomes exhausted.

The four major characteristics of dyes are:

1. Intense color
2. Solubility in water at some point during the dyeing cycle
3. Substantivity for the fiber being dyed
4. Reasonable fastness properties of the dyeing produced

The structures of dye molecules are complex in comparison with those of most common organic compounds. Despite their complexity, dye structures have a number of common features. Most dye molecules contain a number of aromatic rings, such as those of benzene or naphthalene, linked in a fully conjugated system. This means that there is a long sequence of alternating single and double bonds between the carbon and other atoms throughout most of the structure. This type of arrangement is often called the chromophore or color-donating unit. The conjugated system allows extensive delocalization of the p electrons from the double bonds and results in smaller differences in energy between the occupied and unoccupied molecular orbitals for these electrons. At least five or six conjugated double bonds are

Coloration, Textile, Table 1 Classification of dyes according to chemical constitution and usage

Classification of dyes according to chemical constitution	Classification of dyes according to textile usage
Azo	Direct
Anthraquinone	Azoic
Heterocyclic	Vat
Indigoid	Sulfur
Nitro	Reactive
Phthalocyanine	Acid
Polymethine	Basic (cationic)
Stilbene	Disperse
Sulfur	Mordant (metal complex)
Triarylmethane	

required in the molecular structure for a compound to be colored.

Table 1 gives partial classifications of dyes as presented in the Colour Index International [8]. In order to gain an optimum result, the appropriate dye class for the fiber must be used, along with specific dyeing conditions. The ten major dye classes involve acid, metal complex, mordant, direct, vat, sulfur, reactive, basic, disperse, and azoic dyes. Some of ten major dye classes shown in Table 1 can be used to dye the same fiber type, but varying conditions are required. For example, acid, metal complex, mordant, and reactive dyes can all be used to dye wool. However, there may be one type of dye that is preferred for a certain dyeing process, for example, disperse dyes for polyester fibers.

There are numerous factors involved in the selection of dyes for coloring textile materials in a particular shade. Some of these are:

- The type of fibers to be dyed
- The form of the textile material and the degree of levelness required – level dyeing is less critical for loose fibers, which are subsequently blended, than it is for fabric
- The fastness properties required for any subsequent manufacturing processes and for the particular end use
- The dyeing method to be used, the overall cost, and the machinery available
- The actual color requested by the customer

The approximate relative annual consumption of the major types of fibers and dyes estimated in the year 2000 indicates that dyes used for cotton (the most widely used natural fiber) and for polyester (the most widely used synthetic fiber) dominate the market. In the case of cellulosic fibers including cotton, reactive dyes due to possessing excellent fastness properties upon fixation and demonstrating bright and brilliant shades occupy the lion's share of the market for this fiber category. Disperse dyes also occupy a large sector of the market due to their use on polyester fibers. Other colorants occupy smaller sections of the market and their applications are specific and less common.

Sorption Isotherms

In order to determine the thermodynamics of dye sorption by various fibers three main models have been proposed. These are known as Nernst, Langmuir, and Freundlich sorption isotherms [9]. These models determine the relationship between the concentration of dye in fiber and that in solution under isothermal dyeing conditions. In the simplest form the relationship is linear (Nernst). The Nernst isotherm indicates that the distribution of dye between fibers and bath is simply due to the partition of the dye between two different solvents until one becomes saturated. In a more complex scenario, dye "sites" are present in the fiber which result in an apparent saturation value. These sorption processes are defined by the Langmuir isotherm. In some cases an empirical power function, represented by the Freundlich model, can be used to determine the relationship between the amount of dye in solution and that in fiber. An examination of this model shows that the exhaustion of dye drops towards the end of the dyeing cycle. The drop in percent exhaustion with an increase in the depth of dyeing is well known and reflects the loss of activity of the dye in solution with increasing concentration. According to this model there appears to be no saturation value for the fiber, and as more and more dye is added to the bath, more and more is taken up. There is of course a practical limit as to how much dye may be placed on the fiber. An example of a system that can be

described by the Nernst isotherm is the dyeing of polyester fibers with disperse dyes. Acid dyes on wool or basic dyes on acrylic are attracted to specific dye sites with opposite charge, and these would exhibit Langmuir-type relationships. The adsorption of direct dyes on cotton may be described by a Freundlich model however, where no specific dye sites are present in fiber but a gradual decrease in the overall rate of dye adsorption is witnessed over time. In general Freundlich models are indicative of nonionic or relatively weak bonding possibilities between dye and fiber.

Summary

Many aspects of dyes and dyeing have not been covered. This is due to the complexity of the process and the large number of variables and processes involved. Textile coloration is a large industry, and a number of excellent resources are available that cover the fundamentals of coloration and the dyeing of specific types of fibers. The reader should refer to additional resources for a detailed examination of the topics.

Cross-References

- [Coloration, Fastness](#)
- [Coloration, Mordant Dyes](#)
- [Colorant, Natural](#)
- [Colorant, Textile](#)
- [Colorant, Environmental Aspects](#)
- [Dye](#)

References

1. Vickerstaff, T.: *The Physical Chemistry of Dyeing*, 2nd edn. Oliver and Boyd, Edinburgh (1954)
2. Morton, W.E., Hearle, J.W.S.: *Physical Properties of Textile Fibres*, 3rd edn. Textile Institute, Manchester (1993)
3. Shamey R., Zhao X., *Modelling, Simulation and Control of the Dyeing Process*, The Textile Institute, Woodhead Publishing, Cambridge (2014)
4. Johnson, A.: *The Theory of Coloration of Textiles*. Textile Institute, Bradford (1989)

5. Zhao, X.: Modelling of the Mass Transfer and Fluid Flow in Package Dyeing Machines, Ph.D. Thesis, Heriot-Watt University (2004)
6. Society of Dyers and Colourists: Colour Index International. Society of Dyers and Colourists, Bradford (1989)
7. Carbonell, J., Knobel, R., Hasler, R., and Walliser, R.: Mathematische Erfassung der Zusammenhänge zwischen Farbekinetik und Flottdurchfluss in bezug auf die Homogenität der Farbstoffverteilung auf der Faser in Systemen mit Flotenzirkulation. *Melliand Textilber.* **54**, 68–77 (1973)
8. Cegarra, J., Puente, P., Valldeperas, J., Pepio, M.: Dyeing by Integration. *Text Res J* **58**, 645–653 (1988)
9. Aspland, R.: Textile Dyeing and Coloration. American Association of Textile Chemists and Colorists (1997) Research Triangle Park, NC, USA

Coloring

► Coloration, Textile

Color-Magnitude Diagrams

Michael H. Brill
 Datacolor, Lawrenceville, NJ, USA

Definition

A color-magnitude diagram is a scattergraph of astronomical objects showing the relationship between each object's absolute magnitude and its estimated surface temperature or between optical or perceptual proxies for these quantities.

Historical Antecedents

Humankind has always wanted to understand the bodies in the night sky, and one step to understanding them is to categorize them.

The first tool available to assign these categories was the human visual system – the unaided eye. Hipparchus (c. 190 BCE–c. 120 BCE) developed a scale for stars based on visual brightness, which eventually became quantified as stellar

magnitude. The convention that dimmer stars have higher magnitude is a historical precedent that dates from Hipparchus (1). However, whereas Hipparchus attached magnitude 1 to the brightest star within each constellation, Ptolemy (c. 140 AD) refined the system so that the brightest stars had magnitude 1 and the barely visible stars had magnitude 6 [1].

From the time of Galilei (1564–1642) and Kepler (1571–1630), the unaided eye became augmented by telescopes, whose main function in astronomy is to gather light from a much larger area than is available at the pupil of the eye. Telescopes were at first very limited in their light-gathering ability but later developed enough aperture so that the stars became accessible to color vision. At that point, a new categorization was possible, in which color and brightness could be conjoined.

Of course, stars have spectra that are close to black-body radiators, and such radiators have only two physical variables: intensity and temperature [2]. In perceptual terms, the color ranges from reddish through bluish through yellow and white, on a curve in the chromaticity diagram called the black-body locus. It is therefore possible to render the visible character of a star by two numbers: the brightness (magnitude) and a single variable of color (or temperature).

Diagram for Stars

At the beginning of the twentieth century, aided by the mature technology of telescopes, Danish astronomer [7] and American astronomer [8] developed the first color-magnitude diagram, called the **Hertzsprung–Russell diagram (H-R diagram)** [3–5]. Originally the diagram was based on visual estimation of magnitude and color, and it was a research tool to help characterize stellar evolution before the mechanism of nuclear fusion was understood. After about the 1930s, the H-R diagram became based on objective measurements but was used less as a research tool and more as a way to illustrate the theoretically predicted evolution of stars through trajectories in the diagram.

The objective measurements that replaced direct human perception were as follows: A star's absolute magnitude is the attenuation (in factors $10^{-0.4}$) of the star Vega's power (as received at 10 pc viewing distance [32.6 light years]) to equal that of the star (also corrected to 10 pc).

A star's surface temperature is estimated in one of three ways: by the observed color (an old way), by a comparison of two sensor outputs such as blue and violet (a newer way), or by a model prediction of the temperature of a black-body radiator with the same radiation power per unit star-surface area (the most modern way). The third way requires independent inference of the star's radius but assumes the star has an emissivity of 1. To acknowledge the lack of compensation for the true emissivity, the temperature on an H-R diagram is called "effective temperature."

It is important to note the coordinate conventions of the H-R diagram: Temperature on a log scale (5) increases from right to left and magnitude (i.e., dimness) increases from bottom to top. Hence dim red stars (red giants) appear near the upper right of the diagram, and bluish bright stars (such as white dwarfs) appear in the lower left. A long cluster of stars called the Main Sequence extends from the upper left to lower right. Higher-mass stars occur at the upper left of this sequence, and the Sun appears approximately in the middle. The Main Sequence is composed of stars that are currently dominated by hydrogen that is fusing into helium. According to currently accepted theory of stellar evolution, such stars will eventually migrate either to the red-giant or white-dwarf domain of the H-R diagram.

H-R diagrams are often depicted in color, either pseudo-color with a thermal code to show the temperature or coded according to star categories such as cluster membership. No matter how the measurements were obtained, the data representation for the end user returns to visual perception as the way to see the color and brightness relationships of stars.

Diagram for Galaxies

One further stage in telescope evolution occurred in 1990 when the Hubble Space Telescope was

launched into orbit. The Hubble Space Telescope was able to render high-contrast images outside the Earth's atmosphere without encountering the absorption, scattering, and haze that beset Earth-bound telescopes. In digital images conveyed to earth from the Space Telescope, it was possible to see the colors, not only of stars, but of whole galaxies that are so distant as to be too dim to measure on the Earth. The color-magnitude diagram developed by Hertzsprung and Russell then evolved to accommodate galaxies. Thus was born the **galaxy color-magnitude diagram** [6]. The construction of such a diagram is similar to that of the Hertzsprung–Russell diagram, but the interpretation in terms of physical properties is not as precise.

To compute a galaxy's absolute magnitude, it is treated as a point-like object, whereupon its radiation is corrected (by the inverse-square law) to a distance of 10 pc, and the absolute magnitude is computed as the number of $10^{-0.4}$ attenuations of similarly compensated power from a reference to achieve a match.

The temperature is also the same "effective temperature" as is used in the H-R diagram. However, galaxies are visible from much farther away than individual stars, so a galaxy's recessional red shift exerts an appreciable influence on its effective temperature. Because light from farther galaxies (and greater red shift) requires more time for light to travel to the Earth, the galaxy color-magnitude diagram, if it were not red-shift-corrected, would depict an earlier universe at the red end of the temperature scale. However, in practice, the temperature is red-shift-corrected to be in the "rest frame" of the galaxy. In this case, the galaxies divide into two clusters: bright and reddish as opposed to dim and bluish. The existence of such clusters places strong constraints on theories of galactic formation.

Color-magnitude diagrams are not always defined in terms of fundamental physical properties such as those described above. In one embodiment, the galaxy color-magnitude diagram coordinates are defined by the log outputs of two sensors, typically called g and r, the sensors having different spectral sensitivities: the abscissa records r (which stands in lieu of the absolute

magnitude) and the ordinate records $g - r$ (which relates to the temperature). At the top of the diagram, one finds the red sequence of galaxies (elliptical galaxies), and at the bottom, one finds the blue cloud (spiral galaxies).

Cross-References

- [Galaxy Color Magnitude Diagram](#)
- [Hertzsprung-Russell Diagram](#)

References

1. Astronomical Society of South Australia: Stellar photometry. <http://www.assa.org.au/articles/stellarphotometry>. Accessed 30 May 2013
2. Wyszecki, G., Stiles, W.S.: Color Science, 2nd edn. Wiley, New York (1982)
3. Gribbin, J.: Companion to the Cosmos. Little, Brown, New York (1996)
4. Pasachoff, J.M.: A Brief View of Astronomy. Saunders College Publishing, New York (1986)
5. Wikipedia, Hertzsprung-Russell Diagram: http://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram. Accessed 30 May 2013
6. Shapley, A.E.: Physical Properties of Galaxies from $z = 2 - 4$. <http://arxiv.org/abs/1107.5060v2>. Accessed 26 Feb 2013
7. Hertzsprung, E.: Über die Sterne der Unterabteilung c und ac nach der Spektralklassifikation von Antonia C. Maury. *Astron Nachr* 179(4296), 373–380 (1909)
8. Russell, H.N.: Relations Between the Spectra and Other Characteristics of the Stars. *Pop Astron* 22, 275–294 (1914).

Color-Opponent Processing

- [Color Vision, Opponent Theory](#)

Color-Separation Overlay

- [Compositing and Chroma Keying](#)

Color-Word Interference

- [Stroop Effect](#)

Colouring

- [Coloration, Textile](#)

Colour-Opponent Processing

- [Color Vision, Opponent Theory](#)

Combustion Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Flame light lamps](#)

Definition

Lamps that produce light as a result of an exothermic reaction between the vapor of a solid, liquid, or gaseous fuel, consisting of hydrocarbons and oxygen.

Types of Combustion Lamps

Torches, oil lamps, candles, and gas lamps all are combustion types of lamps. The light comes from the flame that is the result of a reaction between oxygen and the vapor of a solid fuel in the case of candles, of a liquid fuel in the case of oil lamps, and of a gaseous fuel in the case of gas lamps. A spark is needed for starting the reaction.

Working Principle

In all combustion types of lamps, after a spark has initiated the process, the combustion reaction takes



Combustion Lamp, Fig. 1 Flame of a candle

place between the gaseous state of hydrocarbons of the fuel and oxygen from the air. The products of the reaction are carbon dioxide (CO_2), water vapor (H_2O), heat, and light in the form of a flame. In oil lamps and candles, the wick draws, by its capillary action, the fuel to the flame against the force of gravity. It is not the liquid part of that fuel that takes part in the combustion reaction but the evaporated fuel around the wick. With all combustion lamps, carbon particles of high temperature (soot), resulting from incomplete combustion, are brought to incandescence and contribute to the total light radiation. The blue light at the bottom of the wick (Fig. 1) is the result of the combustion reaction, the yellowish part of the flame is the result of incandescent light from the carbon particles.

The combustion reaction vitiates the atmosphere by consuming oxygen and returning carbon dioxide (CO_2). Non-complete combustion leads to the emission of harmful gasses like carbon monoxide (CO), sulfur oxides (SO_x), and nitrogen oxides (NO_x).

Oil Lamps

Oil lamps represent the oldest form of artificial light. Scraped-out stone oil reservoirs (Fig. 2)



Combustion Lamp, Fig. 2 Stone oil lamp found in Lascaux, France, 17000 BC (Photograph: Sémhur: Creative Commons 3.0 unported)

with a wick have been in use for lighting purposes since prehistoric times some 20,000 years ago [1, 2]. The basic conception of oil lamps has remained unchanged although the materials used and the type of construction have been advanced quite a bit. The Dutchman Jan van der Heyden, for example, developed in 1663 a closed oil reservoir for street lanterns [3]. He not only made a detailed description of the production process of the oil lamp and lantern but also of the set of maintenance material the lamp lighters crew needed (Fig. 3).

The industrial revolution some 200–250 years ago asked for artificial lighting in industrial premises, and in that period a boom in new developments in the technology of oil lamps is seen, followed by developments of completely new lighting products, such as gas lighting and later electrical lighting. Until the end of the nineteenth century, oil lamps have been in general use, especially for domestic lighting.

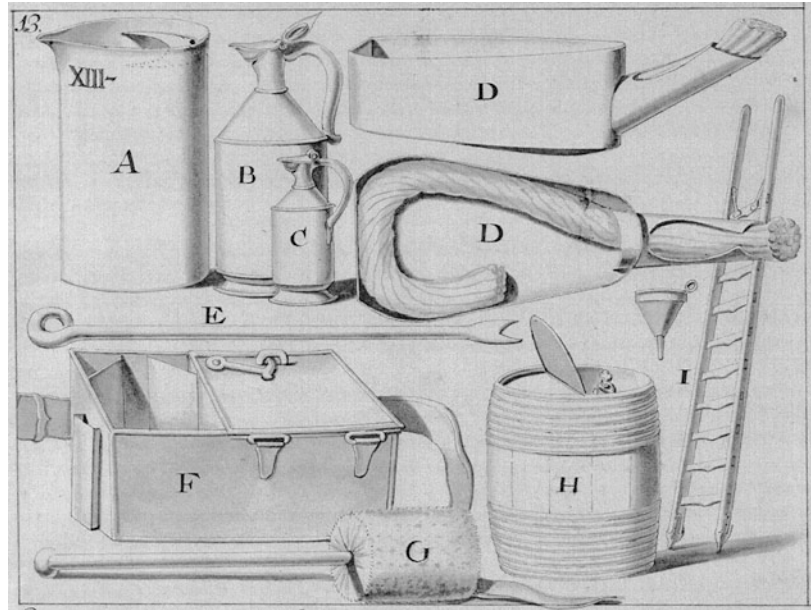
Materials and Construction

The Oil

Vegetal or animal oil, rich in carbon, is used. The type was dependent on the availability in the actual region. In warm areas vegetal oil was made from olives, coconuts, and palms and in more moderate climate regions, from colza, linseeds, and rapeseeds. Animal oil was obtained from fish, whale, or domestic animals. Especially sperm whales were hunted and slaughtered for oil obtained from their head cavities. Around 1850

Combustion Lamp,

Fig. 3 Maintenance set for a lamp lighter of oil street lighting installations around the end of the seventeenth century (Drawing van der Heijden, 1663 [3])



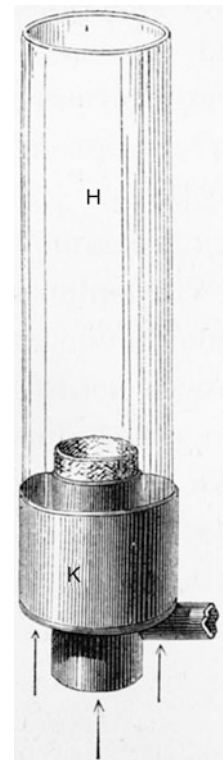
kerosene (also referred to as paraffin oil) was produced from crude oil (also referred to as petroleum) through a refining distillation process. Quickly it became the standard fuel for oil lamps which became known under the names of paraffin, kerosene, or petroleum lamps.

The Wick

In the early oil lamps days, the wick was made from bark, moss, or plant fibers that were twisted together. Later cotton was usually used for wicks. Some oil lamps had multiple wicks, up to 16 in certain Greek and Roman types. Around 1780, experiments with the shape of the wick resulted in a burner that consisted of two concentric tubes between which a tubular wick is located (Fig. 4). Through the open central tube, air is drawn so that the combustion is improved and consequently the light output as well (equally to some ten candles) while reducing smoke and smell. This oil lamp type is named, after its Swiss inventor, the Argand lamp. A further improvement of this type constituted of a glass chimney placed over the flame that increased the upward air draft through the hollow tube (see again Fig. 4).

Combustion Lamp,

Fig. 4 Argand oil burner with tubular wick inside a hollow cylinder equipped with a glass chimney [4]





Combustion Lamp, Fig. 5 Brass spout oil lamp with gutter to catch spilled oil

The Fuel Reservoir or Lantern

Originally the fuel reservoir was a simple open tray made out of stone, seashell, or earthenware in which the wick was free-floating or laid in a groove in the rim of the tray. Later the reservoir was provided with a nozzle or spout through which the wick was led. Sometimes underneath the spout a gutter was mounted to catch spilled oil to avoid pollution and for reuse (Fig. 5).

The material used for the reservoir becomes gradually more advanced: brass, copper, silver, pewter, glass, or porcelain. The reservoirs made out of these materials are closed usually with a lid, although with some versions filling of the reservoir had to be done via the spout. Around 1,800 double-reservoir oil lamps came into use on the initiative of, again, Argand who earlier introduced the double concentric tube burner. These lamps had the advantage that the supply of oil to the wick was relatively constant from a small reservoir that continuously was filled by the force of gravity by



Combustion Lamp, Fig. 6 Double-reservoir oil lamp

a secondary larger reservoir fitted somewhat higher (Fig. 6).

The Carcel lamp had a clockwork that operated a pump to raise the oil to the wick. The light output of the Carcel lamp was so constant that its horizontal intensity was for some time used as the unit of intensity. That one Carcel equals approximately 9.8 cd illustrates the high light output of the Carcel lamp relative to that of a candle. The less complicated “moderator lamp” with a spring-loaded piston to pressure-feed the burner became popular for general lighting purposes. The last improvement in oil lamps dates from around 1900 and in fact comes from a technology then already in use for gas lamps. It combines the use of a gas mantle (see a further section “gas lamps”) with a hand pump that pressurizes the fuel liquid to force it into the burner for better combustion where its flame brings the mantle to incandescence. The resulting light is brighter and has a cooler color. These types of lamps are still produced today for emergency lighting purposes and for outdoor use (Fig. 7).



Combustion Lamp, Fig. 7 Modern, double-mantle, pressurized kerosene lantern

Properties

Simple oil lamps have a lumen output of about 10 lumen and a luminous efficacy (based on heat release) of some 0.1 lm/W [5, 6]. The mixture of light from incomplete combustion and incandescence of carbon particles results in a correlated color temperature of approximately 2,000 K [7].

Argand and moderator type of oil lamps have a lumen output in the range of 50–200 lumen (comparable to 5–25 W incandescent lamp) with an efficacy of 0.1–0.3 lm/W.

Oil lamps equipped with a gas mantle may raise the lumen output to more than 500 lumen with an efficacy of 0.5–1 lm/W (more than 1.5 lm/W for the pressurized types). The correlated color temperature increases to some 2,700 K.

Candles

Candles came in use much later than oil lamps. Spillage of oil and the associated risk of fire have

always been a problem with oil lamps. With the invention of the candle made of non-fluid material, the spillage problem was solved albeit not the risk for fire. The candle was less fragile than the oil lamp and therefore more easily portable. The Romans, from the time of the birth of Christ onwards, have been responsible for the dissemination of the candle throughout Europe and the Middle East [1]. Today candles are mainly used for devotion and for ambience lighting. Annual retail sales in the USA of candles, today, exceed five billion pieces (calculated based on [5]).

Materials and Construction

The Candle Substance

The first candles were made of hard animal fat (tallow) or of beeswax. Wax candles were of much better quality but also much more expensive. From the end of the eighteenth century, the use of relatively cheap fat from the spermaceti organ in the head of sperm whales improved the quality of candles. Standardized candles on this basis were used as standards for light intensity: the candlepower (only in 1948 the SI unit candela came into use with a value roughly equal to one candlepower). Around 1830 stearin, obtained by chemical treatment of animal or vegetal fat or oil, was added which increased the melting point and consequently reduced dripping of candle fat. Around 1850 candles became much cheaper because of the use of solid paraffin, a distillation product from crude oil. At that time however gas lighting was already, at many places, the source of lighting.

Candles are produced in some different ways:

- By dipping the wick repeatedly in molten candle substance
- By pouring molten candle substance in a glass container
- By pouring molten candle substance in molds
- By drawing soft candle substance through a hole (machinically)

The Wick

The first candles had a wick made of a piece of wood, cord, or animal skin. Around 1800 the

braided cotton wick was introduced that reduced the disturbing smoke that accompanied burning candles. The braided wick can have a stiff core, originally made of lead and later of zinc, of paper, or today of synthetic fibers. Most wicks are impregnated with wax to facilitate ignition. Early wicks had to be trimmed regularly. Later wicks got such a structure that they bend and their residues dip into the molten fuel and are completely consumed so that trimming is not needed.

The Lantern

To reduce the risk of fire but especially to enable the use of candles outside, candles were used in lanterns made of perforated metal sheet or with windows of animal horn or glass. Only from the eighteenth century onwards candle lanterns sometimes were equipped with mirrors or lenses to concentrate the light in certain directions.

Monumental buildings and homes of the rich were lit with luxurious candle chandeliers, sometimes decorated with pieces of cut glass that made the light sparkle. In contrast to oil lamps, candles were only for a very limited period used for street lighting.

Properties

The light and color properties of the candle are similar to those of simple oil lamps. The lumen output of a candle flame is approximately 10 lm and its luminous efficacy 0.1 lm/W [5, 6]. The correlated color temperature is around 1,900 K [7].

Gas Lamps

Oil lamps and candles are light sources where the energy or fuel is stored in the lamp itself. Gas lamps were the first lamps where the energy is distributed to the lamps from a central energy depot at a centralized remote location. The challenge was not only to develop the lamp itself but also the development of town-sized gas production and distribution systems. In 1785 demonstrated the Dutch Jan Pieter Mickelers the use of coal gas to produce light by lighting his lecture room at the university of Leuven with gaslight.

The first to make a public demonstration of gas lighting was the Scot William Murdoch when he in 1802 installed a gas pipe network with gas burners for the lighting of the facades of a range of buildings of James Watt's Soho factory in Birmingham [1, 2, 6]. Oil and candle lighting was quickly replaced by gas lighting, first in street lighting and industrial premises, quickly followed in domestic areas of the rich. In 1875 the new Paris opera got some 45 km of gas pipe with 960 gas lights connected to it.

With the introduction of electric incandescent lamps in 1880, some 75 years after the first use of gas lighting, the use of gas lighting decreased quickly. Even today however, in some cities in the Western world, gas lighting for street lighting is still in use. The city of Berlin, for example, uses some 40,000 gas lanterns.

Materials and Construction

The Gas

Gas for lighting was sometimes produced by heating animal fats or vegetal oil but mostly by heating coal. When the latter is done in an air-free atmosphere, in cast iron retorts, the percentage of methane in the gas is higher, resulting in a better combustion and therefore in more light. After its



Combustion Lamp, Fig. 8 Bicycle carbide lantern working on acetylene gas

Combustion Lamp,

Fig. 9 From *top to bottom*: lime light gas burner with piece of limestone, box with limestones, theater spotlight projector in which the lime light gas burner is fitted (Photographs: Stage Lighting Museum, Israel, creator Dan Redler)



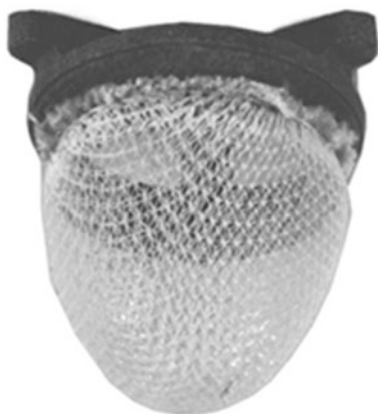
original application, this type of gas was named “illuminating gas.” Illuminating gas produces whiter light than the blue light of natural gas (only after the invention of the gas mantle – see next section – also natural gas could be used for gas lights). The gas was stored in huge gas storage tanks or gasometers, in some cities preserved as historic landmarks.

A special kind of gas was used for carbide gas lamps on bicycles and early automobiles: acetylene gas obtained by dripping water from a small reservoir in another reservoir under it that contains calcium carbide (Fig. 8).

The Burner

Open Flame Burner The first burners, so-called open flame burners, were simply a hole or a series of holes at the end of a pipe. The double concentric tube burner with glass chimney, as used since the end of the eighteenth century by Argand for his oil burners, was often used for open flame gas burners as well. It was realized that the higher the temperature of the flame, the larger the lumen output of the flame would be. In 1858 William Sugg therefore introduced burners made from non-heat-conducting steatite that became hotter than heat-conducting metal burners.

Gas Mantle Already in 1825 it was known that by putting solid material into the flame, this material could be brought to incandescence. In this way both the light output and color of the flame could be influenced. A piece of lime from limestone brought into the flame resulted in an



Combustion Lamp, Fig. 10 Gas mantle

extremely intense and compact light source that was used as spotlight in theaters (Fig. 9). The expression “in the limelight” originates from this.

In 1885 the Austrian Carl Auer von Welsbach succeeded in using a knitted mantle of solid material around the gas flame (Fig. 10). The material consists of a pear-shaped net of fabric impregnated with the nitrates of the rare-earth metals thorium and cerium. After drying, the fabric is burned off and what remains is a fragile, inflammable structure: the gas mantle existing of the oxides of thorium and cerium that is brought to incandescence in the gas burner. Because of its fragility the mantle has to be replaced regularly.

The gas mantle improved both the light output and efficacy of gas lamps considerably because thorium and cerium oxides produce more light in the visible spectrum range than a black body at the same temperature does (selective radiation). The gas mantle invention has stretched the actual use

Combustion Lamp, Fig. 11 Gas lamp chandelier with gas mantles and with thin secondary gas pipes connected to pilot lights for ease of ignition (Photograph: Pat Cryer, taken at the Museum of Welsh life, Cardiff, Wales)



Combustion Lamp,

Fig. 12 Multiple gas mantle street lighting lanterns as in use in 2012 in Dusseldorf, Germany (Photograph: Wout van Bommel)



of gas lighting well into the era of the first 50 years of electrical lighting.

Regenerative Burners Around 1880 the so-called regenerative gas light systems further improved the efficiency of gas lamp systems. The heated air produced by the flame is guided such that it preheats the incoming air needed for combustion. Often this principle was applied with inverted gas mantle burners in which the heat of the flame is better retained in the mantle.

The Lantern

The design of the early gas lanterns was very much based on the design of oil lanterns of that

period. Chandeliers for gas lighting were equipped with pull-chains for switching the gas supply on and off and to control, through the quantity of gas supply, the light output. They also often had thin secondary gas pipes connected to the so-called pilot lights that were always on so as to enable the main gas lights to be ignited without the use of matches (Fig. 11).

Lanterns for gas lights with gas mantle always had a glass cover to protect the fragile mantle and to restrict glare. The glass cover, of course, was open at the bottom to permit inlet of air needed for combustion. The top part of the lantern had a ventilation shaft working as chimney (Fig. 12). Most lanterns for use in street lighting and for the lighting of



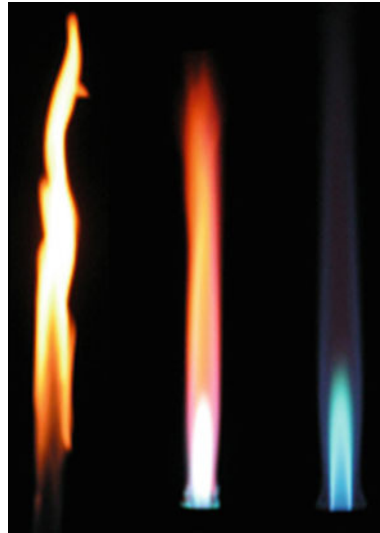
Combustion Lamp, Fig. 13 Lamplighter (Photograph: Klearchos Kapoutsis, Foter)

industrial premises had multiple gas mantle burners varying in number from 2 to more than 10.

Ignition Control

In a large part of the nineteenth century, gas street lighting lanterns were ignited each day at dusk by a patrolling lamp lighter (Fig. 13). With a stick on which a torch was fitted, he opened a trap door in the bottom of the lantern, opened the gas valve with a hook on the stick, and ignited the flame with the torch.

Later, gas lanterns had for remote ignition a pilot lamp connected through a bypass to the main gas pipe. Gas supply was remotely controlled by gas valves actuated by gas pressure or by a mechanical clockwork built into the lantern. From 1900 onwards remote gas igniters came on the market which did not need a pilot light. A platinum sponge, existing of a porous mass of finely divided platinum, initiates a flame by catalytically combining oxygen with hydrogen from the gas when the gas valve is opened. Later, also battery-operated remote igniters came into use.



Combustion Lamp, Fig. 14 Different colors of gas flame depending on type of gas and oxygen supply. From left to right: more complete combustion (Photograph adapted from: WarX: Creative Commons 3.0 unported)

Properties

A single gas burner without mantle has a lumen output of 10–50 lumen with an efficacy of 0.2–0.5 lm/W [5, 6]. The spectrum of radiation depends on the type of gas and the oxygen supply that together determine the quality of the combustion reaction and the soot particles that take part in the incandescence process. Figure 14 shows different colored gas flames resulting from different types of gas and different oxygen supply. Open flame burners on “illuminating” gas have the color of flame as shown on the left: yellowish white light from incomplete combustion and incandescence of soot particles. Modern gas ovens have flames as illustrated on the right: a blue flame from complete combustion, without soot particles taking part in the process. The correlated color temperature of the open burner gas flame, burning on illuminating gas, lays around 2,400 K [8].

Gas burners with a gas mantle have per mantle a lumen output of 200–600 lumen (comparable to that of a 25–60 W incandescent lamp) at 1–2 lm/W [5–7]. It is interesting to compare this luminous efficacy with that of Edison’s first incandescent bulb: 1.4 lm/W. The correlated color temperature is dependent on the

composition of the mantle. Typical values are 2,700–2,900 K [8]. The lifetime of a gas mantle was some 50–200 h. Modern gas mantles, as used today in professional street lighting lanterns, have a life of up to 4,000 h.

The effect of the mains pressure on the performance of gas lamps was much less than the effect of mains voltage on the performance of most electric lamps. From the last quarter of the nineteenth century, most gas lanterns are equipped with a small pressure governor that keeps the outlet pressure within acceptable limits.

Cross-References

► [Incandescence](#)

References

1. Stoer, G.W.: *History of Light and Lighting*. Philips Lighting Division, Eindhoven (1986)
2. Rebske, E.: *Lampen Laternen Leuchten*. Franckh'sche Verlagshandlung, Stuttgart (1962)
3. Van der Heijden, J.: *Het licht der lamplantarens*. Town library, Haarlem (1668)
4. Figuier, L.: *Les Merveilles De La Science, ou Description Populaire Des Inventions Modernes*. Hachette Livre, Paris (1891)
5. Dillon, S.E.: Characterization of candle flames. *J Fire Protect Eng* **15**, 265–285 (2005)
6. Luckiesh, M.: *Artificial Light Its Influence Upon Civilization*. The Century Books of Useful Science, New York (1920)
7. Waldram, J.M.: *Street Lighting*. Edward Arnold, London (1952)
8. Macbeth, N.: Color temperature classification of natural and artificial illuminants. *Trans. IES* 23:302–324 (1928)

Comparative (Cross-Cultural) Color Preference and Its Structure

Miho Saito
Department of Human Sciences, Waseda
University, Mikajima, Tokorozawa, Saitama,
Japan

Synonyms

[“Blue-Seven Phenomenon”](#); [Color preference](#); [Cross-cultural color preference](#); [Culture and](#)

[color](#); [Preference for white in Asia](#); [Structure of color preference](#)

Definition

Surveys on color preference can be found among the very first psychological experiments, with several factors thought to be responsible for color preference, such as age, gender, and geographical area of residence. Although numerous studies have investigated age and gender differences in color preference, very few have concentrated on geographical regions, especially from a cross-cultural perspective. Data from early surveys indicated the existence of cultural differences, especially in Asia where white was commonly and strongly preferred by Japanese, Koreans, Chinese, and Indonesians. Subsequent studies have shown that blue has been consistently preferred in many countries for many years. The term “Blue-Seven Phenomenon” is used to indicate that blue is the universally favorite color. The phenomenon refers to Simon’s finding that subjects selected “blue” when asked to name a color and selected “seven” when requested to choose a number from zero to nine and has been widely researched in many countries. Generally, the associative images which were assumed to be responsible for color preference and the subjects’ reasons for selecting colors that tended to be liked or disliked regardless of time or place were closely connected with the feelings of pleasantness and unpleasantness. Cognitive studies suggest that the amygdala is closely connected with preference in relation to the feelings of pleasantness and unpleasantness, suggesting that the feelings of “pleasantness” and “unpleasantness” also play an important role in determining color preference. Based on an analysis of the results of such surveys, a general structure of color preference was suggested in a three-layered diagram, with preferred feelings of “pleasantness” and “unpleasantness” forming the nucleus or the innermost first layer, preferences based on individual factors composing the surrounding second layer, and preference based on environmental factors forming the outermost third layer.

Historical Background of Studies on Color Preference

The study of color preference is of current interest, especially among cognitive psychologists and neuroscientific researchers because preference is a basic human trait which regulates everyday behavior. There are numerous studies which have attempted to clarify the mechanism of preference and to isolate the factors which influence one's preference or taste. While there are various research areas concerning this topic, this entry focuses on the mechanism underlying the preference for color.

Surveys on color preference can be found among the very first psychological experiments. Some studies have been carried out on the preference for colors associated with particular objects. Many, however, have investigated the affective appeal of color, not in combination, but separately, so as to evaluate single colors themselves without the influence of other variables.

Several factors are thought to be responsible for color preference, such as age, gender, and geographical area of residence. Although numerous studies have investigated age and gender differences in color preference, very few have concentrated on geographical regions, especially from a cross-cultural perspective.

Eysenck [1] suggested that there was a general order of preference for fully saturated hues in the order of blue, red, green, purple, and orange, with yellow ranking last. As this order did not differ between Caucasian and other races, he concluded that there was no cross-cultural difference in the preference for colors. Choungourian [2] reported the preferences of American, Lebanese, Iranian, and Kuwaiti university students in Beirut. While red and blue ranked highest in preference value for the American subjects, those colors ranked lowest for Kuwaitis. Blue-green was ranked as being the least preferred among the Americans, but was most preferred by both the Iranian and Kuwaiti subjects. He concluded that cultural variables were an underlying factor in determining color preferences. On the other hand, a factor analysis study by Adams and Osgood [3] found similarities in feelings about colors among 23 cultural groups.

Saito [4] demonstrated cross-cultural differences and similarities in color preference among nine cultural groups. The groups were Americans, Germans, Danes, Australians, Papua New Guineans, South Africans, Japanese-Americans living in the USA, non-Japanese living in Japan, and Japanese. Four hundred subjects were asked to choose the colors they liked and disliked from among 65 colored chips. Results showed that vivid blue was the only color that was commonly preferred highly by all groups, suggesting that cultural variables are indeed involved in color preference.

One significant finding emerging from Saito's study was the distinct Japanese preference for white. One out of every four of the Japanese subjects selected white as their first, second, or third choice, while no such high preference for white was observed in other countries.

In factor analysis and cluster analysis studies [5, 6], a detailed investigation of color preference was carried out on 1600 Japanese in four large cities, considering subjects' age, gender, area of residence, and lifestyle. This study also suggested that white was the highest preferred color, regardless of age, gender, or area of residence, further indicating the high preference for white by Japanese.

To investigate whether this tendency was unique to Japan, if it may be observed in other Asian areas and if the preference is influenced by environmental factors such as cultural and geographical aspects, Saito [7] replicated the study in Seoul (Korea). The fact that white was preferred highly not only in Tokyo but also, even more so, in Seoul led Saito and Lai [8] to conduct the same survey in Taipei (Taiwan), which is close to Japan both geographically and culturally, to further test the hypothesis that the strong preference for white is based to some degree on geographical and cultural variables. The result of the survey indicated that the high preference for white was common in Taipei as well.

The preference for white in China has long been noted in studies, beginning with those by Chou and Chen [9] and Shen [10]. Chou and Chen postulated two possible explanations for this preference: association influence and tradition

influence. As the most frequently used color word in Chinese literature was the character for white, they postulated that the subjects' preference was based on familiarity, i.e., frequency of association. The second possibility, tradition influence, was that their preference derived from the color of their national flag. (It is to be noted that the colors of the flag were not the same then as the present-day flags). Shen, however, questioned Chou and Chen's explanations and offered an alternative explanation which combined Chou and Chen's concept of association frequency with language influence. For example, he noted that the Chinese character for white is associated not only with pureness but also with everything open, clear, and unselfish, while grayness (gray was the least preferred color in their study) is a symbol for everything negative, disappointing, discouraging, or pessimistic.

Saito [11] extended the area of investigation to Tianjin in China and Jakarta in Indonesia in order to investigate the preference for color in more detail with special emphasis on the preference for white to establish whether or not a strong preference for white is common to Asian areas which have both geographical and cultural proximity. As a result, it was found that while white was strongly preferred in both areas, the reasons for the preference were different. In Japan, white was mostly preferred because of its associative image of being clean, pure, harmonious, refreshing, beautiful, clear, gentle, and natural. In China, the reasons for the choice were mainly in association with chastity or purity. Chinese also preferred white because it was elegant, clean, beautiful, and "pure white." It was also found that white is also a symbol of sacredness for them. Several subjects were reported as stating that white was the source of every color suggesting it to be substantial and unique. In Indonesia, white was reported as being mostly preferred for its image of being clean, chaste, neutral, and light.

The associative images stated above were assumed to be responsible for the strong preference for white. However, it should be noted that in China, white was found to be sometimes disliked, especially by male subjects because of its

lifelessness, emptiness, loneliness, and image of death. And some Indonesian subjects were reported to have also disliked white, although only very slightly, because it was too light, too easy to become dirty, and too simple. However, for Indonesian subjects, it was found that white did not have the image of death as it did for the Chinese.

Another possible explanation of the preference for white in Japan is that literature on ancient Japanese religion and mythology states that ancient people believed in the power of the Sun. This belief can still be found in Japanese Shintoism. The Sun Goddess is called Amaterasu-ōmikami. As white represented the color of the Sun or sunshine, people accepted it as a sacred color. This is shown by Shinto priests wearing holy white costumes and also holding a sacred wand called a *gohei*, a purifying implement with white strips of paper used while they pray (Fig. 1). Such items which imply that white is a sacred color can still be seen throughout the country.



Comparative (Cross-Cultural) Color Preference and Its Structure, Fig. 1 Photograph of a Gohei (white strips of paper) for a purification (With permission from Office of Public Relations, Waseda University)

There are examples, especially in folklore, of even white objects or animals becoming objects of worship at times. In this way, white had special meaning for people who revered the sun. For those people, this may explain why the color quite naturally came to be favored and admired.

Studies on the “Blue-Seven Phenomenon”

The “Blue-Seven Phenomenon” has been widely researched in many countries since it was first reported by Simon in 1971. It refers to Simon’s finding that over 40 % of American subjects selected blue when asked to name a color and over 30 % selected “seven” when requested to choose a number from zero to nine. This phenomenon has been confirmed by studies in the USA, Australia, and Kenya. In order to investigate color and number preferences in Japan, Saito [12] asked 586 university undergraduates (239 men and 347 women, average age = 20.85) (1) to name the color which first comes to mind, (2) to name his or her favorite color, and (3) to select his or her preferred number from zero to nine.

Japanese students were reported to have selected blue (33.50 %) most frequently followed by red (28.02 %) when they were asked to name a color (question 1). Blue, red, white (11.06 %), and black (6.18 %) together accounted for approximately 80 % of the responses. Further, a gender difference was reportedly found in the selection of colors, with blue and black being preferred more by men and red and white being preferred more by women. On the other hand, in response to question 2, the top four colors were the same, but red was not chosen as frequently as blue as the preferred color (red, 11.09 %; blue, 37.08 %). A gender difference was also obtained in question 2.

As for the preferred number, the subjects in Saito’s [12] study selected “seven” most frequently (22.50 %), supporting Simon’s [13] finding of the “Blue-Seven Phenomenon.” The reasons given for the choice showed that “seven” was associated with “lucky seven” and was considered “a lucky number” and to “represent happiness” among Japanese students. Other highly preferred numbers were found to be “three” (16.24 %), “five” (13.03 %), and “one” (11.84 %). Odd numbers accounted for 68.35 %

of the responses. Male students selected the number “one” more often (men, 15.67 %; women, 9.07 %), the main reason given being that it represented “number one” or “top.” Female students, on the other hand, preferred “five” (men, 9.66 %; women, 15.30 %), because they “just liked the number” or because it was “a birth date,” “a good cutoff point,” or “a shapely number.” A gender difference was also found in number selection. Numbers were sometimes preferred for their “visual appearance.”

The results of Saito’s study consequently indicated the existence of the “Blue-Seven Phenomenon” among Japanese students. It should be noted that the top four colors (blue, red, white, and black) have consistently been found to be preferred highly in Japan by the method of choosing a favorite color from a color chart, as reported in related color preference studies [14]. Moreover, it has also been found to be the only color not likely to be taboo in most cultures [15].

Further study will be needed to determine whether the predominant selection of blue is a natural, spontaneous human response or something that is based on personal preference. In other words, it is believed that when humans and other living organisms show some reaction, they do not show the same tendency to react with every object-stimulus, but instead show a type of selectivity. In that sense, “seven” and “blue” may be the number and color, respectively, where response is likely to be concentrated.

There are several factors that may be connected with this tendency. For example, the subject’s tendency to respond in a certain way may be increased by factors from his or her cultural background (e.g., the belief that “seven” is a lucky number or the positive image of blue as seen in the paintings of Europe and America) or simply by the individual’s tastes. Therefore, if, after carrying out investigations in various areas of the world, it is found that there is a multiregional tendency toward certain responses, such as the tendency toward the selection of “seven” and “blue,” then these findings should not be dismissed simply as phenomena to be noted. Instead, studies from the viewpoint of human science or interdisciplinary studies should be

carried out to investigate whether the cause of such common tendencies in human response is related to an innate cognition style or due to a cognition style acquired through experience.

Cognitive Implication and the Structure of Color Preference

During the course of the analysis of color preference, it has been found that there are preferences which have remained relatively unchanged for many years and those that have been changeable. In addition, it seems to be found that there are preferences that are common universally and those that seem to be distinctive to a specific region.

For example, “blue” tended to be preferred very highly in all regions in all years surveyed. Similarly, as report above, in an early cross-cultural 1963 study of taboo colors over the world, Winick reported that he did not find “blue” to be a taboo color in any country.

Moreover, the subjects’ reasons for selecting colors that tended to be liked or disliked regardless of time or place were closely connected with feelings of “pleasantness” and “unpleasantness.” According to the results of those surveys, the three principal images most frequently associated with pleasant feelings were “beautiful,” “agreeable,” and “bright,” while those most frequently associated with unpleasant feelings were “dirty,” “disagreeable,” and “dark.” These associations were observed commonly in all regions regardless of the year of the survey.

In the field of cognitive psychology, it is said that when information reaches certain centers of the brain, the corresponding sensory system is stimulated. The various stimuli that reach the sensory centers in the cerebral cortex do not simply remain there, but are transmitted to the amygdala and activate the memory circuit connecting the amygdala and the thalamus. The involvement of emotion is thought to occur because the union of the amygdala and the hypothalamus adds emotion to past experience and memory. This may be why, for example, when we hear a certain sound, we may remember nostalgic scenes or sense colors or smells that we associate with that sound or that we usually find accompanying it.

Thus, the union of the amygdala and the hypothalamus may also be involved in outcomes such upon seeing a certain color, we may feel an emotion, form an image in our minds, or make a specific association. Our perception of color, in other words, is not a simple sensation. Rather, we add on psychological elements such as emotion as part of our act of “seeing.” This is what the cognition of color involves, and the integrated performance of this cognition is suggested as being the result of the integrated union between the amygdala and the hypothalamus.

Of further significance is the close connection of the amygdala with our feelings of preference. In the physiology of the brain, the amygdala has always been thought to be involved in our judgment of whether something is “safe” or “dangerous.” This was then transferred to the feelings of “pleasantness” and “unpleasantness,” with “safe” being equated with “pleasantness” and “dangerous” with “unpleasantness.” Ultimately, this became transformed into the feelings of like and dislike.

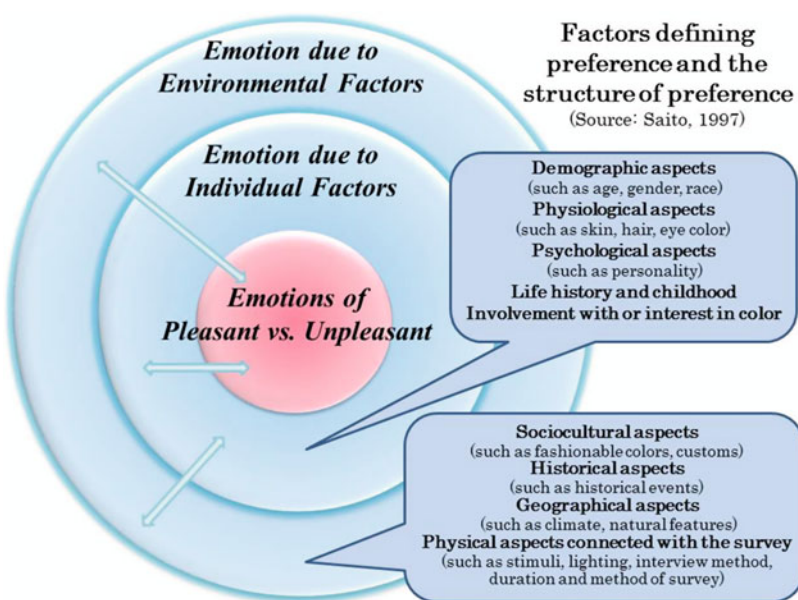
As mentioned above, the amygdala is closely connected with preference in relation to the feelings of “pleasantness” and “unpleasantness,” and it seems apparent that the feelings of “pleasantness” and “unpleasantness” are closely related to the basis of color preference as well. Based on an analysis of the results of those surveys, a diagram has been made (Source: Saito et al. [16]) of what it is believed to be one of the general structures of color preference (Fig. 2).

As shown in the diagram, color preference has a three-layered structure, with preference due to feelings of “pleasantness” and “unpleasantness” forming the nucleus or the innermost first layer, preference due to individual factors composing the surrounding second layer, and preference due to environmental factors the outermost third layer.

The individual factors involved in the second layer include demographic aspects such as age and gender, physiological aspects such as skin color and eye color, psychological aspects such as personality, and other elements such as the individual’s life history, including his or her childhood. These are believed to be the principal individual factors that affect color preference.

Comparative (Cross-Cultural) Color Preference and Its Structure,

Fig. 2 Diagram – factors defining preference and the structure of preference (Source: Saito, M. (1997). “Shikisaino Kouzouni Kansuru Shinrigakuteki Kenkyu—Kokusai Hikaku Kenkyuwo Toushite— (A Psychological Study of a Structure of Color Preference—Though Cross-cultural Studies—),” an unpublished doctoral dissertation, the Graduate School of Human Sciences, Waseda University., Fig. 4.1, p. 264)



The environmental factors in the third layer affecting color preference are thought to include sociocultural aspects such as current fashion trends or custom, geographical aspects such as climate and natural features, historical events, and physical aspects related to the survey itself.

The closer the preference is to the center of this structure, the more stable it is, and the more it is a preference that is common to all people, being relatively unaffected by differences in geographical area of residence or year of survey. The further away the preference is from the center, the more liable it is to change with the individual and the environment surrounding that individual.

As has been shown, the factors influencing comparative color preference are varied and diverse. Further studies are necessary to clarify other factors which may influence this phenomenon, because color preference is such a fundamental human trait.

References

1. Eysenck, H.J.: A critical and experimental study of color-preferences. *Am. J. Psychol.* **54**, 385–394 (1941)
2. Choungourian, A.: Color preference and cultural variation. *Percept. Mot. Skills* **26**, 1203–1206 (1968)
3. Adams, F.M., Osgood, C.E.: A cross-cultural study of the affective meaning of color. *J. Cross-Cult. Psychol.* **7**, 135–157 (1973)
4. Saito, M.: Shikisai Shikouno cross-cultural research (A cross-cultural study on color preference). *Bull. Grad. Div. Literat. Waseda Univ.* **27**, 211–216 (1981)
5. Saito, M., Tomita, M., Kogo, C.: Nihonno Yontoshiniokeru Shikisai Shikou (1) Inshibunsekiteki Kenkyu (Color preference at four different districts in Japan (1) – factor analytical study). *J. Color Sci. Assoc. Jpn.* **15**(1), 1–12 (1991)
6. Saito, M., Tomita, M., Yamashita, K.: Nihonno Yontoshiniokeru Shikisai Shikou (2) Kurasutabunsekiteki Kenkyu (Color preference at four different districts in Japan (2) –classification of characteristics of life style by cluster analysis). *J. Color Sci. Assoc. Jpn.* **15**(2), 99–108 (1991)
7. Saito, M.: Ajianiokeru Shikisai Shikouno Kokusaihi-kakukenyu (1) Nikkanhikaku Shiroshikouni Chakumokushite (A cross-cultural survey on color preference in Asian countries (1) – comparison between Japanese and Koreans with emphasis on preference for white). *J. Color Sci. Assoc. Jpn.* **16**(1), 1–10 (1992)
8. Saito, M., Lai, A.C.: Ajianiokeru Shikisai Shikouno Kokusaihi-kakukenyu (2) Nittaihi-kaku Shiroshikouni Chakumokushite (A cross-cultural survey on color preference in Asian countries (2) – comparison between Japanese and Taiwanese with emphasis on preference for white). *J. Color Sci. Assoc. Jpn.* **16**(2), 84–96 (1992)
9. Chou, S.K., Chen, H.P.: General versus specific color preference of Chinese students. *J. Soc. Psychol.* **6**, 290–314 (1935)

10. Shen, N.C.: The color preference of 1368 Chinese students, with special reference to the most preferred color. *J. Soc. Psychol.* **8**, 185–204 (1937)
11. Saito, M.: Comparative studies on color preference in Japan and other Asian Regions, with special emphasis on the preference for white. *Color Res. Appl.* **21**(1), 35–49 (1996)
12. Saito, M.: “Blue and seven phenomena” among Japanese students. *Percept. Mot. Skills* **89**, 532–536 (1999)
13. Simon, W.E.: Number and color responses of some college students: preliminary evidence for a “Blue Seven Phenomenon”. *Percept. Mot. Skills* **33**, 373–374 (1971)
14. Japan Color Research Institute (ed.): 12th annual report on consumers’ color preference. Japan Color Research Institute, Tokyo (1992)
15. Winick, C.: Taboo and disapproved colors and symbols in various foreign countries. *J. Soc. Psychol.* **59**, 561–568 (1963)
16. Saito, M.: Shikisaino Kouzouni Kansuru Shinrigakuteki Kenkyu—Kokusai Hikaku Kenkyuwo Toushite (A psychological study of a structure of color preference—though cross-cultural studies). A doctoral dissertation, The Graduate School of Human Sciences, Waseda University, 297 p (1997)

Comparative Color Categories

Romann M. Weber¹ and Mark Changizi²

¹Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA

²Institute for the Study of Human and Machine Cognition, Boise, ID, USA

Synonyms

Cone fundamentals; Dichromacy; Primate color vision; Trichromacy

Definition

Comparative color categories are salient partitions in the perceived color spaces of primates that are known to vary when encoded by signaling systems that differ across species.

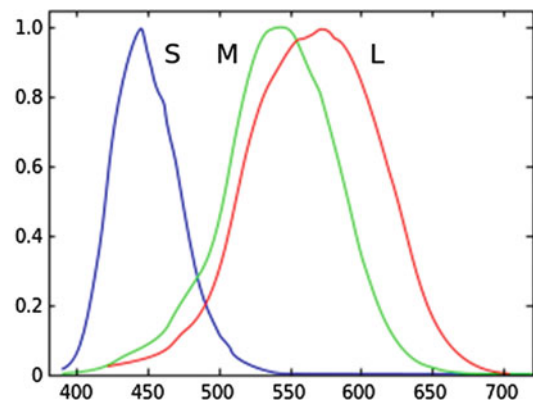
The Neurophysiology of Color Vision

The entire range of human color perception, in all its vibrancy, is due to the operation of just three

types of photosensitive retinal cone cells and the neural mechanisms responsible for interpreting their signals. These cone cells contain photosensitive opsin pigments that preferentially absorb light at different wavelengths. The short-wavelength (S) pigment has peak absorption of light at about 430 nm, roughly corresponding to blue light; the medium-wavelength (M) pigment has peak absorption at about 530 nm, corresponding to green light; and the long-wavelength (L) pigment has peak absorption at about 560 nm, corresponding to red light (see Fig. 1).

Dysfunction in the cone cells or in the coding of their pigments leads to various types of color blindness. In humans, genes for S-type cones are located on chromosome 7, while those for the M- and L-type cones are located on the X chromosome. This largely explains why color blindness is a much more common phenomenon in men, as women possess independent copies of the M- and L-pigment genes on their two X chromosomes.

The M- and L-type pigments are nearly identical in their genetic specifications. The small difference in their light-absorption profiles is due to a difference in only 3 out of 364 amino acids that code for their respective proteins [8]. The S-type pigment, on the other hand, is more distinct on the molecular level.



Comparative Color Categories, Fig. 1 Sensitivity curves for human photopigments (Attribution: Vanessaekowitz at [en.wikipedia](https://en.wikipedia.org))

Each type of cone cell responds over a range of wavelengths, and the intensity of a cone's response to a light signal depends not only on that signal's wavelength but also its intensity. This creates a situation in which the wavelength of a light signal cannot be uniquely identified based on the response of a single type of cone cell. Unique color information is only recoverable through the simultaneous and concerted action of different types of cone cells operating within a small neighborhood of each other.

In a sense, the response profile of each type of cone cell can be viewed as a "basis vector" that, along with the other cones that make up the basis, defines a multidimensional color space. This analogy is made more concrete when considering the *color-matching* experiment that is foundational to the field of color science. In this experiment, observers are presented with a test light with an arbitrary spectral power distribution on one side of a bipartite white screen. The observers' task is then to try to match the appearance of the test light by individually adjusting the intensities of a series of primary lights focused on the other side of the screen. The lights on each side of the screen, while physically different in their spectral distributions, are perceived to be the same. Appearance-matched lights produced in this way are called *metamers*.

For people with normal color vision, exactly three primary lights are required to perform the color-matching task. As a result, normal human color vision is said to be three-dimensional or *trichromatic*. Establishing the dimensionality of color vision in nonhuman animals is an often difficult task requiring extensive training and testing of the animals [7]. A frequent practice in the literature – arguably justified – is to equate an animal's color-vision dimensionality with the number of distinct color photopigments it expresses in its retina. However, the tarsier provides an example of a primate with a fairly uneven retinal distribution of its two cone types, challenging classical views about how color identification works through local networks of opponent cone cells [5]. The tarsier's labeling as a dichromat based on a photopigment-type count alone should therefore be done cautiously [7].

The Evolution of Primate Color Vision

It is believed that gene duplication and mutation of cone pigments, which ultimately made multidimensional perceptual color spaces possible, traces back to the time of the first vertebrates [1]. Almost all vertebrates have an S-type pigment very similar to that possessed by humans [8]. Most mammals have a single X-chromosome cone opsin gene, typically coding for a single long-wavelength visual pigment, and are *dichromats*. But even among primates, only the Old World monkeys, apes, and humans are *trichromats*, having a second – and highly homologous – cone opsin gene on the X chromosome, which codes for a medium-wavelength visual pigment [8, 12]. The generation of these red- and green-pigment genes is thought to have occurred following a duplication event after the split between the New World monkeys and the *catarrhines*, a group of higher primates made up of the Old World monkeys and apes (including humans), about 40 million years ago [12].

Opsin-gene and cone-pigment complements appear to be largely the same among the catarrhines [8, 12]. Psychophysical studies of macaques provided much early information on the color-discrimination abilities of the Old World monkeys [7], and much of this research showed performance very similar to that of humans, suggesting a concordant perceptual color space [15]. Studies of chimpanzees also showed wavelength discrimination similar to human performance [4]. For a review, see [6, 7].

Among the placental mammals, trichromacy appears to be found only in certain primates [7]. However, it is possible to enhance color-discrimination behavior in dichromatic mammals through genetic engineering. Behavioral tests have shown strong evidence for novel color vision in knock-in mice expressing the human L-opsin gene [9]. This suggests that the visual systems of traditionally dichromatic mammals can readily make use of higher-dimensional color signals.

Why, then, when traditionally dichromatic animals can be easily made to demonstrate three-dimensional color vision, is trichromacy seemingly restricted to certain primates? Further, why does trichromacy in primates manifest the way

that it does, namely, with such a large spectral overlap between the M- and L-pigment sensitivities? Several hypotheses have been advanced to answer these questions. They are of two basic types: one focusing on the potential advantage of three-dimensional color vision in finding food and the other suggesting that trichromatic color vision was selected for perceiving subtle, behaviorally relevant cues on exposed skin.

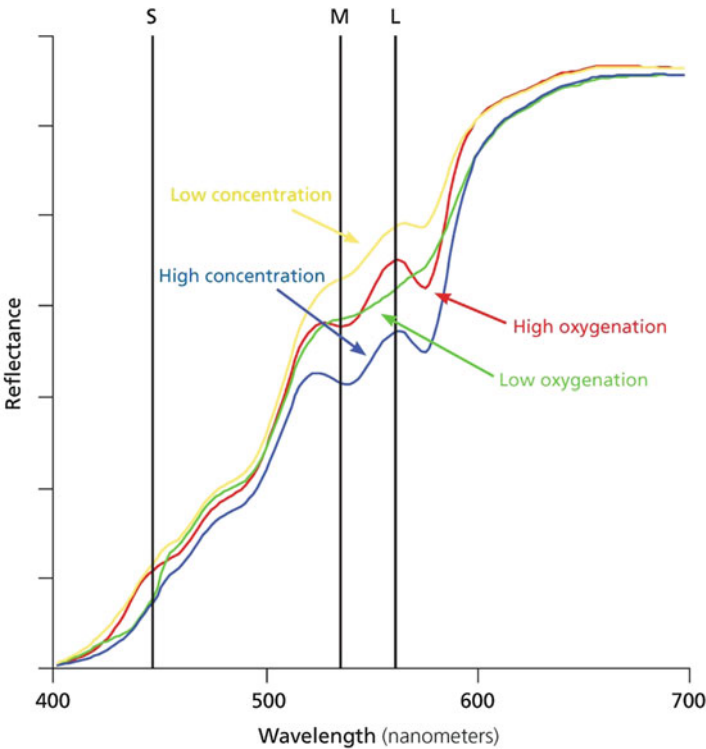
The experience of humans with various forms of reduced color vision suggested the hypothesis that color vision was selected for its benefit to detecting targets in visual scenes. Specifically, it has been conjectured that color vision is especially useful for detecting ripe fruit and edible leaves [11]. Consistent with this hypothesis is the observation that ripe fruit is more easily detected against a background of leaves in a trichromatic perceptual space [13]. Also, in one study, several species of frugivorous (fruit-eating) primates tended to eat and disperse seeds from fruit whose reflectance spectra are tightly clustered in the color space described by trichromatic catarrhine photopigments, while non-preferred

fruit is more diffusely distributed in this color space [14]. In another study, folivorous (leaf-eating) trichromatic primates that were surveyed tended to eat leaves shifted more toward the red end of the spectrum relative to leaves eaten by non-trichromatic primates [10].

The food-based hypothesis suggests an obvious evolutionary advantage for primates. However, a question remains as to why catarrhine trichromacy is so uniform in its cone sensitivities given the wide diets – not all of them strictly frugivorous or folivorous – of these primates. An alternative hypothesis suggests that the specific tuning of primate color vision is selected for its advantage in detecting changes in the concentration and oxygenation of blood in conspecifics as indicated by subtle color shifts displayed on exposed skin [3].

Skin reflectance can be modulated over two dimensions according to hemoglobin oxygen saturation and the concentration of hemoglobin in the skin [2, 3]. The effect of changes over the oxygenation-concentration dimensions result in shifts in skin reflectance that are predictable (see Fig. 2).

Comparative Color Categories, Fig. 2 Skin reflectance changes due to varying blood oxygenation and concentration



Particularly notable in this pattern of reflectance changes is that they occur exclusively in the region of color space at which the M and L pigments are maximally sensitive. That is, the characteristic “W” shape nestled within skin’s reflectance profile shifts its position in the region of peak sensitivity of the most recently evolved photopigments, which are nearly optimal for detecting this shift [3]. These shifts are also consistent across all skin types.

It has been observed that trichromatic primates tend to have either bare faces or large areas of exposed skin, such as a bare rump, a finding supportive of the skin-based hypothesis [3]. Further, dichromats are demonstrably poor at detecting changes in skin coloration [2]. Also, whereas the reflectance spectra of the collective primate diet vary widely, the shifts in skin reflectance across the oxygenation-concentration dimensions are stable across all primates [3].

While the typical human experience is that skin tone is itself barely noticeable as a color, subtle deviations from baseline coloration are obvious to people with normal trichromatic vision [3]. Consider, for instance, the “blueness” of an arm vein below the surface of the skin within the context of the surrounding skin. Outside of that context – namely, without reference to the surrounding skin, as though cropped from an image – the “blue” patch simply appears to be of normal skin tone.

More general shifts of skin color are often linked with various conditions of health and emotional status [3]. The skin-based hypothesis for trichromatic-vision selection is unique in that it suggests a *social* advantage for catarrhine color vision. This advantage applies not only to maintaining the integrity of the social group through a sensitivity to health-related cues but also to perceiving threats from conspecifics, such as signs of aggression in the form of increased oxygenation of blood.

Cross-References

- ▶ [Color Categorical Perception](#)
- ▶ [Cone Fundamentals](#)
- ▶ [Environmental Influences on Color Vision](#)
- ▶ [Metamerism](#)
- ▶ [Trichromacy](#)

References

1. Allman, J.M.: *Evolving Brains*. Scientific American Library, New York (2000)
2. Changizi, M.A.: *The Vision Revolution*. BenBella Books, Dallas (2009)
3. Changizi, M.A., Zhang, Q., Shimojo, S.: Bare skin, blood and the evolution of colour vision. *Biol. Lett.* **2**(2), 217–221 (2006)
4. Grether, W.F.: Chimpanzee color vision. I. Hue discrimination at 3 spectral points. *J. Comp. Psychol.* **29**, 167–177 (1940)
5. Hendrickson, A., Djajadi, H.R., Nakamura, L., Possin, D.E., Sajuthi, D.: Nocturnal tarsier retina has both short and long/medium-wavelength cones in an unusual topography. *J. Comp. Neurol.* **424**, 718–730 (2000)
6. Jacobs, G.H.: *Comparative Color Vision*. Academic, New York (1981)
7. Jacobs, G.H.: Primate color vision: a comparative perspective. *Vis. Neurosci.* **25**, 619–633 (2008)
8. Jacobs, G.H., Nathans, J.: The evolution of primate color vision. *Sci. Am.* **300**, 56–63 (2009)
9. Jacobs, G.H., Williams, G.A., Cahill, H., Nathans, J.: Emergence of novel color vision in mice engineered to express a human cone photopigment. *Science* **315**, 1723–1725 (2007)
10. Lucas, P.W., et al.: Evolution and function of routine trichromatic vision in primates. *Evolution* **57**, 2636–2643 (2003)
11. Mollon, J.D.: Tho she kneel’d in that place where they grew. . . . *J. Exp. Biol.* **146**, 21–38 (1989)
12. Nathans, J., Thomas, D., Hogness, D.S.: Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* **232**, 193–202 (1986)
13. Osorio, D., Vorobyev, M.: Colour vision as an adaptation to frugivory in primates. *Proc. R. Soc. B.* **263**, 593–599 (1996)
14. Regan, B.C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., Mollon, J.D.: Fruits, foliage and the evolution of primate colour vision. *Philos. Trans. R. Soc. B.* **356**, 229–283 (2001)
15. Sandell, J.H., Gross, C.G., Bornstein, M.H.: Color categories in macaques. *J. Comp. Physiol. Psychol.* **93**(3), 626–635 (1979)

Complementary Colors

Paul Green-Armytage
School of Design and Art, Curtin University,
Perth, Australia

Synonyms

[Contrary colors](#); [Contrasting colors](#)

Definition

Complementary colors are colors that “complete” each other. This completion can be understood in terms of some physical relationship or in terms of how the colors are related in their appearance. There are different ways of establishing these relationships, two being widely accepted: By one definition, two paints, inks, or colored lights are complementary if their mixture can yield a neutral black, gray, or white. This is a physical relationship that can be demonstrated. By another definition, two colors are complementary if the afterimage of one color has the same hue as the other color. This is a phenomenal relationship that can also be demonstrated. It is also common simply to claim that colors opposite to each other on a color circle are complementary, without further explanation or justification. Complementary relationships can be helpful when mixing paints to produce particular results. Complementary relationships also feature in theories of color harmony.

Introduction

The notion of complementary colors is caught between science and art, between what can be measured and what cannot. But this notion can also be seen as a bridge. If there are objective ways of establishing complementary color pairs, and if such pairs are found to be pleasing, then complementary colors may be a key to color harmony – harmony could be subject to measurement. One difficulty here is that different ways of establishing complementary relationships do not yield exactly the same results. A number of these ways are described and illustrated in the sections that follow. There is also the problem of what comes first, the objective methods or the experience of harmony. The idea that some color pairs are more pleasing than others is older than any demonstration of particular physical or phenomenal relationships of the kind associated with complementary colors. Perhaps the earlier judgments are endorsed by the later demonstrations, or perhaps the demonstrations show color

relationships that are now widely agreed to be harmonious.

Color Harmony

Harmony is a slippery word. Most definitions deal with harmony in music, but there it has more than one meaning. It can mean a combination of notes which have a pleasing effect [1]. It can also mean a combination of notes organized according to a system of structural principles [2]. The same definitions could be applied to color combinations. According to the first definition, only those color combinations that are found to be pleasing are harmonious. This could make color harmony a private matter; harmony, like beauty, would be in the ears of the listener or the eyes of the beholder. Only some kind of consensus could establish a wider claim for some color combinations to be accepted as more harmonious than others. Whether the judgment is made by an individual or a group, this definition depends on evaluation. With the second definition, there are also limitations. Only with certain relationships between notes or colors can a combination be called harmonious. The relationships can be measured so the definition depends on description. Philosophers might argue about the possibility or impossibility of any link between description and evaluation, but for most people, to say that there is an interval of a third between the notes C and E when they are sounded together in a chord is description, while to say that the chord is pleasing is evaluation. Similarly, to give Munsell or NCS notations for the two colors in a particular combination is description while to say that the combination is beautiful is evaluation. The idea of a link between description and evaluation as the basis for some theories of color harmony is discussed in a paper that was presented at a conference in Gothenburg in 1996 [3].

Pleasing Color Combinations

Before the invention of the color circle, or any notion of a special relationship between colors that are opposite to each other on a circle, there was a recognition that certain color pairings are

more satisfying than others. Leon Battista Alberti, in fifteenth century Italy, claimed that “there is a kind of sympathy among colours whereby their grace and beauty is increased when they are placed side by side. If red stands between blue and green, it somehow enhances their beauty as well as its own” [4, p. 85]. Leonardo da Vinci introduced the notion of “contrary colors” and suggested that pairs of contrary colors enhance one another. “The colours which go well together are green with red or purple or mauve, and yellow with blue” [5, p. 73]. Martin Kemp points out that “this account of colours which ‘go well together’ comes close to the doctrine of ‘complementary colours’ as defined in the eighteenth and nineteenth centuries, but lacks the systematic base provided by the later colour wheels” [6, p. 284].

To find out whether people might intuitively choose particular colors as going well together, without appeal to a color wheel or any theories of complementary colors, a preliminary investigation was carried out at Curtin University of Technology in 1994. Students worked with colors of equal nuance from the Natural Color System (NCS). Each was assigned a particular color and asked to find a color to go with it so that the combination would be the most beautiful, most exciting, or most harmonious. The intention was to go back to the kind of judgments made by Alberti and Leonardo, to find colors that were most “sympathetic” or “contrary.” Students generally agreed on which color pairings were the most “beautiful” and the most “exciting”, but there were two schools of thought about what kind of color relationship is “harmonious” – colors of similar or contrasting hue. And it was not always the same colors that were selected as contrasting. There was a small range comparable to the “red or purple, or mauve” as nominated by Leonardo for “going well” with green. A more rigorous study, with a larger number of participants, might lead to firmer conclusions. But it does not seem likely that an approach, from this direction, to the identification of complementary colors would yield precise results. The “complementary” of a given hue would not be a single hue but a narrow range of similar hues. This can be set beside the imprecise results when the

approach is from the other direction, starting with the theories.

Ways of Establishing Complementary Colors

With different ways of establishing complementary relationships yielding different pairs, no color can be said to have a single complementary hue unless a decision is first made about which way of establishing complementary relationships is “correct.”

Colored Shadows

In one of his diagrams, Leonardo da Vinci illustrates a spherical body (a) being lit by two lights (d) and (e). Two shadows are cast (b) and (c). He explains that “the shadow formed by the light *e*, which is yellow, will tend towards blue, because the shadow of the body *a* is formed on the floor at *b*, where it is exposed to the blue light, and accordingly the shadow made by the light *d*, which is blue, will be yellow at the location *c*, because it is exposed to the yellow light” [5, p. 74].

Later observers noticed that a blue shadow would be cast by a yellow light even if another light, shining on the shadowed area, was not blue but white. A systematic study of colored shadows was carried out by Count Rumford who reported his results to the Royal Society of London in 1794. Rumford concluded that the phenomenon was subjective. He went further and suggested that his results might serve as a guide to artists in “the magic of colouring,” and he introduced what was then a new term: “one shadow may with propriety be said to be the *complement* of the other” [7, p. 295]. Moses Harris, in 1766, also describes the colored shadow phenomenon and points out how the colors of the light and the shadow are opposite to each other on his color circle [8, p 8]. Johann von Goethe explains how to conduct the colored shadow experiment with a candle and the full moon as light sources. And he has a beautiful account of how he experienced a sequence of colored shadows on a walk in the Harz Mountains. The ground and trees were covered in snow and hoar frost. Toward the end of the

day, the shadows cast by the trees were “decidedly blue, as the illumined parts exhibited a yellow deepening to orange.” Then, as the sun began to set, it “began to diffuse a beautiful red colour over the whole scene around me, the shadow colour changed to a green, in lightness to be compared to a sea-green, in beauty to the green of the emerald. The appearance became more and more vivid: one might have imagined oneself in a fairy world, for every object had clothed itself in the two vivid and so beautifully harmonising colours” [9, pp. 34–35]. Colored shadows may be a subjective phenomenon, but it is possible to record them in photographs, as shown in Fig. 1.

Afterimages

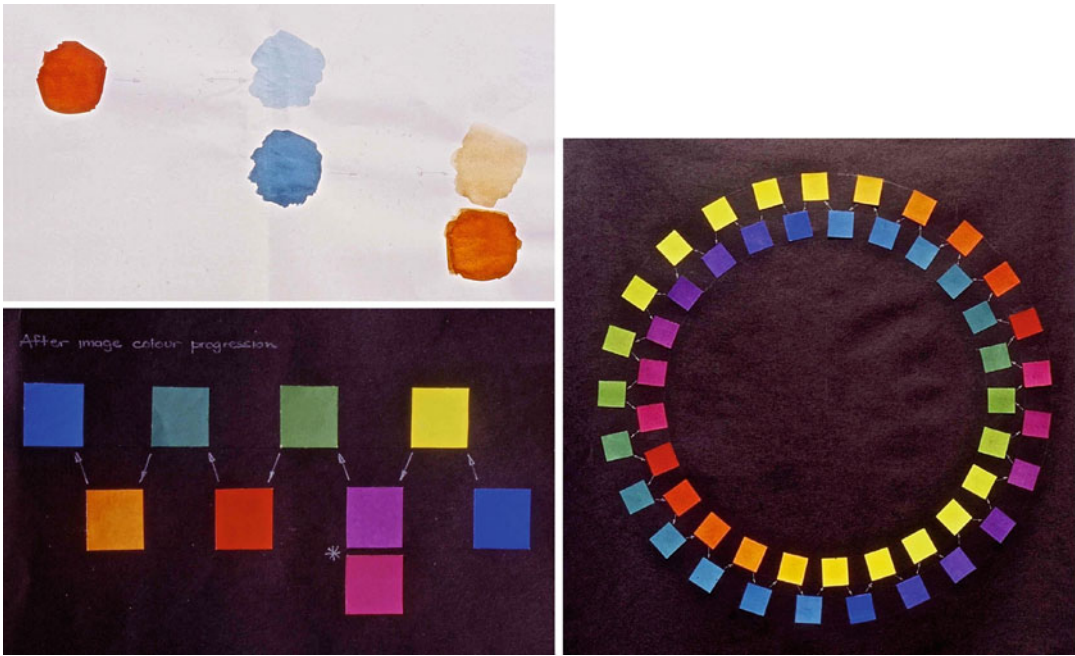
Colored shadows may be linked to the experience of afterimages. Martin Kemp identifies the Comte

de Buffon as “the pioneer in the study of subjective colour.” Buffon described colored shadows and also afterimages at the French Academy of Sciences in 1742–1743 [7, p. 294–295]. Afterimages were introduced by Moses Harris when he explains how the world will look if green-tinted spectacles are worn for about 5 min and then removed: “every scene and object will look of a fiery red, opposite to green you will find red” [8, p. 8]. In the Harris color circle, colored shadow and afterimage color pairs are to be found opposite to each other. For Goethe, afterimages were the key to color harmony. He saw, in the experience of an afterimage, the eye’s search for completeness, and he concludes that “in this resides the fundamental law of all harmony of colours” [9, p. 317]. Goethe had only six colors in his color circle so, within each segment, there could be a



Complementary Colors, Fig. 1 Demonstration of complementary colors as seen in colored shadows. At *top left* the candlestick is lit by two white lights and casts gray shadows. In the other photographs, colored filters were

placed in front of the light shining from the right while neutral filters were placed in front of the other light to equalize the illumination. The shadow colors can be described as being complementary



Complementary Colors, Fig. 2 Afterimage color progression as recorded from her observations by Sally Douglas

range of different hues under the same name. Red would range from orange red to purple red and in that range the afterimage for a particular green might be found. If more precise identification of afterimage hues is required, there are two questions to be considered: If a particular red is the afterimage of a particular green, does it follow that the same green will be the afterimage of that red? Is the afterimage phenomenon reciprocal in this way? The second question takes account of afterimages as subjective experiences peculiar to the individual. So is it possible to know whether everyone experiences the same hue as the afterimage of a given color?

Afterimage Color Progression

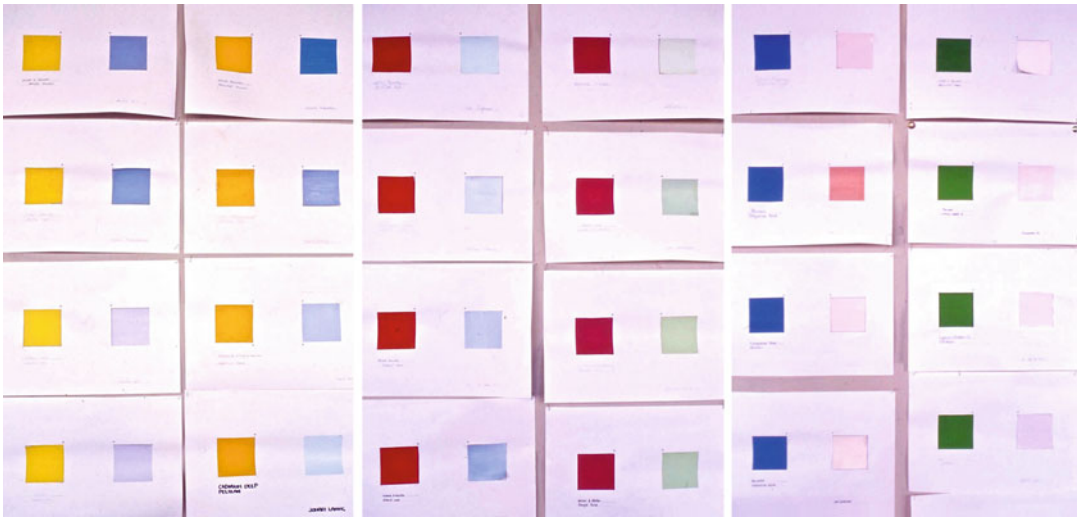
The assumption that two colors could be each other's afterimage was investigated by Nathan Cabot Hale. Cabot Hale claimed to have found what he called "after-image color progression" [10, p. 262]. Having experienced a yellow as the afterimage of a blue, he found that the afterimage of the yellow was a violet. Then the afterimage of the violet was a green and this progression continued right round the circle. Cabot Hale's claim

was investigated, in turn, by Western Australian artist Sally Douglas. Douglas did not find such strong hue shifts as those reported by Cabot Hale, but she did find progression all round the circle. Her method and results are shown in Fig. 2.

Having painted a patch of reddish orange, Douglas stared at it, looked away, and matched its afterimage which she saw as a pale turquoise. She then increased the intensity of the turquoise, being careful to retain its hue, and repeated the process for the turquoise. The afterimage of the turquoise was slightly more yellowish than the original reddish orange. In a carefully controlled operation, where such color relationships were established in random order, she found, when she put it all together, a progression right round the circle.

Individual Variations in the Experience of Afterimages

To see if individuals have the same afterimage experiences from given colors, students at Curtin University of Technology were assigned specific paints and asked to match the colors they experienced as the afterimages. Some of the results are shown in Fig. 3.



Complementary Colors, Fig. 3 Afterimage colors as recorded by individual students at Curtin University of Technology

The slight variations in hue may be due to inaccurate mixing but studies by Marian-Ortolf Bagley have confirmed that there are, indeed, variations in how individuals experience afterimages [11]. Rather than have the participants in her study match the afterimages in paint, she asked them to identify their afterimage colors from the samples in the Munsell book of color.

Colors That Are Least Like Each Other

Wilhelm Ostwald points out how “if one moves away from a given hue in the hue circle, the colors become increasingly dissimilar” [12, p. 33]. Since the hue circle is continuous, there will come a point where this dissimilarity is at its greatest, and beyond which there will be a progressive return to similarity. For Ostwald, “there exists for every hue in the hue circle another that is most different from it. This relationship is mutual. The entire hue circle is filled with such pairs of contrasting colors, which shall be called complementary colors” [12, p. 34]. However, the determination of least similarity must depend on the judgment of observers, and such judgments are likely to vary. Ostwald preferred precision. Accordingly he appeals to mixture and proposes an alternative definition: “Complementary colors are colors which in an optical mixture yield a neutral gray” [12, p. 35].

Colors that Mix to a Neutral

When discussing color mixture, it is important to distinguish between the paints, inks, or lights that are being mixed on the one hand and, on the other hand, the appearance of those paints, inks, or lights and the resultant mixture. If a paint that appears red is mixed with a paint that appears blue, the mixture will appear purple, more bluish, or more reddish depending on how much of each paint is in the mixture. There are different ways of mixing: subtractive, additive, and partitive. The results of one way of mixing are not always a guide to the likely results of the other ways.

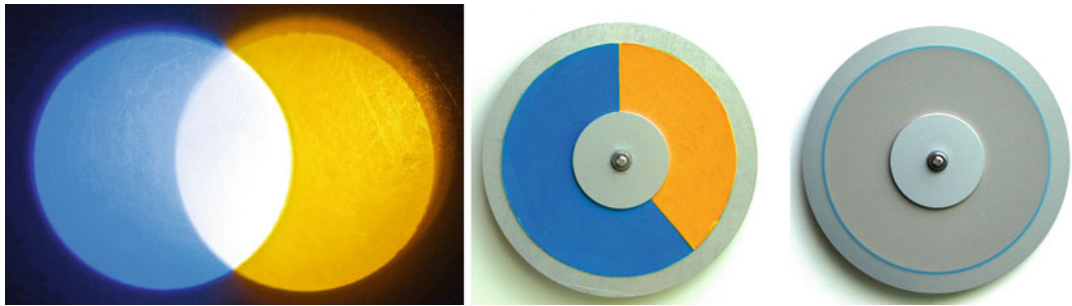
Subtractive Mixture

Subtractive mixture occurs when paints are mixed in the palette, when inks are premixed before being used on the press, and when transparent inks or paints are applied in layers one over the other. Subtractive mixture also occurs when two differently colored filters are placed in front of a single light source. Experiment with different paint combinations can lead to the discovery of a pair of paints that can be mixed to a dark gray as shown in Fig. 4.

Moses Harris claims that if such colors (i.e., paints) are mixed and are “possest of all their powers, they then compose a deep black” [8,



Complementary Colors, Fig. 4 Dark gray in the middle of a sequence of colors mixed from varying amounts of two paints which appear blue and orange when used straight from the tube



Complementary Colors, Fig. 5 Similar colors in filtered lights and on painted paper discs can mix additively and partitively to a neutral – white or gray

p. 7]. But he concedes that no pigments that are generally available, like those used for Fig. 4, have such “powers.” Instead of deep black, the result here is a neutral dark gray.

Additive and Partitive Mixture

Additive mixture occurs when beams projected from two filtered light sources overlap on a screen. Partitive mixture occurs when a disc with segments painted in different colors is spun at high speed. The results of additive and partitive mixture are shown in Fig. 5.

Partitive mixture was the method used by Ostwald to establish complementary pairs which he then placed opposite to each other on his circle. Ostwald’s circle has yellow opposite blue, as in Fig. 5, where Harris has orange as in Fig. 4.

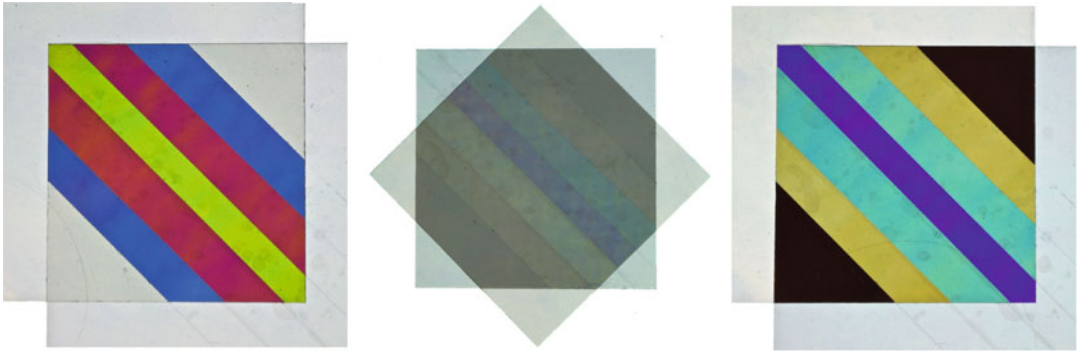
Possibilities from Physics

In the introduction to his book *The Principles of Harmony and Contrast of Colors and their Application to the Arts*, Michel-Eugène Chevreul suggests a precise definition: “if we re-united the total quantity of the coloured light *absorbed* by a coloured body, to the total quantity of coloured

light *reflected* by it, we should reproduce white light: for it is this relation that two differently coloured lights, taken in given proportions, have of reproducing white light, that we express by the terms *Coloured lights complementary to each other, or complementary colours*” [13, p. 54]. From this one can imagine a spectral reflectance curve which would serve as a kind of template for a second curve – where one curve had peaks, the other would have valleys. This may be a theoretical ideal; it is doubtful whether two surfaces could be found with such perfectly matched reflectance curves.

Newton’s Rings

Color relationships that could be regarded as complementary can be seen in a close examination of Newton’s rings. Isaac Newton describes the phenomenon that bears his name in his book *Opticks*. He found that “By looking through . . . contiguous object glasses . . . that the interjacent air exhibited rings of colours, as well by transmitting light as by reflecting it. . . Comparing the coloured rings made by reflection, with those made by transmission . . . I found that white was opposed to black,



Complementary Colors, Fig. 6 Colors revealed when cellophane tape is sandwiched between two polarizing filters. The axes of the filters are parallel on the *left*, at 45° in the *center*, and at 90° on the *right*

red to blue, yellow to violet and green to a compound of red and violet” [14, book 2, part I, observ. 9].

Polarized Light and Cellophane Tape

Polarized light can also reveal striking color contrasts when cellophane tape is sandwiched between two polarizing filters. The filters only transmit light that is vibrating in one plane which means that each filter has what might be called an axis. When two filters are superimposed, they will transmit more or less light depending on the relationship between the axes. If the axes are at right angles to each other, no light is transmitted. The cellophane tape has the effect of modifying the light as it passes through so that certain wavelengths, visible as colors, are transmitted even when the axes are at right angles. This is demonstrated in Fig. 6.

The colors that are seen in the stripes depend on the relationship between the axes and on how many layers of tape there are – in this case one, two, and three layers, the three-layer stripes being the ones in the center. The colors change radically as one filter is rotated in relation to the other. The colors on the left in Fig. 6 could be described as being complementary to those in the corresponding positions on the right. A more detailed, but simple, explanation of this effect is provided for the “Polarized Light Mosaic” which is included in the *Science Snackbook* compiled by the Exploratorium Teacher Institute [15, p. 78].

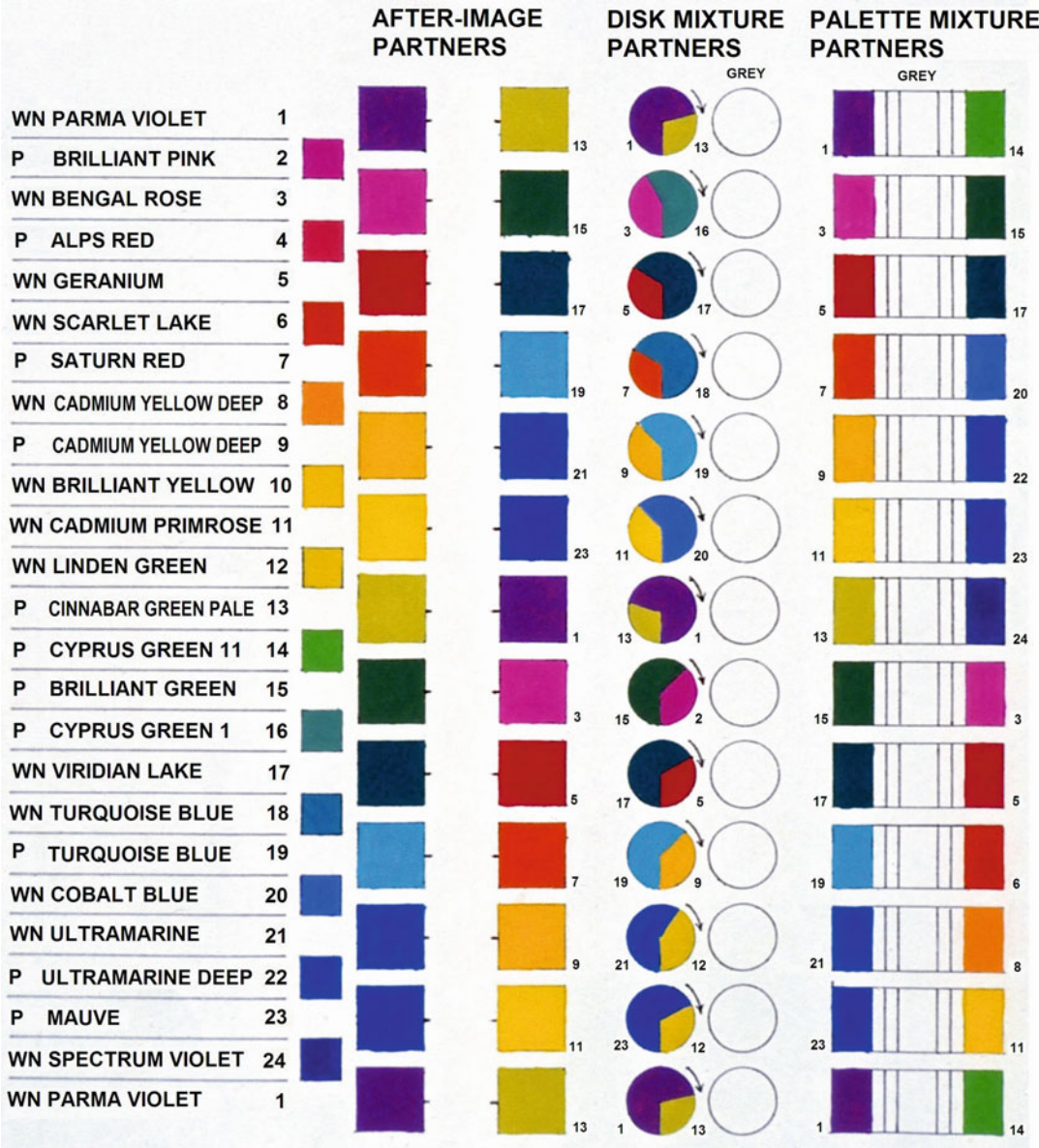
Contradictory Complementaries

It is clear from the above that there is no single complementary color for each hue in the color circle. A study to establish complementaries by different definitions was carried out in 1981. Gouache paints, manufactured by Winsor and Newton (WN) and Pelikan (P), were used for this study. The results are illustrated in Fig. 7.

While carrying out the study, it was assumed that two colors can be each other’s afterimage. Given the findings of Sally Douglas, shown above in Fig. 2, such a color circle cannot be constructed except to suggest that opposite colors are close to being each other’s afterimage. The afterimage of unique blue may be a yellow orange, but the afterimage of that yellow orange may be a blue that is very slightly reddish.

An Elastic Color Circle

The issue of contradictory complementaries is discussed in the entry on the color circle in this encyclopedia. An elastic color circle is proposed as a way of illustrating how the number of steps between the unique hues would need to be increased or decreased to bring differently defined complementary pairs opposite to each other. The illustration from the color circle entry is repeated here as Fig. 8.

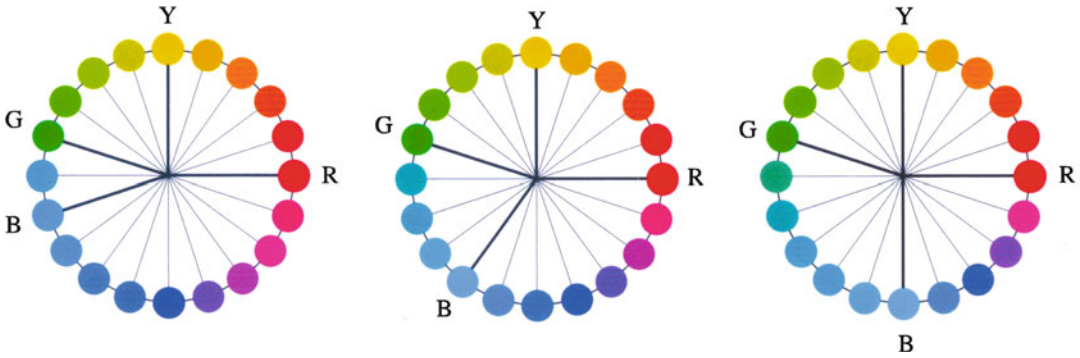


Complementary Colors, Fig. 7 Chart showing the results of a study to find complementary color pairs (partners) according to different definitions

Complementary Colors as a Guide to Mixing

Complementary colors can be used as a guide to mixing paints, inks, or lights. Given that the results of subtractive and additive/partitive

mixture are not the same, the elastic color circle can help to clarify the situation. Figure 9 shows an extreme example of the way in which the same two paints can be used to produce radically different results from subtractive and partitive mixture. Here the blue and the yellow



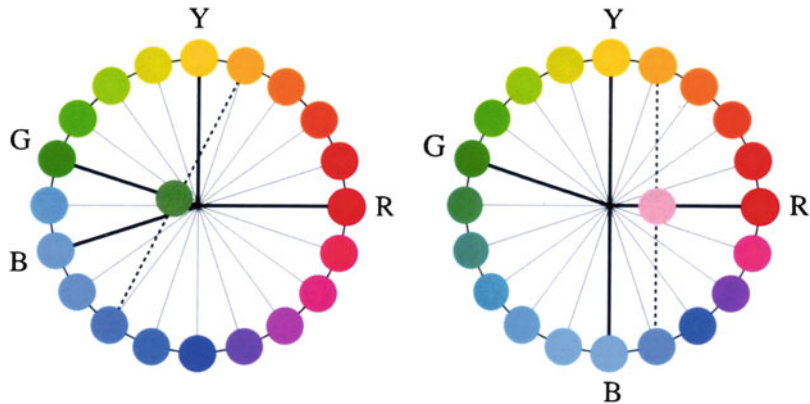
Complementary Colors, Fig. 8 Color circles based on subtractive mixture complementaries (*left*), afterimage complementaries (*center*), and additive mixture

complementaries (*right*). The color circles illustrated here were developed from the results of the study shown above in Fig. 7



Complementary Colors, Fig. 9 The same two paints mix subtractively to a dull green and partitively to dull pink

Complementary Colors, Fig. 10 Dotted lines trace the possible results from subtractive (*left*) and additive/partitive mixtures (*right*) of a reddish blue and a reddish yellow



are more reddish than the blue and yellow in Fig. 5. The result of subtractive mixture is a dull green, but on the spinning disc, the result is a dull pink.

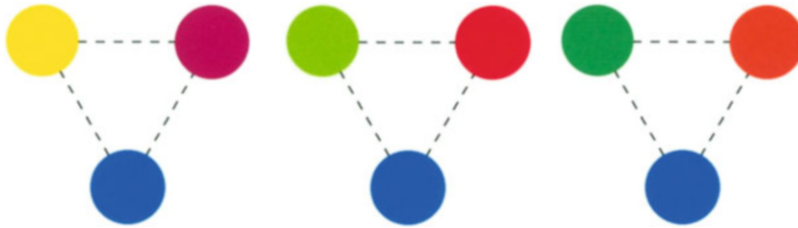
Lines connecting that blue and yellow pass though green on the subtractive circle but through

pink (red) on the additive/partitive circle, as shown in Fig. 10.

Painters are often advised to use complementary colors, rather than black, white, or gray, to modify the appearance of colors in their painting that they feel are too vivid. Black,



Complementary Colors, Fig. 11 Blue with three different complementary colors for three different definitions: (*left to right*) subtractive mixture, afterimages, and additive/partitive mixture



Complementary Colors, Fig. 12 Blue as one of three colors in combinations found by an equilateral triangle placed inside each of three color circles where

complementaries are established by, from *left to right*, subtractive mixture, afterimages, and additive/partitive mixture

white, and gray paints are felt to have a deadening effect.

Complementary Colors as a Guide to Color Harmony

Complementary colors underpin many theories of color harmony, notably the theories of Johannes Itten, author of what may still be the most influential and widely used books on color theory for students of art and design today [16, 17]. Itten argues that “the concept of color harmony should be removed from the realm of subjective attitude into that of objective principle” [17, p. 19]. He is unequivocal when he claims that “we can make the general statement that all complementary pairs . . . are harmonious” [17, p. 21]. And he supports this claim by offering three objective means of establishing complementary relationships: “the after-image always turns out to be of the complementary color . . . colors (lights) are harmonious if they mix to give white . . . colors (paints or pigments) are . . . harmonious if their mixture yields a neutral gray” [17, pp. 19–20]. These three definitions cannot coexist in a single color circle. There would need to be separate circles, one for each

definition, if complementary colors are to be opposite to one another, and that would mean three possible frameworks for harmonious color combinations. If blue is to be one of the colors in a harmonious two-color combination, the other color would be orange, orange yellow, or yellow, depending on which circle was used. These combinations are shown in Fig. 11.

According to Itten, any regular geometric figure can be placed inside the color circle, and its vertices will connect with colors that will form a harmonious combination. An equilateral triangle can be used to find three-color combinations, or triads. Again, with blue as a constant, three rather different combinations would be found in the three circles. These are shown in Fig. 12.

Choosing a Single Definition as the Key to Harmony

Robert Hirschler discusses the potential for confusion arising from different ways of defining complementary colors. Hirschler quotes David MacAdam who was concerned with problems of measurement: “only that definition which states that the optically additive mixture of two complementary colors must match some arbitrarily

assigned ‘neutral’ stimulus is sufficiently specific” [18, p. 1]. This way of establishing complementaries appealed to Ostwald and he used it as the basis for his theories of color harmony. And it is certainly true that optical mixture with spinning discs (partitive mixture) was the easiest way to establish precise complementary relationships during the study described above. A large range of vivid and consistent colors can be produced when gouache paints are used straight from the tube. Paper discs were painted and interleaved in very many combinations of two. The proportions were adjusted until particular combinations in particular proportions were found which would spin to appear neutral gray. These are illustrated in Fig. 7. However, it does not follow that the degree of precision obtainable with this process means that this must be the only “correct” way of establishing complementary color pairs which would be, by definition, harmonious. For Goethe, the key to harmony was afterimages, “a natural phenomenon immediately applicable to aesthetic purposes” [9, p. 320]. Afterimages, as subjective experiences, are not so easy to measure but they involve the observer more intimately, and the experience of an afterimage suggests that the visual system is seeking some form of completeness. Chevreul imagined two spectral reflectance curves which would produce white light when combined. A color stimulus and its afterimage could be the perceptual equivalent.

Complementary Colors as a Fuzzy Concept

Artists and designers, who are looking for some formula that will lead to harmonious color combinations, may be perplexed to find that the color circle is unstable as it adjusts to comply with different definitions of complementary colors. Rather than worry about which definition, and its corresponding color circle, is “correct,” it might be more fruitful to accept complementary colors

as a fuzzy concept and make decisions based on personal judgment. After studying the color combinations illustrated in Figs. 11 and 12, a choice can be made. If one combination seems more satisfying and harmonious than the other two, the corresponding color circle could be adopted and Itten’s ideas applied. This could give the artist or designer a feeling of ownership while still providing a sense of security with a clear framework for developing harmonious color combinations. Alternatively a more flexible approach could be taken. If information about the likely results of mixing is needed, then reference could be made to the subtractive or additive circles as appropriate. But as a framework for developing harmonious color combinations, the color circle could be regarded as offering a choice within limits. While a given color may have more than one complementary, as different definitions are applied, those complementary colors will not be radically different in hue. Even in the most extreme example illustrated here, the alternative complementaries for blue, there is not a large spread between yellow and orange. This range can be considered as offering a choice with the decision being left to the judgment of the artist or designer. It is worth noting that Alberti and Leonardo both nominated more than one color as being “sympathetic” or “contrary.” Alberti mentions red in relation to blue and green and Leonardo lists red or purple or mauve as going well with green.

This more flexible attitude to the color circle is recommended by Deryck Healey. Having introduced a 12-hue color circle based on subtractive mixture, with the printers’ cyan, magenta, and yellow as primaries (the so-called process inks), Healey illustrates combinations that deviate somewhat from strict adherence to Itten’s rules. He advises the reader to “study these examples of harmony and contrast. Few are strictly precise triads, split complementaries and so on; rather they illustrate how such themes may be subtly interpreted and still retain the desired characteristics of their general category” [19, p. 40].

Conclusion

Complementary colors, as defined by mixture to a neutral or by afterimages, provide reference points for the relationships between hues in a color circle. When the relative positions of hues are determined by mixture, the different results from subtractive and additive/partitive mixtures can be recognized and the appropriate circle used as a guide. More problematic is the role of complementary colors in theories of color harmony. Different definitions lead to different relationships between the hues in the color circle. This could lead one to question the theories or abandon them altogether. Nevertheless, the theories are interesting and can be helpful, especially for those who lack confidence in their own intuition. If a choice has to be made between the different circles, there are arguments in favor of objectivity and subjectivity. On the side of objectivity, there is the circle based on additive/partitive mixture where the color pairs are most readily measurable. This has the support of MacAdam and Ostwald. On the side of subjectivity is the circle based on afterimages which relates more directly to the personal experience of color and the judgments of beauty and harmony as made by the observer. This has the support of Goethe. In the end it is up to the artist or designer to choose a single circle or to take a more flexible approach as recommended by Healey. An advantage of the elastic color circle is that it allows one to stretch the rules without breaking them.

Cross-References

- [MacAdam, David L.](#)
- [Munsell, Albert Henry](#)
- [Newton, \(Sir\) Isaac](#)
- [Ostwald, Friedrich Wilhelm](#)
- [Palette](#)
- [Pigments](#)
- [Unique Hues](#)

References

1. Thompson, D. (ed.): The Concise Oxford Dictionary. Oxford University Press, Oxford (1995)
2. Sadie, S. (ed.): The New Grove Dictionary of Music and Musicians. Macmillan, London (1980)
3. Green-Armytage, P.: Complementary colours – description or evaluation? In: Sivik, L. (ed.) Colour and Psychology: From AIC Interim Meeting 96, Colour Report F 50, pp. 205–209. Scandinavian Colour Institute, Stockholm (1996)
4. Alberti, L.B.: On Painting. Penguin, London (1991). 1435
5. da Vinci, L.: Leonardo On Painting. Yale University Press, New Haven (1989 [1518])
6. Kemp, M.: Note. In: Leonardo on Painting. Yale University Press, New Haven (1989)
7. Kemp, M.: The Science of Art. Yale University Press, New Haven (1990)
8. Harris, M.: The Natural System of Colours. Whitney Library of Design, New York (1963 [1776])
9. von Goethe, J.W.: Theory of Colours. MIT Press, Cambridge, MA (1970 [1810])
10. Cabot Hale, N.: Abstraction in Art and Nature. General Publishing Company, Toronto (1993 [1972])
11. Bagley, M-O.: Color in design education. In: Arnkil, H., Hämäläinen, E. (eds.) Aspects of Colour. University of Art and Design, Helsinki (1995)
12. Ostwald, W.: The Color Primer. Van Nostrand Reinhold, New York (1969 [1916])
13. Chevreul, M.-E.: The Principles of Harmony and Contrast of Colors and their Application to the Arts. Reinhold Publishing, New York (1967 [1839])
14. Newton, I.: Opticks. Impression Anästaltique, Culture et Civilisation, Brussels (1996 [1704])
15. Institute Exploratorium Teacher: Science Snackbook. The Exploratorium, San Francisco (1991)
16. Itten, J.: The Art of Color. Van Nostrand Reinhold, New York (1961)
17. Itten, J.: The Elements of Color. Van Nostrand Reinhold, New York (1970)
18. Hirschler, R.: Teaching colour wheels and complementary colours. In: Kortbawi, I., Bergström, B., Fridell Anter, K. (eds.) Colour – Effects and Affects, Interim Meeting of the AIC, Proceedings, paper 98. Scandinavian Colour Institute, Stockholm (2008)
19. Healey, D.: Living with Colour. Macmillan, London (1982)

- [Chevreul, Michel-Eugène](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [Color Circle](#)
- [Color Combination](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Color Mixture](#)
- [Goethe, Johann Wolfgang von](#)
- [Itten, Johannes](#)

Compositing and Chroma Keying

Jorge Lopez-Moreno
GMRV Group, Universidad Rey Juan Carlos,
Móstoles, Madrid, Spain

Synonyms

Blending; Color-separation overlay; Deep compositing; Layering; Matting

Definition

Compositing is the act of combining two images or video sequences, producing a new image or video sequence. There are several techniques to composite two images (or video frames): alpha matte, gradient-domain blending, deep compositing, multiply, overlay and screen modes, etc.

Chroma keying is the name of one of such techniques which uses color hues (chroma) to guide the compositing process. A portion of the target image is selected (based on a color or a range of colors) and substituted by the image to be inserted. This technique is widely adopted in video editing and postproduction.

Origin of the Terms

Traditionally, both in TV and films, there have been four basic compositing techniques: matting, physical compositing, background projection, and multiple exposure.

- **Matting** is currently the most widespread technique and corresponds to the general definition given above. Digital compositing relies entirely on variations of this technique: *chroma keying* is an example of matting.
- **Physical compositing**: When capturing the background image, an object is physically introduced into the scene to be composited as foreground image. For instance, a *glass shot* consists of recording a scene through a

transparent glass where some elements (or most of it) are painted on the glass (some background buildings, for instance). The area of the frame where the action happens is left clear.

- **Multiple exposure**: One of the earliest compositing techniques ever developed, achieved by recording multiple times with the same film, but exposing a different part each time with the help of a mask over the lens. Georges Méliès, pioneer of visual effects, used it to obtain multiple copies of himself in the film the *One-man band* (Fig. 1).
- **Background projection**: This technique, currently fallen in disuse, is based on projecting the desired video or image onto a background screen with the foreground elements (actors, objects) between the camera and this screen. The development of digital compositing and its own complexities (synchronization issues, illumination constraints) rendered this method obsolete.

Techniques

This section describes some of the standard techniques in image compositing together with the latest advances in the last decade. For further exploration of digital compositing, visual effects, and specific examples in the film industry, please refer to the work of Ron Brinkmann [1].

Chroma Keying

This method renders a range of colors in the foreground image as transparent, revealing the image below. Blue and green are the most used color in films, videogames, and TV due to their hue distance from the human skin tone. The main disadvantage of chroma keying is that it requires a lighting set with a (sometimes rather large) chroma screen which has to be as evenly illuminated as possible in order to minimize the range of color variations (noise) in the background (Fig. 2). A secondary issue is that the object (or person) to be inserted cannot have any of the hue values used for chroma keying (e.g., a green dress).

Compositing and Chroma Keying,

Fig. 1 One of the first examples of compositing in films. A frame from George Méliès *One-man band* movie (1900)



Compositing and Chroma Keying,

Fig. 2 Example of chroma key compositing (*green screen*). The actress is overlaid with a synthetic background in real time on camera (Source: Devianart artist: AngryDogDesigns)



Green is generally preferred over blue due to the lower energy required to produce an even illumination over the chroma screen. However, sometimes blue is used whenever there is a risk of green tones appearing in the foreground layer (e.g., outdoor scenes with vegetation which produce green interreflections). In computer graphics, chroma keying is usually obtained as a function of RGB values which measures the differences from the range of colors used for chroma keying (which is analogous to finding the distance to a closed 3D surface in space color):

$$\begin{aligned}\alpha(p) &= f(R_0, G_0, B_0) \\ &= d(R_0, R_{ck}) + d(G_0, G_{ck}) + d(B_0, B_{ck})\end{aligned}$$

where d is a distance function (e.g., *Euclidean distance*) and both (R_0, G_0, B_0) and (R_{ck}, G_{ck}, B_{ck}) are the red, green, and blue channel values, respectively, of pixel p in the image and the color used for chroma keying.

Alpha Blending

Compositing is based on the information stored in the alpha channel of the inserted image.

Introduced for the first time in the image editor *Paint3* in 1970, the *alpha channel* [2, 3] stores a value between 0 and 1, 0 being a pixel which is completely transparent and 1 completely opaque. This channel in a 2D image is a grayscale image by itself called *matte*, which can be visualized and edited. Multiple file formats support alpha-extended data (like RGBA): PNG, TIFF, TGA, SGV, and OpenEXR.

Although binary alpha masking (exclusively 0 or 1) is generally suitable for several compositing scenarios, complex visual phenomena such as transparency or translucency require a subtle gradation of alpha values (e.g., separating a dog from its background requires dealing with strands of hair which capture and scatter the environment light).

Still an open problem, researchers have developed complex methods to obtain these subtle matte masks with the minimal user input. For instance, Levin et al. [4] propose a method called *spectral matting*, which finds clusters of pixels with affinity properties (such as X, Y, and/or RGB distances) by relying on spectral analysis (with matte Laplacian) to evaluate automatically the quality of a matte without explicitly estimating the foreground and background colors. This method requires less user input than most approaches (four strokes on average) in order to distinguish the foreground from the background.

Blend Modes

The majority of the image compositing software includes multiple blending modes, that is, different functions to mix each pair of pixels from two overlaid images depending on their RGB values and the layering order. There are several well-known modes [5]: multiply, screen, overlay, soft light, hard light, dissolve, color dodge, addition, subtraction, darken, lighten, etc. Some of them might differ in implementation and formulation, but some of the most established and agreed modes are described here:

- **Multiply:** Both pixel values are multiplied. The result tends to be darker than any of the original images: dark (black) values are preserved, while white values have no effect in the

final composition. This mode is especially useful when combining black and white line drawings with color images.

- **Screen:** Considered as the opposite of the previous mode. Both layers are inverted and multiplied. The final result is inverted again. It tends to produce brighter images than the original, as bright values are preserved and black pixels have no effect. If a denotes the background pixel and b the foreground pixel, the following equation shows how the screen blend mode is applied:

$$f(a, b) = 1 - (1 - a)(1 - b)$$

- **Overlay:** A hybrid mode based on a combination of multiply and screen modes, guided by the value of the background layer. Foreground colors overlay the background while preserving its highlights and shadows:

$$f(a, b) = \begin{cases} 2ab & \forall a \leq 0.5 \\ 1 - 2(1 - a)(1 - b) & \forall a > 0.5 \end{cases}$$

Gradient-Domain Compositing

Gradient-domain techniques aim to merge two images, making the boundary between them imperceptible. They rely on image gradients (differences of pixel values instead of absolute values), looking for the composite that would produce the smoothest fuse of the gradient field (see Fig. 3). These approaches were introduced by Perez et al. in 2003 [6], being now the standard in commercial software (like the *healing brush tool* in Adobe Photoshop).

The latest research advances combine gradient domain with multi-scale methods and visual transfer to produce compositions which mimic even the structural noise at different spatial levels [7].

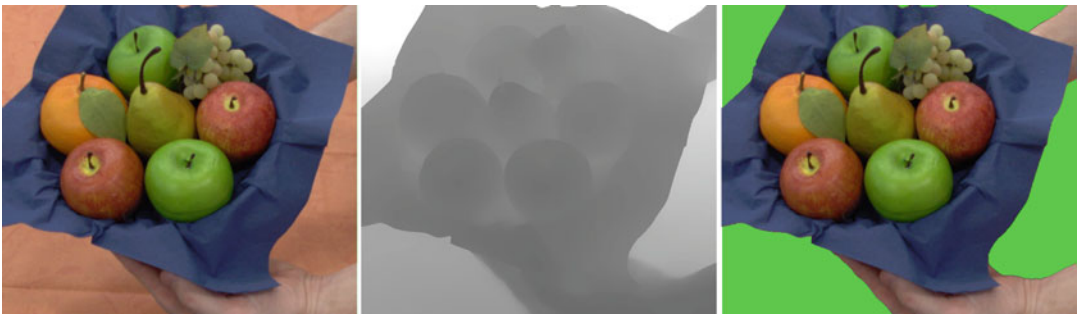
Deep Compositing

Deep compositing techniques, in addition to the usual color and opacity channels, take into account depth information stored at each pixel along the z-axis, perpendicular to the image plane (see Fig. 4). In opposition to traditional compositing, which arranges 2D layers in 3D



Compositing and Chroma Keying, Fig. 3 Example of image compositing based on gradient domain and proximity matching (Photoshop). An area from the input image (left) is selected with a loose stroke (middle), from which a

radius is derived to analyze gradient and color affinity. The result is shown in the right image. Some artifacts are still noticeable and additional user input might be required



Compositing and Chroma Keying, Fig. 4 Example of deep compositing [8]. The fruit basket in the input image (left) is extracted from the background (right image, colored in green) by using the associated depth range data (middle image)

with a single depth value for each element, deep compositing stores a range of depth values for each object, extending the gamut of editing possibilities. For instance, if the compositing artist aims to integrate an actor with a 3D rendered column of billowing smoke, without 3D deep data in the smoke element, the actor will appear to be in front or behind the column. However, with varying density values in the Z-axis, the actor can be placed *inside* the smoke and show both correct occlusion and partial visibility effects.

In order to mix and compose deep images, a proper depth buffer is required. This data is available for traditional 3D rendered graphics, as each pixel has a depth value associated (Z-buffer), whereas for footage obtained from camera, reliable depth values are available only at some arbitrary specified points. This sparsity is due to the limitations of the capturing device (stereo, time-of-flight cameras, etc.). The remaining pixels are

then obtained through interpolation. For instance, Richardt et al. [8] incorporate a time-of-flight IR camera to a consumer-level video camera in order to capture a rough depth map which is subsequently filtered to obtain a high-quality depth image by means of spatiotemporal denoising and an upsampling scheme (see Fig. 4).

Although one of the first uses was Pixar RenderMan's deep shadow technique [9], this technology has been progressively adopted by the media industry with the upsurge of 3D cinema and TV. Nowadays, most companies like Weta or DreamWorks rely on deep compositing pipelines and tools such as Nuke (The Foundry) to create their final compositions.

Visual Transfer and Relighting

The concepts of visual transfer and relighting refer to a set of techniques that aim to transfer some visual properties from the background to the image to be inserted. Illumination (shading,

Compositing and Chroma Keying,

Fig. 5 Example of image capture in a Lightstage [9]. The actor is illuminated with an even omnidirectional light in order to capture the base reflectance of his face. In further shots at high speed, the light sources are switched on and off individually to capture the interaction with each light source for future interpolation



shadows, and highlights) is one of the main visual factors to be considered to homogenize a composition. When real objects have to be introduced into CGI environments with known illumination, this illumination is mimicked with actual light sources in a stage. In the early years of cinema, this technique was necessary when performing background projection (e.g., when recording an actor driving a car by night, the lights from cars or street lamps reproduced in the back screen were synchronized with actual lamps illuminating the actor in the studio). More sophisticated systems like Lightstage [10] have been developed to capture an actor performance.

By combining high-speed cameras and structured light sources from many directions, it is possible to create a database of views of the actor under different light environments which can be interpolated to re-render under arbitrary lighting conditions for compositing in any background (see Fig. 5).

2. Ray Smith A.: Image Compositing Fundamentals. Microsoft Tech Memo 4 (1995)
3. Porter, T., Duff, T.: Compositing digital images. In: SIGGRAPH' 84: Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques, pp. 253–259 (1984)
4. Levin, A., Rav-Acha, A., Lischinski, D.: Spectral matting. *IEEE Trans. Pattern. Anal. Mach. Intel.* **30**, 1699–1712 (2008)
5. Grasso A. (ed.): SVG Compositing Specification. W3C Working Draft (2011)
6. Perez, P., Gangnet, M., Blake, A.: Poisson image editing. *ACM Trans. Graph.* **22**(3), 313–318 (2003)
7. Sunkavalli, K., Johnson, M.K., Matusik, W., Pfister, H.: Multi-scale image harmonization. *ACM Trans. Graph.* **125**, 1–125 (2010). SIGGRAPH, 10
8. Richardt, C., Stoll, C., Dodgson, N., Siedel, H.-P., Theobalt, C.: Coherent spatiotemporal filtering, upsampling, and rendering of RGBZ videos. *Cmp. Graph. Forum. Proc. of Eurographics. Eurographics Association, Cagliari, Sardinia* (2012)
9. Lokovic, T., Veach, E.: Deep shadow maps. In: *ACM SIGGRAPH 2000*, pp. 85–392. (2000)
10. Debevec, P.: Virtual cinematography: relighting through computation. *IEEE Comput.* **39**(8), 57–65 (2006)

Cross-References

► [Global Illumination](#)

References

1. Brinkmann, R.: *The Art and Science of Digital Compositing*. Morgan Kaufmann, San Francisco (1999)

Computer Depiction

► [Non-Photorealistic Rendering](#)

Computer Modeling

► [Color Category Learning in Naming-Game Simulations](#)

Cone Fundamentals

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

Synonyms

CIE fundamental color matching functions; Cone spectral sensitivities

Definition

The three cone fundamentals are the spectral sensitivities of the long- (L-), middle- (M-), and short- (S-) wavelength cones measured relative to light entering the cornea. They are also the fundamental color matching functions (CMFs), which in colorimetric notation are referred to as $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$. The simple identity between the cone fundamentals of color matching and the cone spectral sensitivities depends on phototransduction and its property of *univariance*. The absorption of a photon produces a photoreceptor response that is independent of photon wavelength, so that all information about wavelength is lost. With one cone type, vision is monochromatic. With three cone types, vision is trichromatic.

Trichromacy means that for observers with normal color vision, the color of a test light of any chromaticity can be matched by superimposing three independent primary lights (with the proviso that one of the primaries sometimes must be added to the test light to complete the match). The amounts of the three primary lights required to match test lights as a function of test wavelength are the three CMFs for those primary lights (usually defined for matches to test lights of equal energy). Because the cone spectral sensitivities overlap, it is not possible for any real light to stimulate just one of them. However, the cone fundamentals are the three CMFs for the three *imaginary* primary lights that would uniquely stimulate individual cone types (i.e.,

imaginary lights that produce the three “fundamental” sensations that underlie color vision). The cone fundamental CMFs define all other CMFs and must be a linear transformation of them.

The cone fundamentals can be determined directly by measuring cone spectral sensitivities using “color-deficient” observers lacking one or two cone types and/or special conditions to isolate single-cone responses. They can also be derived by the linear transformation of CMFs measured using real primary lights, but for that the coefficients of the linear transformation must be known.

Current estimates of the cone fundamentals [1–3] use spectral sensitivity measurements to guide the choice of the coefficients of the linear transformation from a set of measured CMFs to the cone fundamental CMFs.

Overview

The spectral properties of the cone fundamentals and that of color matching, in general, are determined principally by the way in which the cone photoreceptors interact with light at the very first stage of vision. They depend, in particular, on photon absorptions by the cone photopigment and on how the probability of photon absorption varies with wavelength. To understand that, we start at the molecular level.

Phototransduction

The cone photopigment molecule is made up of a transmembrane opsin, a G protein-coupled receptor protein, bound to a chromophore, 11-*cis* retinal. The absorption of a photon provides the energy needed to isomerize the chromophore from its 11-*cis* form to its all-*trans* form; this change in shape activates the opsin and triggers the phototransduction cascade and the neural response. The likelihood that a given photon will produce an isomerization depends upon how closely its energy matches the energy required to initiate the isomerization. Crucially, this energy varies with cone type because of differences in key amino acids in those parts of the opsin molecule that surround the chromophore. These amino

acids modify the isomerization energy and thus the spectral sensitivity of the photoreceptor [4]. In most observers with normal color vision, there are four photoreceptor classes: three types of cone photoreceptors, L-, M-, and S-cones, and a single type of rod photoreceptor.

Cone fundamentals, since they are measured behaviorally in terms of energies measured at the cornea, also depend on the absorption of photons by the optical media and on the density of photopigment in the photoreceptor outer segment, both of which vary between observers (see below). Thus, the cone fundamentals are not *precisely* related to the spectral properties of the photopigments (called the absorbance or extinction spectra).

Univariance

The relatively simple relationship between the cone fundamentals and color matching arises because of the way in which cone (and rod) photoreceptors transduce absorbed photons. When a photon is absorbed to initiate the phototransduction cascade, the effect is all or nothing and is consequently independent of photon wavelength. Photoreceptor outputs thus vary *univariantly* according to the number of photons that they absorb [5], as a result of which wavelength and intensity are confounded, and individual photoreceptors are “color blind.” A change in the rate of photon absorption could be due to a variation in light intensity, but equally, it could be due to a variation in wavelength.

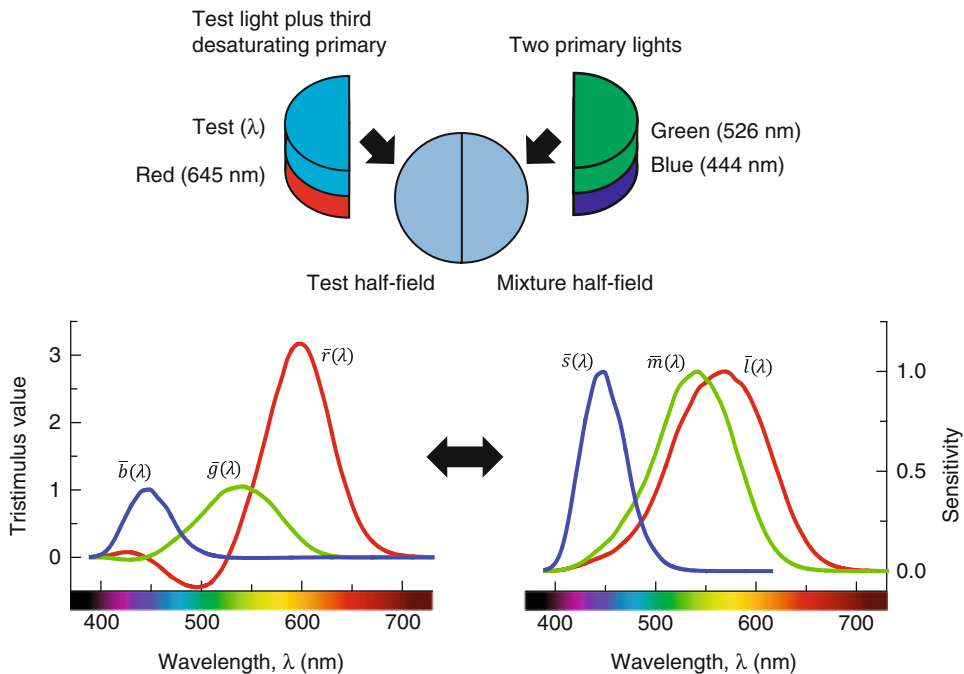
With only one cone type, vision is monochromatic and reduced to a single dimension: two lights of any spectral composition can be made to match simply by equating their intensities (a relationship defined by the cone type’s spectral sensitivity). With only two cone types, vision is dichromatic and reduced to two dimensions: lights of any spectral composition can be matched with a mixture of two other lights. Dichromatic human observers fall into three classes, protanopes, deuteranopes, and tritanopes, depending upon whether they lack L-, M-, or S-cones, respectively. Observers with normal color vision have three classes of cone photoreceptor, and their vision is trichromatic. Color

vision depends on comparing the univariant outputs of different cone types.

Trichromacy

A consequence of trichromacy is that the color of any light can be matched with three specially selected or “independent” primary lights of variable intensity (chosen so that no two will match the third). These primary lights are frequently red (R), green (G), and blue (B), but many other triplets are possible. The upper panel of Fig. 1 shows a typical color matching experiment, in which an observer is presented with a half-field illuminated by a “test” light of variable wavelength λ and a second half-field illuminated by a mixture of red, green, and blue primary lights. At each λ , the observer adjusts the intensities of the three primary lights, so that the test field is perfectly matched by the mixture of primary lights. The lower left-hand panel of Fig. 1 shows the mean $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs obtained by Stiles and Burch [6] for primary lights of 645, 526, and 444 nm. Notice that except at the primary wavelengths one of the CMFs is negative. There is no “negative light”; rather, these negative values indicate that the primary in question must be added to the spectral test light to make a match (as illustrated in the panel for the red primary). Real primaries give rise to negative values because real lights do not uniquely stimulate single-cone photoreceptors (see Figs. 2 and 3). The cone fundamental CMFs are always positive.

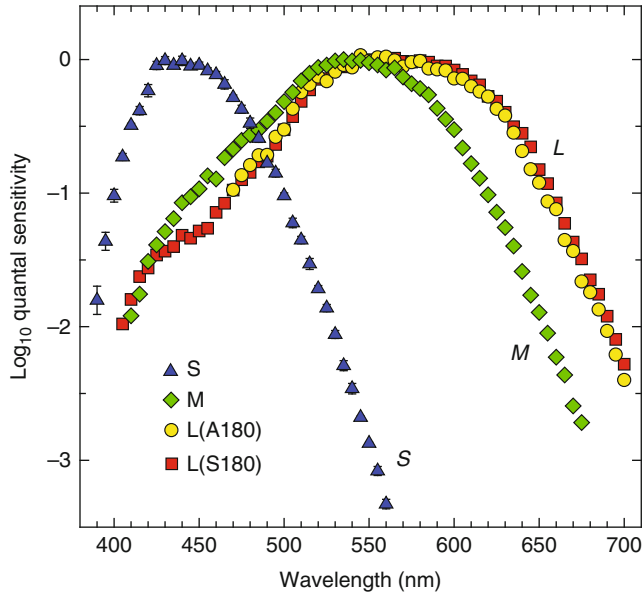
CMFs, such as the ones shown in the lower-left panel, can be linearly transformed to any other set of real primary lights and to the *fundamental* primaries that are illustrated in the lower right-hand panel of Fig. 1. The three fundamental primaries (or “Grundempfindungen” – fundamental sensations) are the three imaginary primary lights that would *uniquely* stimulate each of the three cones to yield the L-, M-, and S-cone spectral sensitivity functions (such lights are not physically realizable because of the overlapping spectral sensitivities of the cone photopigments). All other sets of CMFs depend on the fundamental CMFs and should be a linear transformation of them. Note that the fundamental CMFs shown here are comparable to those shown in Fig. 3 but

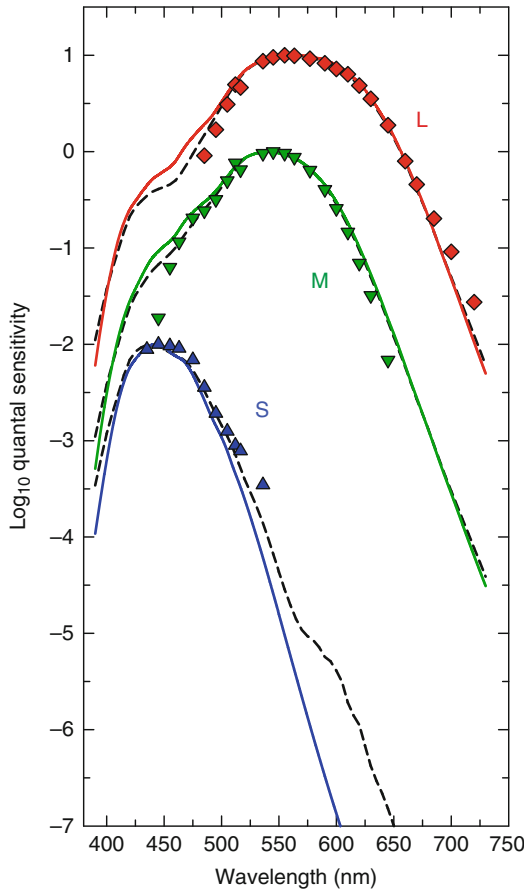


Cone Fundamentals, Fig. 1 A monochromatic test field of wavelength, λ , can be matched by a mixture of red (645 nm), green (526 nm), and blue (444 nm) primary lights, one of which must be added to the test field to complete the match (*upper panel*). The amounts of each of the three primaries required to match monochromatic lights spanning the visible spectrum are the $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs (*red, green, and blue lines, respectively*) shown in the lower left-hand panel. These CMFs were measured

using 10-deg diameter targets by Stiles and Burch [6]. A negative sign means that the primary must be added to the target to complete the match. CMFs can be linearly transformed from one set of primaries to another and to the fundamental primaries. Shown in the lower right-hand panel are the 10-deg $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs linearly transformed to give the $\bar{l}(\lambda)$, $\bar{m}(\lambda)$, and $\bar{s}(\lambda)$ 10-deg cone fundamental primary CMFs (*red, green, and blue lines, respectively*)

Cone Fundamentals, Fig. 2 Mean cone spectral sensitivity data [10, 11]. L-cone data from deuteranopes with either the L(S180) (*red squares, $n = 17$*) or L(A180) (*yellow circles, $n = 3$*) polymorphism, M-cone data from protanopes (*green diamonds, $n = 9$*), and S-cone data (*blue triangles*) from S-cone monochromats (*$n = 3$*) and below 540 nm (measured under intense long-wavelength adaptation) from normal observers (*$n = 5$*)





Cone Fundamentals, Fig. 3 Comparisons between estimates of the 2-deg L-, M-, and S-cone fundamentals by Stockman and Sharpe [2] (solid-colored lines), by Smith and Pokorny [1] (dashed lines), and by König and Dieterici [17] (symbols)

are plotted as linear sensitivities rather than as the more usual logarithmic sensitivities.

The relationship between the fundamental CMFs and a set of real CMFs (obtained using, e.g., red, green, and blue primaries) can be stated formally: When an observer matches the test and mixture fields in a color matching experiment, the two fields cause identical absorptions in each of his or her three cone types. The match, in other words, is a match at the cones. The matched test and mixture fields appear identical to S-cones, to M-cones, and to L-cones. For matched fields, the following relationships apply:

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

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

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

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

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

Since only relative cone spectral sensitivities are required, the eight unknowns reduce to five:

$$\begin{pmatrix} \bar{l}_R/\bar{l}_B & \bar{l}_G/\bar{l}_B & 1 \\ \bar{m}_R/\bar{m}_B & \bar{m}_G/\bar{m}_B & 1 \\ 0 & \bar{s}_G/\bar{s}_B & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix} \quad (3)$$

A definition of the cone fundamental CMFs in terms of real CMFs requires a knowledge of the coefficients of the transformation.

Spectral Sensitivity Measurements

The transformation in Eq. 3 can be estimated by comparing dichromatic and normal color matches [7]. Dichromats confuse pairs of colors that trichromats do not. When these confusions are plotted in a normal chromaticity diagram, they yield characteristic lines of confusion that converge to a different confusion point for each type of dichromat. The three confusion points correspond to the chromaticities of the missing imaginary fundamental primaries, from which the transformation matrix can be derived.

An alternative, straightforward way of estimating the transformation matrix is to measure the three cone spectral sensitivities directly. This can be achieved using steady or transient chromatic backgrounds to selectively adapt one or two of the

cone types to isolate the third [8, 9]. However, cone isolation can more easily be achieved by using chromatic adaptation in observers in dichromats lacking one (or two) of the three cone types. With the S-cones selectively adapted, L- and M-cone spectral sensitivities can be directly measured in deuteranopes without M-cone function and in protanopes without L-cone function. Figure 2 shows the mean spectral sensitivity data obtained from nine protanopes (green diamonds), from seventeen single-gene L(S180) deuteranopes with serine at position 180 of their L-cone photopigment opsin gene (red squares), and from five single-gene L(A180) deuteranopes with alanine at position 180 (orange circles) [2, 10]. L(A180) and L(S180) are two commonly occurring L-cone photopigment polymorphisms in the normal human population that differ in λ_{\max} by about 2.5 nm [10]. Figure 2 also shows the mean S-cone spectral sensitivities obtained from three S-cone monochromats, who lack L- and M-cones, and under intense long-wavelength adaptation at wavelengths shorter than 540 nm obtained from five normal subjects [11]. Importantly, thanks to molecular genetics, we can now choose dichromats or monochromats for these experiments whose remaining cone photopigments are normal.

Cone Fundamentals

The spectral sensitivities shown in Fig. 2 were used by Stockman and Sharpe [2] to find the linear combinations of the $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs that best fit each measured cone spectral sensitivity, allowing adjustments in the densities of pre-receptor filtering and photopigment optical density in order to account for differences in the mean densities between different populations and different target sizes. The transformation matrix for the Stockman and Sharpe 10-deg cone fundamentals is

$$\begin{pmatrix} 2.846201 & 11.092490 & 1 \\ 0.168926 & 8.265895 & 1 \\ 0 & 0.010600 & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} \quad (4)$$

where $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ CMFs are the Stiles and Burch CMFs measured with a 10-deg diameter test field [6]. The Stockman and Sharpe 2-deg estimates

are based on the same transformation but the cone fundamentals have been adjusted to macular and photopigment optical densities appropriate for a 2-deg target field. The 2-deg functions are shown in Fig. 3 as the solid-colored lines. The Stockman and Sharpe 2-deg and 10-deg functions have been adopted by the Commission Internationale de l'Éclairage (CIE) as the 2006 physiologically relevant cone fundamental CMFs [3].

The quality of cone fundamentals depends not only on the correct transformation matrix but also in the CMFs from which they are transformed. The Stiles and Burch 10-deg CMFs [6], which were measured in 49 subjects from approximately 390–730 nm (and in nine subjects from 730 to 830 nm), are probably the most secure and accurate set of existing color matching data. Other CMFs are less secure and typically flawed [12].

Most other cone fundamentals are also given in the form of Eq. 4 but for different underlying CMFs [1, 2, 13–16]. The most widely used have been those by Smith and Pokorny [1]. Their transformation matrix is

$$\begin{pmatrix} 0.15514 & 0.54312 & -0.03286 \\ -0.15514 & 0.45684 & 0.03286 \\ 0 & 0.00801 & 1 \end{pmatrix} \begin{pmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{pmatrix} \quad (5)$$

where $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are the Judd-Vos-modified 2-deg CMFs [15].

Figure 3 shows the Smith and Pokorny estimates as dashed lines and for historical context the much earlier estimates obtained 125 years ago by König and Dieterici [17] as symbols. For the L- and M-cone fundamentals, the discrepancies between the more modern fundamentals are mainly at shorter wavelengths; the discrepancies between the S-cone fundamentals are more extensive.

The functions mentioned here can be found at <http://www.cvrl.org>

Other Factors that Influence Cone Fundamentals

Factors other than the properties of the photopigment also affect the cone spectral sensitivities. They include the density of the pigment in

the lens that absorbs light mainly of short wavelengths, the density of macular pigment at the fovea, and the axial optical density of the photopigment in the photoreceptor. All three factors exhibit individual differences between observers, and the last two vary with retinal eccentricity. These factors should all be taken into account when trying to predict the spectral sensitivities of an individual from standard functions such as those defined by Eqs. 4 and 5 for a given target size and eccentricity. See, for example, Brainard and Stockman [18] for further details.

Cross-References

- ▶ [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- ▶ [CIE Chromaticity Coordinates \(xyY\)](#)
- ▶ [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- ▶ [Comparative Color Categories](#)
- ▶ [Cone Fundamentals](#)
- ▶ [Deuteranopia](#)
- ▶ [Protanopia](#)
- ▶ [Trichromacy](#)

References

1. Smith, V.C., Pokorny, J.: Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Res.* **15**, 161–171 (1975)
2. Stockman, A., Sharpe, L.T.: Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Res.* **40**, 1711–1737 (2000)
3. CIE: Fundamental chromaticity diagram with physiological axes – part 1. Technical report 170–1. Central Bureau of the Commission Internationale de l'Éclairage, Vienna (2006)
4. Deeb, S.S.: The molecular basis of variation in human color vision. *Clin. Genet.* **67**, 369–377 (2005)
5. Mitchell, D.E., Rushton, W.A.H.: Visual pigments in dichromats. *Vision Res.* **11**, 1033–1043 (1971)
6. Stiles, W.S., Burch, J.M.: NPL colour-matching investigation: final report (1958). *Opt. Acta* **6**, 1–26 (1959)
7. Maxwell, J.C.: On the theory of colours in relation to colour-blindness. A letter to Dr. G. Wilson. *Trans. R. Scott. Soc. Arts* **4**, 394–400 (1856)
8. Stiles, W.S.: *Mechanisms of Colour Vision*. Academic, London (1978)
9. Stockman, A., MacLeod, D.I.A., Vivien, J.A.: Isolation of the middle- and long-wavelength sensitive cones in normal trichromats. *J. Opt. Soc. Am. A* **10**, 2471–2490 (1993); Published online Epub Dec
10. Sharpe, L.T., Stockman, A., Jägle, H., Knau, H., Klausen, G., Reitner, A., Nathans, J.: Red, green, and red-green hybrid photopigments in the human retina: correlations between deduced protein sequences and psychophysically-measured spectral sensitivities. *J. Neurosci.* **18**, 10053–10069 (1998)
11. Stockman, A., Sharpe, L.T., Fach, C.C.: The spectral sensitivity of the human short-wavelength cones. *Vision Res.* **39**, 2901–2927 (1999)
12. Stockman, A., Sharpe, L.T. Cone spectral sensitivities and color matching: In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 53–87. Cambridge University Press, Cambridge (1999)
13. Vos, J.J., Walraven, P.L.: On the derivation of the foveal receptor primaries. *Vision Res.* **11**, 799–818 (1971)
14. Estévez, O.: Ph.D., Amsterdam University (1979)
15. Vos, J.J.: Colorimetric and photometric properties of a 2-deg fundamental observer. *Color. Res. Appl.* **3**, 125–128 (1978)
16. Stockman, A., MacLeod, D.I.A., Johnson, N.E.: Spectral sensitivities of the human cones. *J. Opt. Soc. Am. A* **10**, 2491–2521 (1993)
17. König, A., Dieterici, C.: Die Grundempfindungen und ihre Intensitäts-Vertheilung im Spectrum. In: *Sitzungsberichte Akademie der Wissenschaften in Berlin*, vol. 1886, pp. 805–829. (1886)
18. Brainard, D.H., Stockman, A. Colorimetry: In: Bass, M., DeCusatis, C., Enoch, J., Lakshminarayanan, V., Li, G., Macdonald, C., Mahajan, V., van Stryland, E. (eds.) *The Optical Society of America Handbook of Optics. Vision and Vision Optics*, vol. III, 3rd edn, pp. 10.1–10.56. McGraw Hil, New York (2009)

Cone Fundamentals, Stockman-Sharpe

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

Cone Pigments

- ▶ [Photoreceptors, Color Vision](#)

Cone Spectral Sensitivities

- ▶ [Cone Fundamentals](#)

Congenital Color Vision Abnormality

- ▶ [Deuteranopia](#)
- ▶ [Protanopia](#)

Contextual Color Design

- ▶ [Environmental Color Design](#)

Contrary Colors

- ▶ [Complementary Colors](#)

Contrasting Colors

- ▶ [Complementary Colors](#)

Correlated Color Temperatures

- ▶ [Daylighting](#)

Cortical

- ▶ [Color Processing, Cortical](#)

Crawford, Brian Hewson

Robert W. G. Hunt
Department of Colour Science, University of
Leeds, Leeds, UK



1906–1991

Biography

Brian Hewson Crawford was a British physicist who made important contributions to the science of vision, colorimetry, lighting, and color rendering.

Brian Crawford was born in 1906 and died in 1991. His first scientific publication was, as his mother's ghostwriter, the Boy's Own Corner of the Daily Mail. He went to the University College London, graduating with first-class honors in Physics at the age of 19. After a brief spell at the

Rodenside Laboratory of the photographic company, Ilford, he joined the staff of the National Physical Laboratory (NPL) in 1927.

Having very wide interests, Crawford explored languages, painting, and music. His attitude to life was well illustrated by his remark after breaking his wrist in a fall from a bicycle that it had a reset at a more convenient angle for playing the viola. His honors included his Doctorate of Science from the University of London and his Newton Medal from the Colour Group (Great Britain).

Major Accomplishments/Contributions

Crawford worked in the NPL with W. S. Stiles under the leadership of John Walsh. A brilliant series of papers followed in the Proceedings of the Royal Society, on such new topics as equivalent backgrounds, increment thresholds, and, above all, the directional sensitivity of the retina to light and color known as the Stiles-Crawford effects. These were discovered during attempts to build a visual photometer based on pupillometry. The then “usual assumption that the apparent brightness of an object is proportional to the pupil area” was soon demolished [1–3].

Good fortune in research was characteristic of both Crawford and Stiles, but important discoveries fall only to those who deserve them. Crawford would say that you only had to search about in any field and you were bound to find out something interesting. No doubt this was so in his case. His 1947 Proc. Roy. Soc. paper on “Visual Adaptation in Relation to Brief Conditioning Stimuli” [4] is not obviously inspired by temporary blinding effects of gunfire flashes during the Second World War, but that was how “Crawford masking” was discovered. This is the effect whereby “a luminal stimulus begins to rise before the conditioning stimulus is applied to the eye.”

To compare and contrast the large-field colorimeters which Stiles and Crawford had each constructed in adjacent laboratories is an interesting exercise. Stiles’ machine was workshop built to the highest NPL precision and endowed with such facilities as a meteorology station of hygrometers, barometers, and thermometers, to

monitor changes in the refractive index of the air. Crawford’s apparatus belonged to the string and sealing wax tradition, with corks for nonslip knobs and strips of graph paper for scales. Both instruments served their respective purposes admirably.

The successful early partnership and later divergence of Crawford and Stiles are perhaps explained by the creative tension between opposite natures. To draw an analogy from art, Stiles’ science was classic; like Nicolas Poussin, he had neglected nothing. Crawford was a romantic; he was fond of quoting Maxwell’s dictum that it is always worth playing a trombone to a petunia at least once, and you never know what might happen. But Crawford could work to the highest NPL precision and accuracy when required, as he did in determining the average scotopic spectral response of the human eye, which forms the basis of the present CIE definition [5]. Not least was the difficulty of eliminating the effects of minute amounts of stray light and simultaneously discrediting data produced by several investigators who had not been so careful.

In his last years at NPL, Crawford studied the color rendering properties of artificial light sources, and he made a great breakthrough in convincing suspicious experts in art galleries and hospitals that certain fluorescent lamps were suitable for their exacting requirements [6]. Typical of Crawford was his finding that combinations of tungsten filament lamps and “radar blue” fluorescent lamps could imitate almost perfectly any phase of natural daylight. The resultant equipment was essential in providing a transportable reference illuminant for the darker corners of the Victoria and Albert Museum and the Sheffield Royal Infirmary. He also studied color matching and adaptation [7].

Work in the conservation department of the National Gallery was continued by Crawford long after officially retiring from NPL. He also made investigations on color in the laboratories of the University of Edinburgh, the Paint Research Association, Imperial College, and the Institute of Ophthalmology. He went on publishing original papers until the age of 79, thus refuting the fallacy that scientific research is only for young people [8].

References

1. Crawford, B.H.: The dependence of pupil size upon external light stimuli under static and variable conditions. *Proc. R. Soc. Med.* **B121**, 376 (1936)
2. Crawford, B.H.: The change of visual sensitivity with time. *Proc. R. Soc. Med.* **B123**, 69–89 (1937)
3. Crawford, B.H.: Photochemical laws and visual phenomena. *Proc. R. Soc. Med.* **B133**, 63 (1946)
4. Crawford, B.H.: Visual adaptation in relation to brief conditioning stimuli. *Proc. R. Soc. Med.* **B134**, 283 (1947)
5. Crawford, B.H.: The scotopic visibility function. *Proc. R. Soc. Med.* **B62**, 321 (1949)
6. Crawford, B.H.: Measurement of color rendering tolerances. *J. Opt. Soc. Am.* **49**, 1147 (1959)
7. Crawford, B.H.: Colour matching and adaptation. *Vision Res.* **5**, 71–78 (1965)
8. With the kind permission of the Colour Group (Great Britain), this account is based largely on the obituary written by Dr. D. A. Palmer and published in its newsletter. **17**, 1 (1992)

Cross-Cultural Color Preference

- [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)

Culture and Color

- [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)

D

Dalton, John

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



Biography and Major Accomplishments/Contributions

John Dalton was an English chemist, physicist, and meteorologist who is best known for his work

in the development of modern atomic theory and his research into color blindness.

Dalton was born in the county of Cumberland in England in 1766 and received his early education from his father who ran a private Quaker school. He started his own school aged 11, and at age 15 he joined his brother teaching in a school in Kendal [1]. He was barred from attending English universities because he was a dissenter (a Christian opposed to state religion and mandatory membership in the Church of England) but acquired scientific knowledge from John Gough, a blind philosopher with wide-ranging scientific interests. However, in 1793, he was appointed as teacher of mathematics at New College, Manchester (a dissenters' college). On 3 October 1794, Dalton was elected as a member of the Manchester Literary and Philosophical Society, and a month later he presented his first paper entitled, *Extraordinary Facts Relating to the Vision of Colors with Observations*. The society's *Memoirs and Proceedings* has been published continuously since its first edition which, when it was launched in 1783, was the only regular scientific journal in the United Kingdom except for the *Philosophical Transactions of the Royal Society*. As a result of this paper, congenital color vision deficiency is still referred to as "Daltonism" because he was the first to describe the condition in detail, and the paper also stimulated great debate among other investigators [1]. In 1800 he left New College and began private teaching, and in 1808 he published

the atomic theory for which he is most famous. He became President of the Manchester Literary and Philosophical Society in 1817, a post he retained until his death in 1844.

Dalton's Color Blindness

Dalton noticed that only males were afflicted with his condition, and we now know that color blindness, or color deficiency as it may more properly be referred to, is much more prevalent among males than females. He also recognized that the condition must be hereditary since he and his brother had the same condition.

In his 1794 letter, Dalton wrote:

The flower was pink but it appeared to me almost an exact sky-blue by day; in candlelight however it was astonishingly changed, not having then any blue in it but being what I call red – a colour which forms a striking contrast to blue.

Thirty-six years prior to Dalton's famous 1794 letter, Thomas Young had postulated that congenital color vision defects arose from the photoreceptors, but Dalton refused to believe it [1]. As a result of his meticulous observations, Dalton believed that the fluids in his eyes must contain a blue colorant, and he instructed that on his death his eyes be subject to a postmortem analysis. No trace of the blue colorant was found. Interestingly Dalton found 25 others (including, notably, his own brother) who suffered the same failure of color constancy. Dalton also wrote that the red end of the spectrum was "little more than a shade or defect of light."

We now know that Young was essentially correct, that is, that congenital color vision defects result from genetic mutations that mean that one (or more) of the three visual pigments found in the cones of our retina has anomalous spectral absorption properties or is absent entirely. Young believed that Dalton was a protanope – that is, that he was missing the long-wavelength sensitive visual pigment. Although the postmortem examination of Dalton's eye did not reveal the blue colorant that Dalton had predicted, the remaining eye was preserved

between two sheets of glass until it was analyzed by Mollen et al. who concluded that Dalton was indeed a dichromat but was a deuteranope rather than a protanope [2].

Dalton's theory that color deficiency was caused by a long-wavelength absorbing colorant in the fluids of his eyes ultimately proved to be incorrect. He was also not the first to "discover" color blindness. However, he made a substantial contribution to the understanding of color blindness through his meticulous observations, and he also inspired other scientists to study this fascinating phenomenon.

Atomic Theory

Although Dalton's name is indelibly associated with color deficiency, by far his most important work was in the area of atomic theory which he formally first published in 1805. The main points of Dalton's atomic theory were:

1. Elements are made of extremely small particles called atoms.
2. Atoms of a given element are identical in size, mass, and other properties; atoms of different elements differ in size, mass, and other properties.
3. Atoms cannot be subdivided, created, or destroyed.
4. Atoms of different elements combine in simple whole-number ratios to form chemical compounds.
5. In chemical reactions, atoms are combined, separated, or rearranged.

There is uncertainty about how Dalton arrived at his atomic theory [3]. Also, there was no evidence available at the time to scientists to deduce how many atoms of each element combine to form compound molecules so that Dalton, for example, wrongly assumed that water was composed of a single oxygen atom combined with a single hydrogen atom. Nevertheless, the essential features of his theory have survived to this day, and his theory can be said to be one of the bedrocks of modern chemistry.

Other Contributions

Dalton made notable contributions to meteorology (where he made over 200,000 observations, published several important essays on the gas laws (his law of partial pressures became known as Dalton's law), and wrote about rain, the color of the sky, and refraction).

References

1. Dickinson, C., Murray, I., Carden, D.: Dalton – letter. In: Dickinson, C., Murray, I., Carden, D. (eds.) *John Dalton's Colour Vision Legacy*. Taylor & Francis, London (1997)
2. Mollon, J.D., Dulai, K.S., Hunt, D.M.: Dalton's colour blindness: an essay in molecular biography. In: Mollon, J.D., Dulai, K.S., Hunt, D.M., Dickinson, C., Murray, I., Carden, D. (eds.) *John Dalton's Colour Vision Legacy*. Taylor & Francis, London (1997)
3. Nash, L.K.: The origin of Dalton's chemical atomic theory. *Isis* 47(2), 101–116 (1956). The History of Science Society CrossRef

Daltonism

- [Deuteranopia](#)
- [Protanopia](#)

Daylight

- [Daylighting](#)

Daylight Illuminants

Balázs Kránicz
University of Pannonia, Veszprém, Hungary

Definitions

Illuminants having the same or nearly the same relative spectral power distributions as phases of daylight.

Description of Daylight Phases

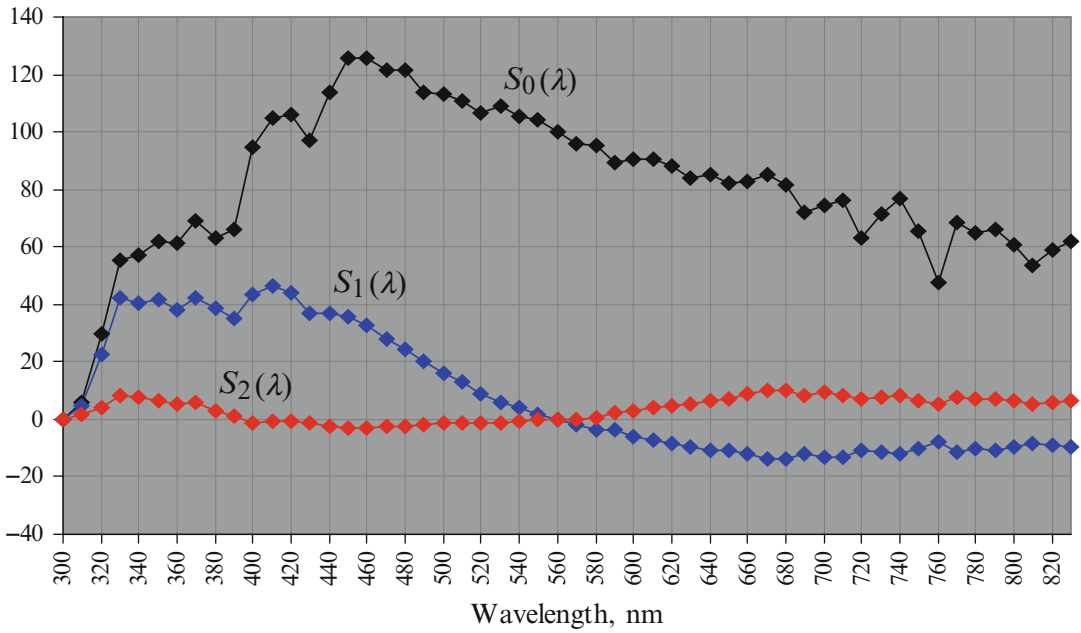
Originally, CIE standardized three illuminants for colorimetric purposes, illuminants A, B, and C. Illuminant C was the illuminant to represent daylight. CIE also defined standard sources for these illuminants, and CIE standard source C was realized by using an incandescent lamp and some liquid filters. This source had, however, less radiation in the ultraviolet spectrum as natural daylight. With the increased use of optical brighteners, it became necessary to define daylight that came nearer to natural phases of daylight.

In the 1960s, the question how daylight power distributions of different correlated color temperatures could be calculated by a simple and explicit formula came to the front. The task was to determine relative spectral power distributions of daylight that could be measured on the surface of the Earth and was modified by the scattering and filtering effect of the sky.

Pioneer importance can be assigned to the publication of Judd, Wyszecki, and MacAdam [1]. The authors and their coworkers performed spectroradiometric measurements, and with proper extrapolation of the wavelength set of a number of observations, a data base was completed from 300 to 830 nm with 10 nm steps.

From the measurement data, the average spectral power distribution (in references mostly denoted by $S_0(\lambda)$) and two characteristic distributions ($S_1(\lambda)$ and $S_2(\lambda)$) were determined; see Fig. 1. The authors themselves called the method applied *characteristic vector method*, and it can be considered as a kind of *principal component analysis*. Of course, more than two characteristic vectors had been determined, but it was obvious that for a given chromaticity coordinate pair, the two abovementioned characteristic distributions could be used. (In principal component analysis, they belong to the two biggest eigenvalues of the covariance matrix of the data.)

Based on the mean vector $S_0(\lambda)$ and the two most important characteristic vectors or distributions, daylight power distribution $S(\lambda)$ of correlated color temperature T_{cp} can be calculated as follows:



Daylight Illuminants, Fig. 1 The mean vector and the first two characteristic vectors to reconstitute spectral power distributions of daylight

$$S(\lambda) = S_0(\lambda) + M_1 \cdot S_1(\lambda) + M_2 \cdot S_2(\lambda). \quad (1)$$

Factors M_1 and M_2 in the formula above are not to be handled as constants; hence, these values depend on chromaticity coordinates x_D , y_D belonging to the given correlated color temperature T_{cp} :

$$M_1 = \frac{-1,3515 - 1,7703 \cdot x_D + 5,9114 \cdot y_D}{0,0241 + 0,2562 \cdot x_D - 0,7341 \cdot y_D} \quad (2)$$

and

$$M_2 = \frac{0,0300 - 31,4424 \cdot x_D + 30,0717 \cdot y_D}{0,0241 + 0,2562 \cdot x_D - 0,7341 \cdot y_D} \quad (3)$$

where in the case of $4,000 \leq T_{cp} < 7,000$

$$x_D = -4,6070 \cdot \frac{10^9}{T_{cp}^3} + 2,9678 \cdot \frac{10^6}{T_{cp}^2} + 0,09911 \cdot \frac{10^3}{T_{cp}} + 0,244063 \quad (4)$$

and in the case of $7,000 \leq T_{cp} \leq 25,000$

$$x_D = -2,0064 \cdot \frac{10^9}{T_{cp}^3} + 1,9018 \cdot \frac{10^6}{T_{cp}^2} + 0,24748 \cdot \frac{10^3}{T_{cp}} + 0,237040. \quad (5)$$

In both correlated color temperature ranges

$$y_D = -3,000 \cdot x_D^2 + 2,870 \cdot x_D - 0,275. \quad (6)$$

The unit of quantity of correlated color temperature T_{cp} in conditions of formula Eqs. 4 and 5 is kelvin (K). The form of Eqs. 4, 5, and 6 and the parameters in them come from curve fitting. The purpose of the curve fitting was to approximate the functional relation between correlated color temperature values and chromaticity coordinates of spectral power distributions of the measurement data set.

On the basis of the facts above, chromaticity coordinates, x_S and y_S , of $S(\lambda)$ determined by formula Eq. 1 have to match chromaticity

coordinates x_D and y_D (from Eqs. 4, 5, and 6) belonging to correlated color temperature T_{cp} .

Nominal and Actual Temperature

The CIE recommended daylight illuminants in 1967. The correlated color temperatures are affected by the numerical value of the radiation constant c_2 . In accordance with the International Practical Temperature Scale, 1948, amended in 1960, which was in use at the time when the procedure for calculating daylight illuminants was adopted by the CIE, the value of c_2 was equal to $1,438\ 0 \times 10^{-2}\ \text{m} \cdot \text{K}$. With this value, the correlated color temperature of illuminant D65 was exactly equal to 6,500 K. The change of c_2 to the value of $1,438\ 8 \times 10^{-2}\ \text{m} \cdot \text{K}$ (International Practical Temperature Scale, 1968) increases the correlated color temperatures of illuminant D65 by the factor 1,438 8/1,438 0. Thus, the correlated color temperature increases by approximately 4 K [2, 3].

A temperature value in the old scale is to be called *nominal temperature* (T_{nom}). Due to the change of constant c_2 , a temperature value belonging to some nominal temperature is called *actual temperature* (T_{act}) and can be calculated as

$$T_{\text{act}} = \frac{1,4388}{1,4380} \cdot T_{\text{nom}}. \quad (7)$$

Note that for the calculation of daylight illuminants or reproduction of definition tables of standards or CIE technical reports, the value of the actual correlated color temperature should generally be rounded to 3 decimal places after the decimal sign; see Table 1. If one wants to reproduce the spectra of daylight illuminants, the following hints should be applied besides that given for the rounding of the actual temperature: chromaticity coordinates at Eqs. 4, 5, and 6 have to be rounded to at least 6 decimal places, while factors M_1 and M_2 have to be rounded to exactly 3 decimal places – with the usual way of rounding.

Illustration of daylight illuminants belonging to different nominal correlated color temperature values is presented in Fig. 2.

Daylight Illuminants, Table 1 Nominal and actual correlated color temperature values of daylight illuminants

Daylight illuminant	Nominal correlated color temperature, K	Actual correlated color temperature, K
D50	5,000	5,002,782
D55	5,500	5,503,060
D65	6,500	6,503,616
D75	7,500	7,504,172

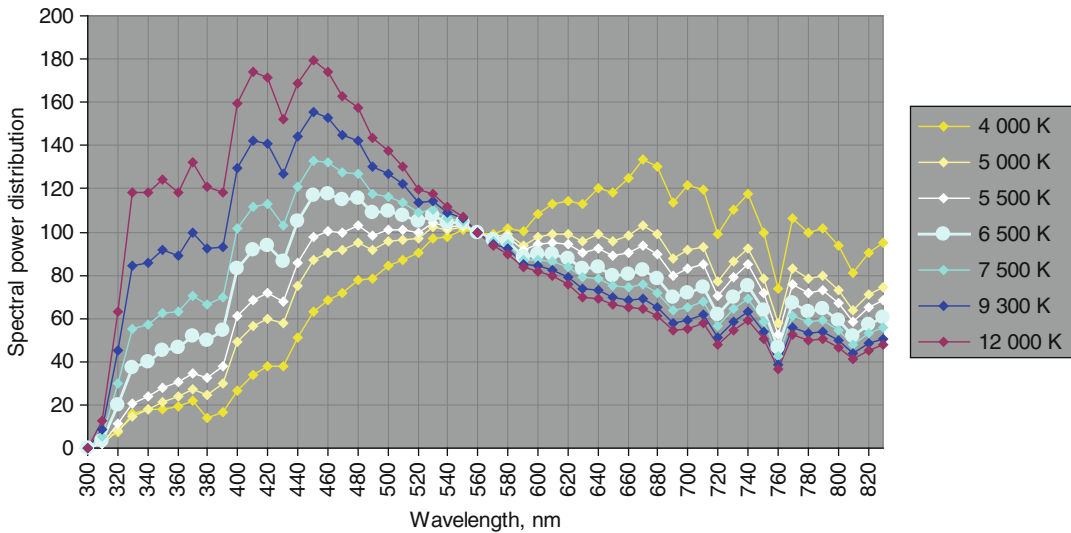
Interpolation of Daylight Illuminants and Its Consequences

In the 1960s, spectroradiometric methods applied in colorimetry often achieved spectral data with a step size of 10 nm. Daylight illuminants had been defined with the same wavelength step. Later, as spectroradiometric methods improved, more and more demands arose to have daylight illuminants with smaller wavelength steps. As the original measurements could not have been repeated, the only way was to apply interpolation. First, distributions $S_0(\lambda)$, $S_1(\lambda)$, and $S_2(\lambda)$ were *linearly* interpolated for a wavelength step of 5 nm [4]. CIE recently published a technical report that recommends nonlinear interpolation and gets to smooth spectral distribution curves [5].

In order to get the appropriate spectral power distributions and chromaticity coordinates, some recommendations contain certain rounding rules for computing M_1 and M_2 ; hence, using *other wavelength steps* than the original one of 10 nm, minor differences arise between chromaticity coordinates belonging to correlated color temperature values (x_D , y_D); see Eqs. 4, 5, and 6) and chromaticity coordinates (x_S , y_S) computed from daylight spectral power distributions themselves (from Eq. 1). These differences do not influence practical applications, but they are very embarrassing from a theoretical point of view. This anomaly had not been explained till 2000 [6].

CIE Standard Illuminant D65 (Shortly CIE D65)

As discussed in the previous section, the detailed spectrum of CIE standard illuminant D65 was



Daylight Illuminants, Fig. 2 Relative spectral power distributions of daylight illuminants belonging to different nominal correlated color temperature values

calculated from the fixed 10 nm values by linear interpolation. Data can be found, e.g., in the book by Wyszecki and Stiles [7] and in the standard for CIE standard illuminant A and D65 [8]. The tabular data were rounded in such a way that 6 significant decimal places were kept, together with the integer part and the digits of the decimal fraction. It should be noted that an unusual rounding rule had been applied: if a number ended with the digit 5, it was rounded downwards, e.g., at 405 nm the piece of data 87,12045 was rounded to 87,1204.

CIE D65 is one of the most important and most frequently used standard illuminants of the CIE. Many spectrophotometers indicate color values of surfaces in such a way as if they had been illuminated by CIE D65; for further details, see Chapter CIE Standard Illuminants and Sources.

CIE D50, D55, and D75

Although CIE suggests using D65 whenever it is possible, three further illuminants are often used. These illuminants are D50, D55, and D75. D50 is often used in graphic arts; D55 and D75 correspond to phases of daylight when the Sun starts

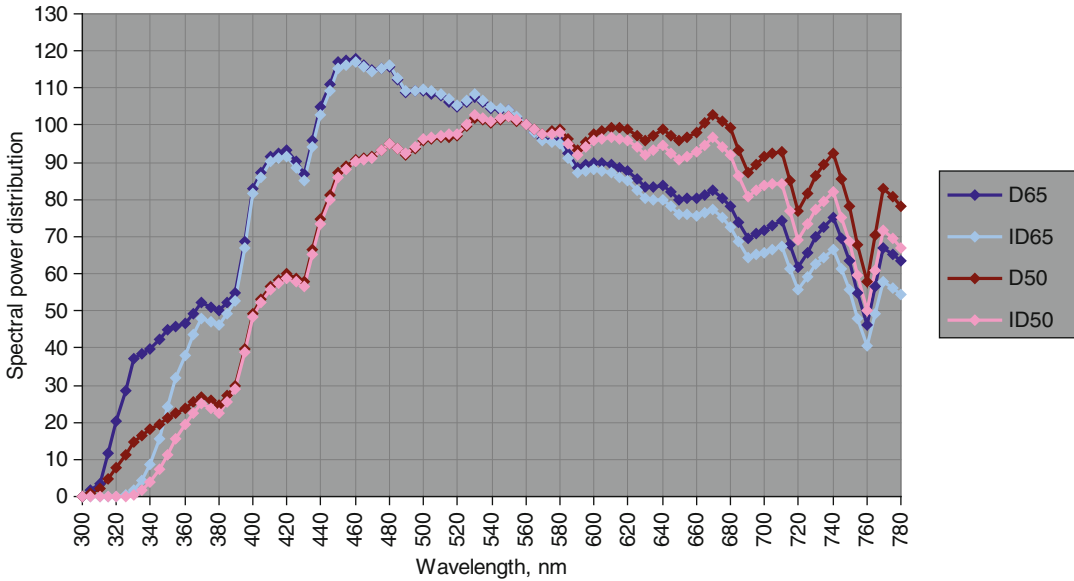
setting and when the blue sky dominates, respectively.

The calculation of D50, D55, and D75 can be performed similarly to that of D65. The definition table in publication CIE 15:2004 [5] contains the wavelength range 300–780 nm with a step size of 5 nm. It should be noted that in this table all numerical data were rounded to 3 decimal places with the usual rounding rule, independently from the number of digits of the integer part of the values.

Recalculation of Daylight Illuminants

The original CIE rule to calculate the spectra of daylight illuminants is that first the spectrum for the 10 nm steps has to be determined and this spectrum has to be linearly interpolated. If one tries to interpolate first the S_0 , S_1 , and S_2 functions, one will get slightly different values.

It turned out that no rounding rules would give a perfect solution for eliminating the problem of differences between chromaticity coordinates calculated by Eqs. 4, 5, and 6 and from the illuminant itself (Eq. 1). It was found that factors M_1 and M_2 in Eq. 1 depend on the wavelength step of distributions or characteristic vectors S_0 , S_1 , and S_2 .



Daylight Illuminants, Fig. 3 Spectral power distributions of D50, D65, ID50, and ID65

Factors M_1 and M_2 were originally determined for the case of a wavelength step of 10 nm. If any type of interpolation of daylight illuminants is applied, this fact leads to computational inaccuracies and theoretical contradictions. Consequently, factors M_1 and M_2 in Eq. 1 have to be *recalculated* in every case when the wavelength step of distributions S_0 , S_1 , and S_2 differs from 10 nm.

A mathematical procedure was worked out by which factors M_1 and M_2 can precisely be recalculated for any case of interpolation of S_0 , S_1 , and S_2 [6]. Based on the recalculation chromaticity coordinates (x_D , y_D) depending on the correlated color temperature of daylight, distributions always match those, (x_S , y_S), calculated from distribution S . CIE publication 204 contains Excel sheets to perform calculations for any correlated color temperature [5].

Indoor Daylight Illuminants

All daylight illuminants dealt with above have power also in the UV range that cannot be neglected in the case of fluorescent samples. However, natural situations like applying daylight

filtered by a window sheet raised a demand for having some indoor versions of daylight illuminants. CIE's technical committee TC 1-66 investigated several plate glass types. Based on their spectral transmittance properties and average thickness, two indoor daylight illuminants were defined: ID50 and ID65 [10]. Figure 3 illustrates D50, D65, ID50, and ID65.

Cross-References

► [CIE Standard Illuminants and Sources](#)

References

1. Judd, D.B., Wyszecki, G., MacAdam, D.L.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1040 (1964)
2. Wyszecki, G., Stiles, W. S.: CIE standard illuminants. In: Wyszecki, G., Stiles, W. S. (auts.) *Color Science*, pp. 143–146. Wiley, New York (1982)
3. Schanda, J.: CIE colorimetry. In: Schanda, J. (ed.) *Colorimetry – Understanding the CIE System*, pp. 42–43. Wiley Interscience, Hoboken (2007)
4. Commission Internationale de l'Éclairage, CIE: *Colorimetry*, 3rd edn, pp. 33–34 CIE **15** CIE Central Bureau, Vienna (2004), Chapter 3.1

5. Commission Internationale de l'Éclairage, CIE: Methods for re-defining CIE D illuminants. CIE **204** CIE Central Bureau, Vienna (2013)
6. Kránicz, B., Schanda, J.: Re-evaluation of daylight spectral distributions. *Color. Res. Appl.* **25**, 250–259 (2000)
7. Wyszecki, G., Stiles, W.S.: Table I(3.3.4). In: Wyszecki, G., Stiles, W.S. (eds.) *Color Science*, pp. 754–758. Wiley, New York (1982)
8. Commission Internationale de l'Éclairage, CIE, Joint ISO/CIE Standard: Colorimetry – Part 2: CIE Standard Illuminants for Colorimetry. CIE S 014-2/E:2006/ISO 11664-2:2007(E) CIE Central Bureau, Vienna
9. Commission Internationale de l'Éclairage, CIE: Colorimetry, 3rd edn, CIE **15** CIE Central Bureau, Vienna (2004), Table T.1, pp. 30–32
10. Commission Internationale de l'Éclairage, CIE: Indoor Daylight Illuminants, 1st edn. CIE **184** CIE Central Bureau, Vienna (2009)

Daylight Spectra

► Daylighting

Daylighting

John Mardaljevic
School of Civil and Building Engineering,
Loughborough University, Loughborough,
Leicestershire, UK

Synonyms

[Correlated color temperatures](#); [Daylight](#); [Daylight spectra](#); [Standard daylight illuminants](#)

Definition

Daylight is the totality of visible radiation originating from the sky and, when visible, the sun during the hours of daytime.

Introduction

The colors, or more specifically, the spectra of the various phases of daylight are closely related to

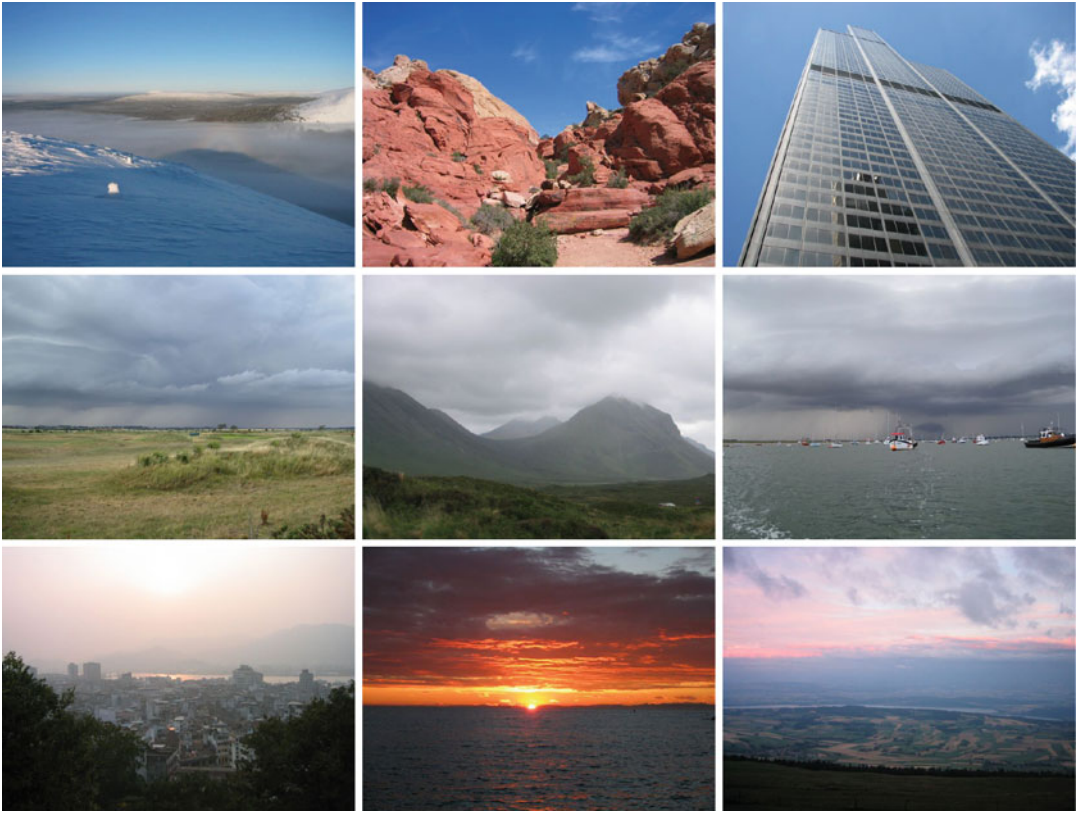
the commonplace experience of naturally occurring sun and sky conditions, e.g., clear blue skies, white clouds, red sunsets, etc. Prevailing conditions can be characterized in terms of a correlated color temperature. For certain typical sun and sky conditions, there are standardized spectral power distributions. From these it is possible to estimate the time-varying spectra of daylight from illuminance values in standardized climate data files. Thus, it becomes possible to predict the daylight spectra experienced by building occupants. This has found application in studies on the nonvisual effects of daylight and the performance of dynamic glazing systems.

The qualitative aspects of the colors, or hues, of daylight are difficult to describe, harder still to codify in any but a superficial manner. Intricately bound with these perceptions are the conditions under which they occur:

- Prevailing degrees of sunniness and cloudiness
- The position of the sun in the sky (when visible)
- Clearness or haziness of the skies
- Occurrence of various cloud types
- Properties of the natural and man-made reflecting environment

Each of this can have some notion of colors associated with it. How daylight can appear to humans found expression in the world of art where certain qualities of illumination became associated with specific locales. For example, the subtle hues in the gently luminescent skies (often reflected over water) painted by the Dutch masters of the seventeenth century. Or the bright, vivid shades used by Cezanne and fellow painters who sought inspiration in the sun-drenched south of France. Commonplace holiday photographs can also reveal much of the variety in experienced daylight conditions, both in natural and man-made settings, Fig. 1.

Notwithstanding the evident variety in the hues of the experienced luminous environment, consideration of the color of daylight has not generally played a role in the daylighting design of buildings – at least not until very recently. In purely aesthetic terms, the color of the daylight



Daylighting, Fig. 1 The “colors” of daylight – light from the sun and/or sky together with that reflected by the environment

is easily noticed, e.g., the warm light from a low sun, the cool light in rooms that are sunless, etc. As a consequence, rooms with a southerly aspect may, in a general rather than just a lighting sense, be perceived as “warm” and those with a northerly aspect as “cool.” This is particularly so in higher-latitude locations, e.g., the Scandinavian countries where the sun is often low in the sky [1]. Designers may even try to compensate somewhat for this, at least for the rooms with a northerly aspect, by selecting paint finishes/decor with warm hues to neutralize somewhat the perception of coolness in the space. There is however little in the way of an empirically based approach to systematically apply such considerations to the color design of spaces.

While it is possible to describe the various types of daylight that occur using chromaticity coordinates [2], it is often simpler and indeed

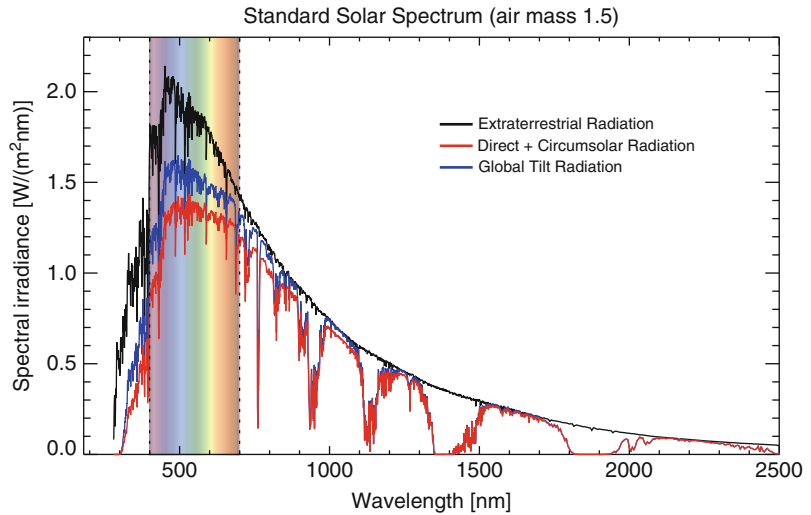
more revealing to consider the spectral power distribution (SPD) of the light rather than reducing it to derived quantities which could hide important details. Recent advances in variable transmission glass technology [3] and in understanding the basis of the nonvisual effects of daylight [4] further suggest that daylight SPDs rather than chromaticity coordinates are needed to progress in these areas. Thus, this chapter will describe daylight and its interactions – reflection and transmission – using spectra rather than any of the color systems. It begins with a consideration of the spectrum of solar radiation arriving at the top of the Earth’s atmosphere.

The Standard Solar Spectrum

The visible radiation is only a part of the total energy associated with daylight radiation. The full spectrum of radiation ranges from about

Daylighting,

Fig. 2 Standardized extraterrestrial radiation with direct and global radiation received at the surface of the Earth (air mass = 1.5)



300 to 4,000 nm, i.e., from the near ultraviolet to the far infrared (both invisible to the eye). The radiation from the sun arriving at the Earth, but before entering the atmosphere, is called extraterrestrial radiation. It approximates a black body radiator with a temperature of 5,800 K, i.e., a large “tail” at long infrared wavelengths gradually diminishing to zero, the peak of output in the visible spectrum, and a steep cutoff at ultraviolet wavelengths, Fig. 2. How that energy arrives at the surface of the Earth depends on various factors including atmospheric.

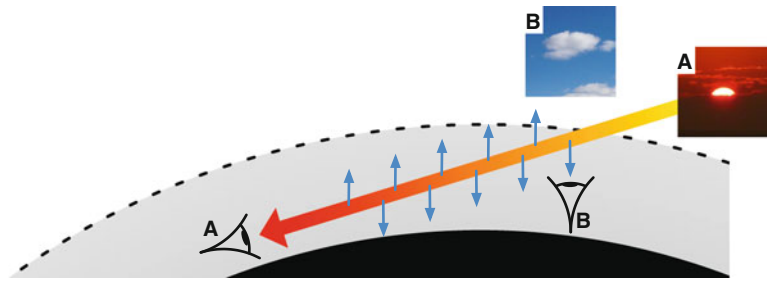
Scattering of sunlight in the atmosphere by air, water vapor, dust, etc. gives the sky the appearance of a self-luminous hemispherical source of light. Sunlight is commonly referred to as direct light since it appears to originate from a small source and can cast sharp shadows. The sky however is an extended source of illumination that casts only soft shadows, and so skylight is commonly referred to as diffuse light. Scattering of solar radiation in the atmosphere does not occur equally across the visible spectrum – shorter, i.e., “blue,” wavelength light is scattered more than the other colors. Thus, on clear days, the sky appears blue – the observer is seeing light (originating from the sun) that arrives via indirect scattered paths. For the same reason, the sun appears yellow because blue light has been scattered out of the beam, i.e., the direct path from the sun to the observer. The lower the angle that the sun appears

in the sky, the greater the distance in the atmosphere through which the sunlight travels. Thus, more of the blue (and green) wavelengths are scattered out of the beam, and the sun appears progressively yellow, orange, and then red as the sun sets. A blue sky and a red sunset are in fact the same optical phenomenon – preferential scattering of short wavelength light in the atmosphere – seen from two different viewpoints, Fig. 3. The air mass for a horizontal view (at sea level) is taken to be 38. Thus, the beam of light from the sun at the horizon has traveled through approximately 25 times more atmosphere than when seen at an altitude of 41.8° (i.e., the condition for AM1.5).

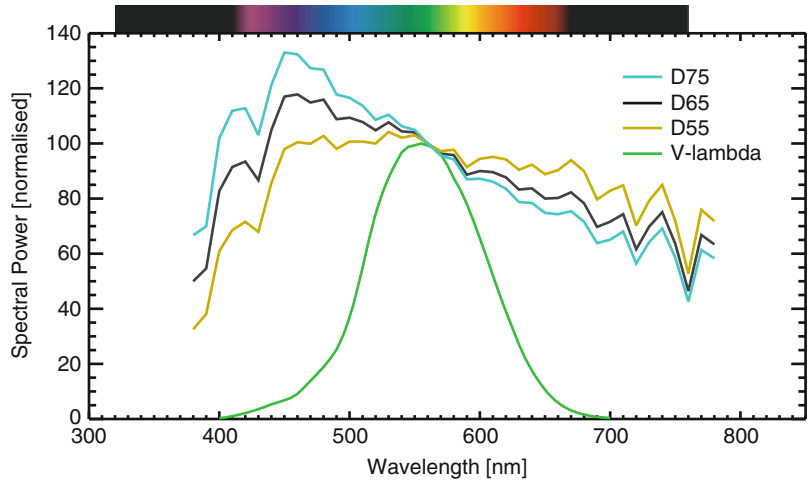
The Standard Daylight Illuminants

The standard solar spectra were largely created for the simulation and testing of energy systems, e.g., photovoltaic panels. Similarly, the standard daylight (or D) illuminants were formulated by the International Commission for Illumination (CIE) to represent various “phases of daylight” [5]. Three of the most commonly used standard D illuminants are D55, D65, and D75. The normalized spectral power distribution (SPD) for each of this is shown in Fig. 4. Also plotted is the visual sensitivity curve $V(\lambda)$ – each curve has been normalized to have the same value at 555 nm. The illuminants are characterized by their nominal correlated color temperature

Daylighting, Fig. 3 A blue sky and a red sunset – two different views of the same phenomenon



Daylighting, Fig. 4 Standardized daylight illuminants and the visual sensitivity curve



(CCT), i.e., the temperature of the Planckian or blackbody radiator whose perceived color most closely resembles that of a given stimulus at the same brightness and under specified viewing conditions. The nominal correlated color temperatures for D55, D65, and D75 are, respectively, 5,500 K, 6,500 K, and 7,500 K. The illuminant SPD curves also exhibit the absorption features present in the “direct” and “global” spectral irradiance curves (Fig. 2), though they are now easier to see with the smaller range in plotted wavelength.

In relative terms compared to D65, illuminant D55 will appear “warmer” and D75 “cooler.” Note that higher correlated color temperatures (CCTs) result in SPDs that are perceived as progressively “cooler,” that is, “bluer” even though the (equivalent) blackbody radiator is of course hotter. Of these three, illuminant D65 is taken as the reference source for daylight, corresponding roughly to the illumination from midday sun and (blue) skylight in temperate European locations:

CIE standard illuminant D65 is intended to represent average daylight and has a correlated colour temperature of approximately 6,500 K. [6]

Daylight and Correlated Color Temperature
Data relating natural illumination conditions to correlated color temperatures is sparse. One of the most comprehensive sources of data is the Kodak Cinematographer’s Field Guide [7]. It gives equivalent correlated color temperatures for 11 daylight conditions, Table 1. Here the daylight condition for the nominal CCT of 6,500 K (i.e., D65) is described as “average summer sunlight (plus blue skylight),” essentially equivalent to the arguably less precise CIE description of the daylight conditions.

Note that the CCT for sunlight varies from 2,000 (sunrise/sunset) to 5,800 K, with “average sunlight at noon” having a CCT of 5,400 K. Thus, illuminant D55 can be assumed to represent the SPD for (direct) sunlight provided that the sun is not too low in the sky. In reality, average summer

Daylighting, Table 1 Correlated color temperatures for various daylight sources (Source Kodak)

Daylight source	CCT
Sunlight – sunrise or sunset	2,000 K
Sunlight – 1 h after sunrise	3,500 K
Sunlight – early morning	4,300 K
Sunlight – late afternoon	4,300 K
Average summer sunlight at noon (Washington, DC)	5,400 K
Direct midsummer sunlight	5,800 K
Overcast sky	6,000 K
Average summer sunlight (plus blue skylight)	6,500 K
Light summer shade	7,100 K
Average summer shade	8,000 K
Summer skylight	9,500–30,000 K

sunlight plus blue skylight will in fact be a combination of SPDs – *D55* for sunlight and some (currently unspecified) SPD for blue skylight with a CCT of up to 30,000 K. The reference illuminant *D65* is therefore a compromise offering an SPD that, for most practical purposes, is equivalent to some combination of *D55* (sunlight) and some unspecified high CCT for blue skylight. The reference daylight illuminant CCT of 6,500 K is closer to the CCT for average sunlight (5,500 K) than the $\geq 9,500$ K for blue skylight because, in most situations, illumination from the sun (direct and reflected) will dominate that from the sky.

Kodak recommends that an overcast sky has an SPD equivalent to a CCT of 6,000 K. In which case, it would appear that the closest match of the three D illuminants is either *D55* or *D65*. Given that *D55* is believed to be a good match for sunlight, the SPD for an overcast sky is better represented by *D65*. That has been confirmed by measurements taken on a wide range of commonly occurring overcast sky conditions [8].

Daylight Quantities in the Standard Climate Datasets

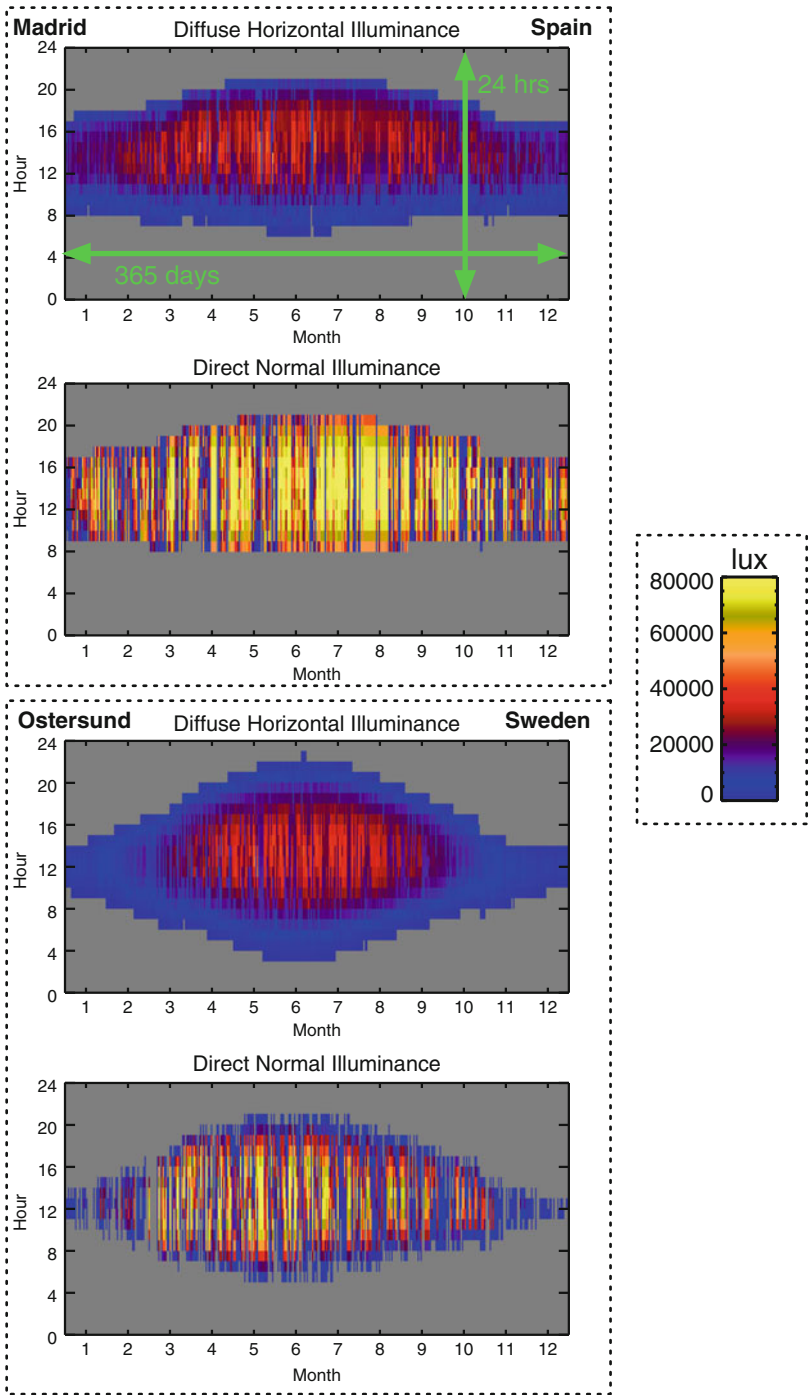
The quantitative nature of illumination from the sun and sky for any particular locale can be appreciated by examining the luminous output from both the sun and the sky over a period of a full year. A full year is needed to capture seasonal

variation in daylight. The principal sources of annual data for daylight are the standard weather files which were originally created for use by dynamic thermal modeling programs [9]. These datasets contain averaged hourly values for a full year, that is, 8,760 values for each parameter. The key daylight parameters stored in the weather files are the diffuse horizontal illuminance and the direct normal illuminance. The diffuse horizontal illuminance is the visible light energy from the unobstructed sky that is incident on a horizontal surface. The direct normal illuminance is the visible light energy from the sun that is incident on a surface which is kept normal to the beam of radiation, i.e., the photocell always “points” directly to the sun. The illuminating power of the sun is measured in this way because when the sun is low its contribution to horizontal illuminance is small even if the absolute intensity is high.

A visualization of the illuminance data from the standard weather files for Madrid (Spain) and Ostersund (Sweden) is given in Fig. 5. The time-series data of 8,760 values has been rearranged into an array of 365 days (x-axis) by 24 h (y-axis). The shading at each hour indicates the magnitude of the illuminance – see legend – with zero values shaded gray. Presented in this way, it is easy to appreciate both prevailing patterns in either quantity or their short-term variability. Most obvious is the daily/seasonal pattern for both illuminances: short periods of daylight in the winter months and longer in summer. The hour-by-hour variation in the direct normal illuminance is clearly visible, though it is also present to a lesser degree in the diffuse horizontal illuminance (i.e., light from the sky).

Comparing the data for the two locales, one immediately notices the pronounced seasonal variation in day length for Ostersund, which is in the north of Sweden. As might be expected, the Madrid data show many sunny days, especially in summer. Note that Spain uses Central European Time despite having a longitude similar to the UK. Thus, there is a marked divergence between clock and solar time in the Madrid data. The diffuse horizontal illuminance reaches a maximum of about 40,000 lx (red shading), and direct normal illuminance under clear sky conditions

Daylighting,
Fig. 5 Illuminance data
from standard climate files
for Madrid (Spain) and
Ostersund (Sweden)



often exceeds 70,000 lx (yellow shading). This is so for both locales. During winter in northern Sweden, the diffuse horizontal illuminance is typically 5,000 lx or even less, with few hours of sun.

Both diffuse and direct illuminances will, in reality, vary over periods much shorter than an hour. Interpolation of the dataset to a time-step shorter than 1 h will provide a smoother traversal

of the sun, which may be necessary when using the data for simulation of daylight. Interpolation alone however will not introduce short-term variability into the values for diffuse horizontal and direct normal illuminance. The patterns of hourly values in the illuminance datasets are unique, and because of the random nature of weather, they will never be repeated in precisely the way shown in Fig. 5. Climate datasets are however representative of the prevailing conditions measured at the site, and they do exhibit much of the full range in variation that typically occurs, i.e., they provide definitive yardstick quantities for modeling purposes. These datasets are used to define the external conditions for climate-based daylight modeling (CBDM), i.e., the prediction of annual profiles of internal (daylight) illuminance due to all the unique sun and sky conditions derived from a standard climate file [10].

For every value in the datasets for diffuse horizontal and direct normal illuminance, it is possible to estimate, with varying degrees of reliability, a corresponding CCT using the entries in Table 1. For example, much of the intense sunlight (i.e., direct normal illuminance) could be taken to have a CCT of around 5,400–5,800 K. Thus, the SPD for the sunlight could be approximated to standard illuminant D55. Sunlight around the hours of dawn and dusk could be assigned lower CCTs, e.g., 2,000–3,500 K, while early-morning and late-afternoon sun could be approximated to a source with a CCT of 4,300 K. Use of climate files to estimate the SPD of the sun and skylight sources of illumination has recently been demonstrated for the purpose of modeling the nonvisual effects of daylight (see section “[The Nonvisual Effects of Daylight](#)”).

Sky Models

A sky model is a mathematical equation that describes the variation in luminance across the hemispherical sky vault. The simplest possible model describes a sky of uniform luminance, i.e., the same brightness is seen wherever one looks above the horizon (in an unobstructed setting). At the start of the twentieth century, it was believed that overcast skies exhibited only moderate variation in brightness across the sky dome,

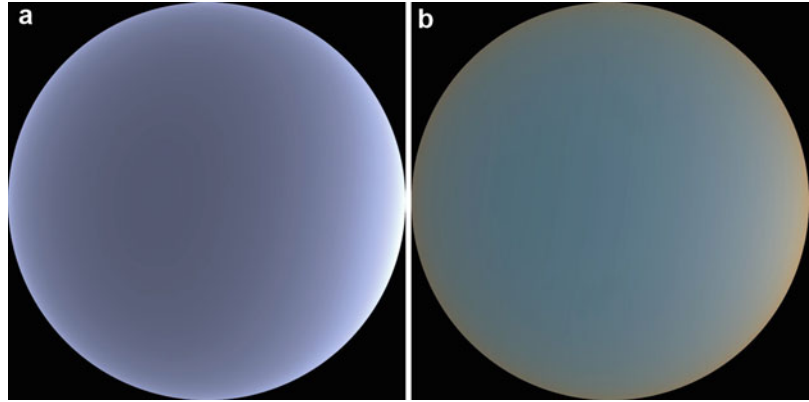
and so they could be considered to be of constant (i.e., uniform) luminance. Measurements revealed however that a densely overcast sky exhibits a relative gradation from darker horizon to brighter zenith; this was recorded in 1901. With improved, more sensitive measuring apparatus, it was shown that the zenith luminance is often three times greater than the horizon luminance for some of the most heavily overcast skies [11]. The formulation for the luminance pattern of overcast skies presented by Moon and Spencer in 1942 was adopted as a standard by the International Commission for Illumination (CIE) in 1955.

The equations describing the luminance pattern for clear or intermediate skies are rather more complex than that for the CIE standard overcast sky. The complexity arises from a number of observed effects that are accounted for in the model. Among these are a bright circumsolar region, a sky luminance minimum that is at some point above the horizon, and a brightening of the sky near the horizon. The scales of these effects are related to the solar position and the relative magnitudes of the illumination produced by the sun and sky [12]. These and similar sky models, e.g., the Perez all-weather model [13], do not include provision in the formulation for any variation in chromaticity of the luminance across the sky vault. If the sky color is required in a daylight simulation, the RGB radiance values applied to the self-luminous “material” used to model the sky can be set to achieve the desired effect, e.g., a small blue “excess” to create a blue hue for the clear sky distribution, Fig. 6a. Notwithstanding the variation in luminance across the sky vault, the chromaticity will remain constant.

More recently, sky models with varying chromaticity have been formulated to depict the subtle color gradations seen in nature [14]. Based on analytical models for the scattering of light in the atmosphere, these formulations can account for effects such as blue sky at high altitude together with warm (i.e., blue deficient) sky just above the horizon, Fig. 6b. It should be noted that data to validate these models is scarce, and, for daylight simulation, their use is primarily to lend realism to renderings rather than for any quantitative modeling, e.g., illuminance prediction.

Daylighting,

Fig. 6 Simulated 180° hemispherical “fish-eye” images of a constant chromaticity clear sky distribution (a) and a variable chromaticity clear sky (b)

**Daylight in Buildings**

Daylight illumination of a space can be considered as a series of transforms – from a light emitted by the luminous source(s) to that received at a point in a space. For daylight performance evaluation, practitioners have largely been concerned with the amount of light received at a point in a space (i.e., lux levels) and have not been concerned with the spectral nature of any of the transforms. For most quantitative modeling purposes – using either the long-standing daylight factor method or more recent climate-based daylight modeling – it is sufficient to describe all opaque surfaces as gray reflectors. And, although tinted glazing materials may be represented in the simulation by their RGB spectral transmittance values, an equivalent gray transmittance would work just as well.

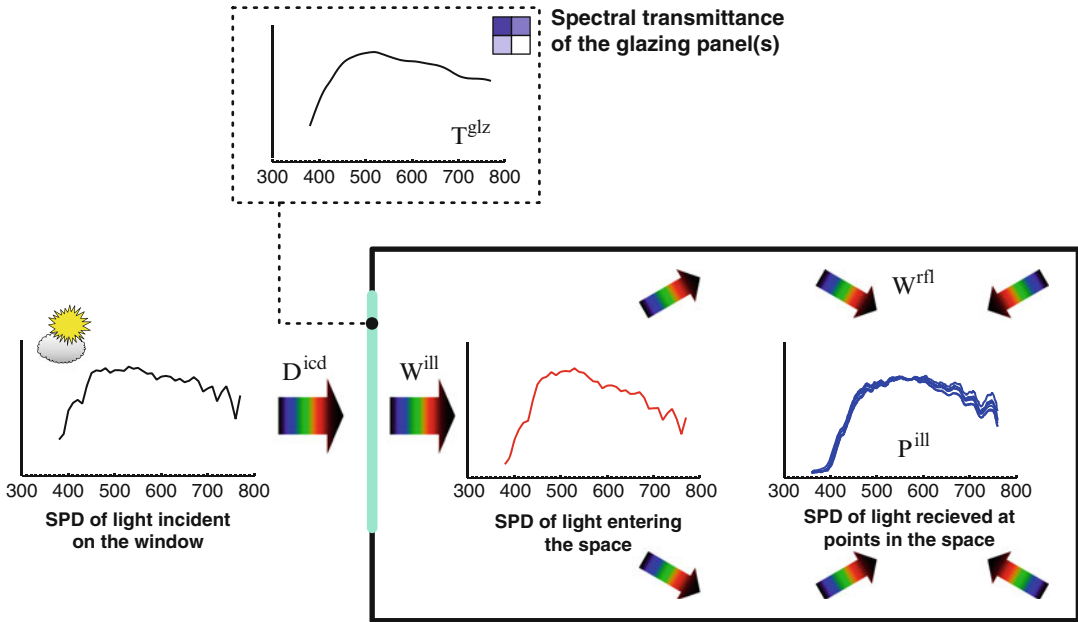
Color can, of course, be modeled in a lighting simulation. Though, depending on the tool used, there may be limitations or restrictions to the degree of precision that can be achieved when modeling the spectral properties of reflection and transmission. Key among these are the limitations of having three channels, as is the case with radiance – generally considered the most widely used and rigorously validated lighting simulation package [15]. The “color” assigned to each of the three channel is arbitrary, though in most instances they are used to model the standard RGB trichromatic system common in computer graphics and digital image display technology. While adequate for many if not most image

display purposes, the three channel RGB approaches are in fact a fairly coarse representation of the reflection and transmission properties of materials. For example, it is possible to have materials with (noticeably) different spectral reflectance curves giving the same RGB values [16]. Thus, although noticeably different to an observer, they would all have the same representation in the simulation, i.e., the same RGB spectral reflectance values.

Transformation of Daylight by Reflection and Transmission

Considering only the normalized spectra (i.e., ignoring absolute values), one can represent the transformations fairly simply as follows. Consider a space with a window receiving direct sun on sunny day with a clear blue sky. The SPDs for the sunlight and the skylight incident on the window are denoted, respectively, by the vectors \mathbf{D}^{sun} and \mathbf{D}^{sky} . The SPD for the light received at the window will be comprised of these two illuminants plus sunlight and skylight reflected off the external surfaces, e.g., the ground and nearby buildings. The vector for the reflected SPD incident on the window is denoted by \mathbf{D}^{ref} . The reflected SPD will be some combination of the SPDs from the sky and (if present) the sun, modified by the spectral reflectance properties of the ground, external buildings etc. The SPD for the light incident at the window \mathbf{D}^{icd} is given by:

$$\mathbf{D}^{\text{icd}} = \mathbf{D}^{\text{sun}} + \mathbf{D}^{\text{sky}} + \mathbf{D}^{\text{refl}} \quad (1)$$



Daylighting, Fig. 7 Transformations in the SPD from source(s) to a point inside a space

If, say, the window is facing the sun direction, then of the total light incident on the glass that arriving directly from the sun will probably form the greatest part. And so the incident SPD could be approximated to the D55 illuminant. As noted previously, only the shape of the spectra is important for this illustration.

The spectrum of light incident on the window will, on passing through it, be modified by the spectral transmittance T^{glz} of the glazing material. The element-by-element (or Hadamard) product of the incident SPD and spectral transmittance vectors give the SPD for the light entering the space:

$$\mathbf{W}^{ill} = \mathbf{D}^{icd} \circ \mathbf{T}^{glz} \quad (2)$$

The vector \mathbf{W}^{ill} may be called the window illuminant SPD since this is the daylight from the window illuminating the space. This light can arrive at a point in the space directly (if the point can “see” the window) and indirectly following one or more reflections. The SPD of the window illuminant will in turn be transformed following every possible reflection in the space depending on the

spectral reflectance properties of the various surfaces. Thus, for every point in the space (x, y, z, θ, ϕ) , the light falling on it \mathbf{P}^{ill} will have a unique SPD comprised of direct and reflected SPDs:

$$\mathbf{P}^{ill}(x, y, z, \theta, \phi) = \mathbf{W}^{ill}(x, y, z, \theta, \phi) + \mathbf{W}^{rfl}(x, y, z, \theta, \phi) \quad (3)$$

In addition to the (x, y, z) position in the space, the orientation of the receiving surface (θ, ϕ) needs to be defined also – the “surface” could be the eye of an observer. If the surface cannot “see” the window directly, then the $\mathbf{W}^{ill}(x, y, z, \theta, \phi)$ term will be zero. This process of spectral transformations is illustrated in Fig. 7.

Practical Simplifications

Under many actually occurring conditions, the SPD of (visible) radiation incident on the windows of a building in a man-made setting can be assumed to be dominated by light from the sky and sun (if present). Thus, one can neglect the transform in the SPD from the reflected component because (a) the ground reflectivity is

invariably low and (b) man-made external surfaces in any case tend to be fairly neutral in color. Under these conditions, it is reasonable to assume that the SPD incident on the windows will be approximately that of the dominant source. For example, if direct sun is incident on the window of a building, we can assume that the incident SPD is that of the standard illuminant D55, i.e., $\mathbf{D}^{\text{led}} = \mathbf{D}^{55}$.

In an actual building with typical neutral decor, the SPD is likely to vary only marginally with position in the space. In other words, inside the space the window illuminant SPD is not transformed significantly on reflection. One can make the approximation that $\mathbf{W}^{\text{refl}} \simeq \mathbf{W}^{\text{ill}}$ since the variation with position (x, y, z, θ, ϕ) in the space is not a significant factor. Thus, one can assume that the (normalized) SPD at any point in the space is approximately equal to the SPD of the light entering the space through the window, i.e., $\mathbf{P}^{\text{ill}} \simeq \mathbf{W}^{\text{ill}}$.

Combining these two simplifications, the SPD of light in the space (\mathbf{P}^{ill}) is now governed by just two factors: the SPD of the incident light and the

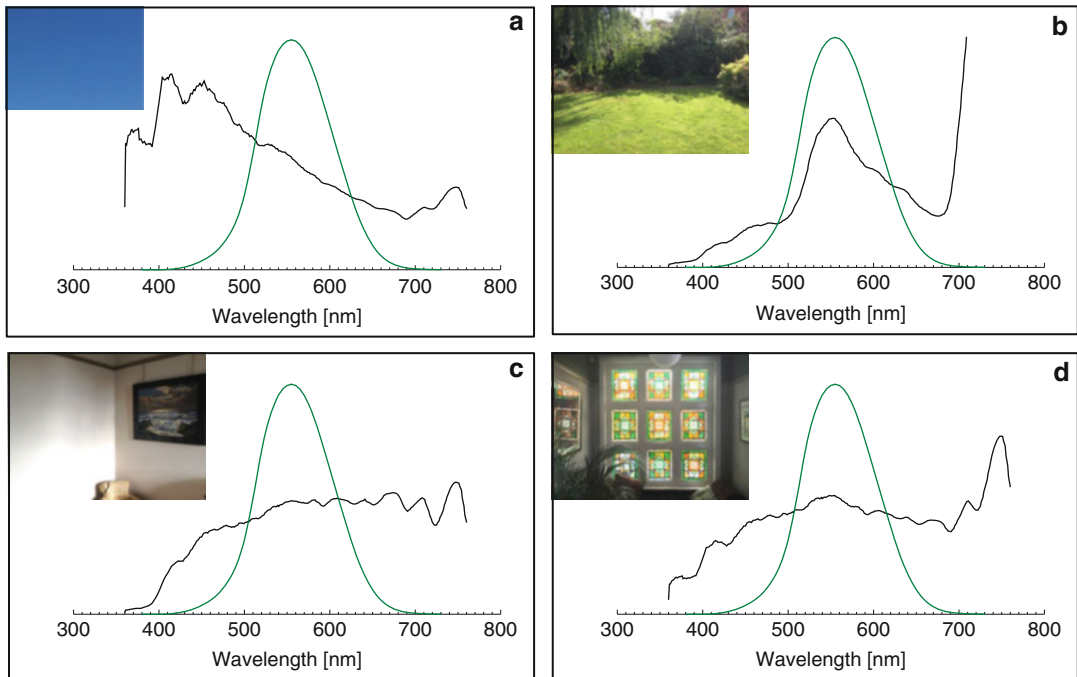
spectral transmittance of the glazing. Using the above example with direct sun incident on the window, the equation for the SPD in the space simplifies to:

$$\mathbf{P}^{\text{ill}} \simeq \mathbf{D}^{55} \circ \mathbf{T}^{\text{glz}} \quad (4)$$

This approximation has been tested and shown to be perfectly adequate for the purpose of predicting the SPD in a space (see section “[Variable Transmission Glazing](#)”).

Measured Daylight Spectra

It is now possible to take reliable measurements of SPDs using relatively inexpensive, handheld spectrometers. These devices measure spectral irradiance, which is approximately that seen by the eye when “pointing” in the same direction. Four measured daylight SPDs are shown in Fig. 8. The inset photograph shows approximately what was “seen” by the sensor, and the visual



Daylighting, Fig. 8 Measured daylight spectra: (a) blue sky, (b) sunlit garden, (c) back of a room with warm/neutral decor and clear glazing on a bright day, and, (d) daylight from a stained glass window

sensitivity curve $V(\lambda)$ is also shown. The SPDs were normalized against $V(\lambda)$ to have the same area in each case, i.e., to have the same lux value irrespective of the shape of the spectra. This allows a like-for-like comparison of the SPDs whatever their absolute values are.

Two Outdoor Daylight Spectra

The SPD for light from the bluest part of a clear sky shows pronounced peaks around 400 nm, Fig. 8a. Comparison with the standard illuminants (Fig. 4) indicates some similarity on the basic form; however, this light would have a CCT significantly higher than the 7,500 K for *D75* (the spectrometer reported a CCT of over 12,000 K for this SPD). The SPD for the sunlit garden shows a strong peak in the green part of the spectrum due to reflection from the foliage and a steep rise at around 700 nm at the start of the infrared region, Fig. 8b.

Two Indoor Daylight Spectra

The SPD taken “looking” at the back of a daylighted room with warm/neutral decor and clear glass windows is shown in Fig. 8c. The measurement was taken on a sunny day, and, although there was no direct sun in the space at that time, the SPD incident on the window would probably have been dominated by reflected sunlight. The decor in the space would be described as warm/neutral since there are no highly saturated colors present. The SPD is fairly flat over most of the visible range, and the illumination in the space would indeed be described as essentially neutral. In the range of 600–700 nm, it is possible to discern some of the features present in the standard illuminants (Fig. 4). Note however that the SPD for daylight indoors cannot in the main correspond to *D65* because all standard glass materials absorb light at short wavelengths.

The SPD for the light from the stained glass window is perhaps rather more neutral than one might expect from the photograph of the window, Fig. 8d. The visual impression of the window suggests that greens and yellows will be dominant in the SPD. However, for the window as a whole, the bulk of the light

transmission occurs through the clear sections of the glass, resulting in a fairly neutral spectrum overall. This is because the tinted glass has a lower transmission than clear glass. The more saturated the color of the stained glass, the lower its transmission, and so the smaller its contribution to overall illumination. Notwithstanding the striking visual impression made by stained glass windows, the overall SPD of the light through them can be surprisingly neutral provided that a proportion of the glass in the window is clear.

Applications

Perhaps, the first practical application of the science of daylight color was in the field of cinematography, where the various phases of daylight were associated with a correlated color temperature (see Table 1). In moviemaking, one scene could take several days to film, and if done so under natural lighting, it is important to maintain color consistency between the various “takes.” Cameramen can use subtle filters, or film stock with different spectral sensitivities, to help maintain a desired CCT between “takes” filmed on days with different daylight conditions [7]. Filming indoors was largely carried out under artificial lighting. Thus, any consideration of the color of daylight indoors was largely an aesthetic concern [1]. However, recent advances in:

- i. Understanding the basis for the nonvisual effects of daylight
- ii. The manufacture of variable transmission glazing

have caused scientists and practitioners to pay attention to the color or, more specifically, the spectra of daylight experienced indoors by the occupants of buildings. The sections below describe some of the first studies to model (section “[The Nonvisual Effects of Daylight](#)”) and measure (section “[Variable Transmission Glazing](#)”) daylight spectra for the two application scenarios noted above.

The Nonvisual Effects of Daylight

It is now well recognized that illumination received at the eye is responsible for a number of effects on the human body that are unrelated in any direct sense to vision. Light has measurable neuroendocrine and neurobehavioral effects on the human body, in particular with respect to maintaining a regular sleep-wake cycle that is entrained to the natural diurnal cycle of night and day [17]. Additionally there is evidence to suggest links between daylight illumination in particular and alertness, productivity, and academic achievement [18]. In the early 1990s, it was discovered that the eye possesses an additional non-rod, noncone photoreceptor named photosensitive retinal ganglion cells or pRGCs [19]. The pRGCs communicate light-induced signals to the suprachiasmatic nuclei (sometimes referred to as the “circadian pacemaker”) in the anterior hypothalamus of the brain. This and related discoveries have led to the emergence of a new area for photobiology research concerned with daylight/light exposure and its nonvisual effects [20].

There are numerous factors operating in a building that influence the sense of well-being or, otherwise, that an occupant experiences. Daylight is known to be one of them. The emergence of the field of photobiology has led to the speculation that we might one day be able to predict for an interior space a measure of the nonvisual effect of the daylight provided by the design of the building. This speculative measure has been referred to as the “circadian potential” or “circadian efficiency” of the space [21].

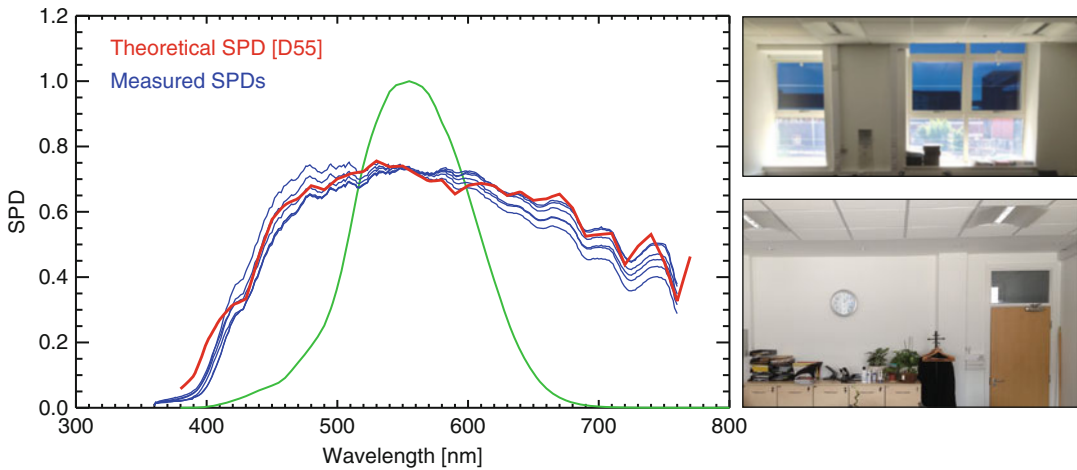
The spectral sensitivity of the photosensitive retinal ganglion cells is shifted to the blue compared to the visual sensitivity $V(\lambda)$ curve. The precise shape is not known, but the peak of sensitivity is around 460–480 nm. The timing, intensity, spectrum, duration, pattern of light received at the eye, and prior light history are the principal factors determining entrainment of the circadian cycle. Although the understanding of the various factors, and their interrelations, is incomplete, there is now sufficient empirical data to begin to experiment with predictive models for the nonvisual effects of light – artificial as well as

daylight [22]. With regard to the spectrum of light received at the eye, for two sources of illumination delivering the same lux value, one that has a blue-rich SPD could produce a stronger nonvisual effect than one which has less power in the 460–480 nm region [23]. Thus, any materials that significantly transform the spectrum of daylight entering a space should, once the scientific basis is sufficiently refined, be assessed for their potential to maintain desired nonvisual effects in addition to providing illumination for task.

Variable Transmission Glazing

The use of daylight in office buildings is generally considered to be a greatly underexploited resource. In large part this is because of the highly variable nature of daylight illumination. The natural, large variability in daylight means that users will often need to use shades to moderate excessive ingress of daylight. Most shading systems act as a “shutter” that is either open or closed, with users rarely making the effort to optimize the shading for both daylight provision and solar/glare control. And blinds are often left closed long after the external condition has changed. A glazing with a transmissivity that varies continuously between clear and dark extremes could offer a much greater degree of control over the luminous environment.

The principle behind variable transmission glazing (VTG) is straightforward: the transmission properties of the glazing are varied to achieve an “optimum” luminous and/or thermal environment. In practice, developing the technology has proven challenging. Of the various VTG formulations undergoing research and development over the last two decades, the most promising appears to electrochromic where a small applied voltage varies the transmission properties of the glass [24]. One of the most advanced of these materials is produced by SAGE Electrochromics Inc., and in 2013, it started being manufactured in high volume. This EC glass varies in visible transmission from 62 % when in the clear state down to 2 % (or even less) when fully tinted. As the glass darkens (i.e., “tints”), the longer wavelengths are diminished proportionally to a greater degree than the shorter wavelengths, giving the EC glazing good solar control properties to help prevent



Daylighting, Fig. 9 Theoretical and measured SPDs for an office with electrochromic glazing – with photographs of the EC window (*upper*) and the rear of the space (*lower*)

overheating. Optically, the consequence of this is to shift the peak in visible transmission to the blue end of the spectrum. However, while control of the luminous and thermal environment is highly desirable, occupants are believed to prefer daylight illumination that is perceived as neutral rather than tinted. Thus, the question regarding the neutrality of the illumination spectrum is an important one that needs to be addressed.

Effectively neutral illumination in spaces with EC glass can be achieved provided that a proportion of the glass remains in the clear state [25]. This was demonstrated by taking measurements of the SPD of ambient illumination in a daylighted office with EC glass under sunny conditions. The measurements were compared with a theoretical model for the predicted SPD. The theoretical SPD for the illumination in the space was predicted using the simplified expression in Eq 4. Of the eight EC panels in the window of the office used for the evaluation, five were set to full tint and three were in the clear state [26]. The spectral transmittance of the eight glass panels in the various states was calculated as a composite spectral transmittance curve based on the individual spectral transmittance curves and their proportions. Thus, in terms of the SPD, the eight panels in the various states were modeled as a single entity.

The measured SPDs were taken at six points in the office: four corresponding to the view positions of the four occupants who normally work in this space, one at the back of the office looking toward the window, and one at the window looking toward the back of the space. The six measured SPDs (normalized as before) are shown in Fig. 9. Notwithstanding the fact that the SPDs were measured at different points across the room – both looking toward and away from the window – the SPDs are in fact very similar. Furthermore, the SPDs are fairly neutral across the visual range, and the $V(\lambda)$ curve is also shown. This is consistent with experience of such spaces where the illumination turns out to be rather more neutral than one might expect from the visual impression of the window – provided that a proportion of the glass is clear. Compare this with the illumination SPD for daylight through the stained glass window (Fig. 8d), where a similar effect is observed. Also shown in Fig. 9 is the predicted SPD calculated using the D55 illuminant and a spectral transmittance properties of the composite [27]. The agreement is remarkably good, and, while these are results from a single study, they provide encouragement for wider and deeper investigations of daylight SPDs experienced by the occupants of buildings.

The Future

The daylighted indoor luminous environment can be quite different to that experienced outdoors. Appreciation of the recently discovered nonvisual effects of illumination has led to a reconsideration of the function of daylight in buildings. The quantification and evaluation of the spectral nature of daylight experienced by building occupants are now achievable using fairly modest apparatus. In the immediate future, it is likely that there will be efforts to characterize the daylight experienced indoors in terms of both the spectrum received at the eye and also the absolute (visual) magnitude of the illumination delivered for task. When the relation between daylight and nonvisual effects becomes better understood, practitioners will then have a basis to refine the appreciation of what constitutes “sufficiency” for daylight provision. That may have far reaching consequences for the drafting of future standards for daylight in buildings.

Cross-References

► Non-Visual Lighting Effects and Their Impact on Health and Well-Being

References

1. Hårleman, M., Werner, I-B., Billger, M.: Significance of colour on room character: study on dominantly reddish and greenish colours in north- and south-facing rooms. *J. Int. Colour. Assoc.* 1 (2007)
2. Hernández-Andrés, J., Romero, J., Nieves, J.L., Lee Jr., R.L.: Color and spectral analysis of daylight in southern Europe. *J. Opt. Soc. Am. A* **18**(6), 1325–1335 (2001)
3. Sanders, H., Andreau, A.: The next generation of dynamic glazing product performance. Proceedings of Glass Performance Days, Tampere, 13–15 June 2013
4. Webb, A.R.: Considerations for lighting in the built environment: non-visual effects of light. *Energy Build.* **38**(7), 721–727 (2006)
5. Commission Internationale de l’Eclairage. Colorimetry. CIE 15-2004 (2004)
6. Commission Internationale de l’Eclairage. Colorimetry – Part 2: CIE Standard Illuminants. 014-2E:2006 (2006)
7. Kodak. Cinematographer’s Field Guide. 14th edn. Aug 2010
8. Lee Jr., R.L., Hernandez-Andrés, J.: Colors of the daytime overcast sky. *Appl. Opt.* **44**(27), 5712–5722 (2005)
9. Clarke, J.A.: *Energy Simulation in Building Design*, 2nd edn. Butterworth-Heinemann, Oxford, UK (2001)
10. Mardaljevic, J., Hesong, L., Lee, E.: Daylight metrics and energy savings. *Light. Res. Technol.* **41**(3), 261–283 (2009)
11. Moon, P., Spencer, D.E.: Illuminations from a non-uniform sky. *Illum. Eng.* **37**, 707–726 (1942)
12. Commission Internationale de l’Eclairage. Standardization of luminance distribution on clear skies. CIE 22 (TC-4.2) (1973)
13. Perez, R., Seals, R., Michalsky, J.: All-weather model for sky luminance distribution-preliminary configuration and validation. *Sol. Energy* **50**(3), 235–245 (1993)
14. Preetham, A.J., Shirley, P., Smits, B.: A practical analytic model for daylight. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques, SIGGRAPH’ 99, pp. 91–100. ACM Press/Addison-Wesley, New York (1999)
15. Ward Larson, G., Shakespeare, R., Mardaljevic, J., Ehrlich, C., Phillips, E., Apian-Bennwitz, P.: *Rendering with Radiance: The Art and Science of Lighting Visualization*. Morgan Kaufmann, San Francisco (1998)
16. Schanda, J. (ed.): *Colorimetry: Understanding the CIE System*. John Wiley & Sons, Inc., Hoboken, New Jersey (2007)
17. Lockley, S.W., Dijk, D.J.: Functional genomics of sleep and circadian rhythm. *J. Appl. Physiol.* **92**, 852–862 (2002)
18. Hesong, L.: Daylighting and human performance. *ASHRAE J.* **44**(6), 65–67 (2002)
19. Foster, R.G., Provencio, I., Hudson, D., Fiske, S., De Grip, W., Menaker, M.: Circadian photoreception in the retinally degenerate mouse (rd/rd). *J. Comp. Physiol. A Sens. Neural Behav. Physiol.* **169**(1), 39 (1991)
20. Brainard, G.C., Hanifin, J.P.: Photons, clocks and consciousness. *Biol. Rhythm.* **4**(20), 314–325 (2005)
21. Pechacek, C.S., Andersen, M., Lockley, S.W.: Preliminary method for prospective analysis of the circadian efficacy of (day) light with applications to healthcare architecture. *Leukos* **5**(1), 1–26 (2008)
22. Andersen, M., Mardaljevic, J., Lockley, S.W.: A framework for predicting the non-visual effects of daylight – Part I: photobiology-based model. *Light. Res. Technol.* **44**(1), 37–53 (2012)
23. Brainard, G.C., Hanifin, J.P., Greeson, J.M., Byrne, B., Glickman, G., Gerner, E., Rollag, M.D.: Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**(16), 6405–6412 (2001)
24. Mortimer, R.J.: Electrochromic materials. *Annu. Rev. Mater. Res.* **41**(1), 241–268 (2011)
25. Mardaljevic, J.: Variable transmission glazing: a viable retrofit technology? 5th VELUX Daylight Symposium, Copenhagen, 4–5 May 2013

26. Kelly, R., Painter, B., Mardaljevic, J., Irvine, K.: Capturing the user experience of electrochromic glazing in an open plan office. CIE Midterm conference – towards a new century of light, Paris, 12–19 Apr 2013
27. Mardaljevic, J., Kelly, R., Painter, B.: Neutral daylight illumination with variable transmission glass: theory and validation. *Light. Res. Technol.* (accepted for publication)

d-Block Elements

► [Transition-Metal Ion Colors](#)

Deep Compositing

► [Compositing and Chroma Keying](#)

ΔE^*_{94}

► [CIE94, History, Use, and Performance](#)

DE*ab

► [CIELAB](#)

ΔE_{00}

► [CIEDE2000, History, Use, and Performance](#)

Demichel Equations

Roger D. Hersch
Ecole Polytechnique Fédérale de Lausanne
(EPFL), Lausanne, Switzerland

Synonyms

[Halftone layer superposition equations](#)

Parts of this contribution have appeared in the *Handbook of Digital Imaging*, Chapter 31: Base models for color halftone reproduction, Ed. Michael Kriss, J. Wiley, 2014, reproduced with permission

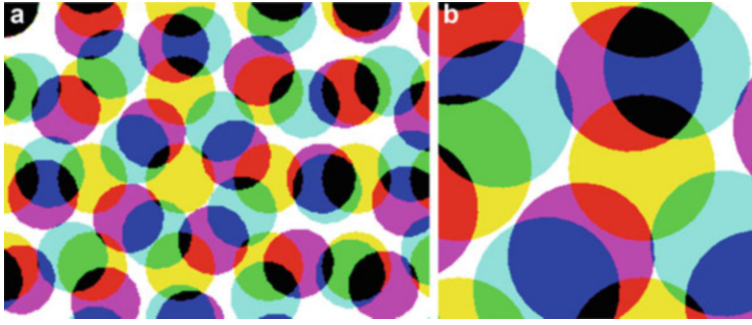
Definition

In color halftoning, the Demichel equations define the relationships between the surface coverages of the printed ink dots and the surface coverages of the Neugebauer primaries, i.e., the surface coverages of the colorants populating the color image. These equations are valid for the independently laid out ink layers that occur in stochastic screening, in error diffusion halftoning, and in mutually rotated clustered dot or line screens. The Demichel equations are used in connection with spectral models predicting the reflectances of color prints in function of the deposited amount of inks [1, 2].

Overview

Professional as well as desktop printers are either able to print with a limited number of pixel dot sizes or with a single pixel dot size. For example, in modern ink-jet printers, several droplets may be printed at the same pixel location, leading to a limited number of pixel dot sizes. In classical offset prints, a pixel location is either printed at a fixed size or remains unprinted. Color variations are created by halftoning, i.e., by creating with the different colored inks halftone dots having predetermined surface coverages.

For the sake of clarity, let us consider a cyan, magenta, and yellow ink print halftoned with classical mutually rotated halftone dots. As shown in Fig. 1, the superposition of ink dots of different colors yields new colors. The superposition of a cyan and a yellow ink dot yields a green dot portion; the superposition of a cyan and a magenta ink dot yields a blue dot portion; the superposition of a magenta and a yellow ink dot yields a red dot portion; and the superposition of a cyan, a magenta, and a yellow dot yields a black dot portion. Like in a puzzle, the space of a color halftone comprises juxtaposed dot portions of the colors white (no ink), cyan (cyan only), magenta (magenta only), yellow (yellow only), red (magenta superposed with yellow), green (cyan superposed with yellow), blue (cyan superposed with magenta), and black (cyan superposed with



Demichel Equations, Fig. 1 (a) Clustered dot color halftone with the cyan layer at orientation 15°, the magenta layer at orientation 75°, and the yellow layer at orientation 0° and (b) enlarged part of the halftone showing the new

colorants red, green, blue in regions where two of the cyan, magenta, and yellow ink dots overlap and black where the three ink dots overlap

magenta and yellow). These colored dot portions are called “Neugebauer primaries” or simply “colorants.”

The *Demichel equations* enable calculating the relative surface covered by each colorant. They assume that the ink layers are printed independently one from another. In the present context, independence means that if a light ray hits a given position on the paper, the probabilities that the different halftone ink layers contain an ink dot at this position are independent one from another. As illustration, consider a first layer (cyan) whose halftone dot grows horizontally and a second ink layer (magenta) whose halftone dot grows vertically.

Let us assume that the total surface of the halftone element is one. The respective surface coverages of the cyan and magenta inks are respectively c and m , having values between 0 and 1. Let us assume that light rays are thrown into the halftone element surface, and that their probability to fall within a specific location of that surface follows a uniform distribution. Then the probability $P(c)$ to fall on a cyan ink dot of surface coverage c is equal to c and the probability $P(m)$ to fall on a magenta ink dot of surface coverage m is equal to m .

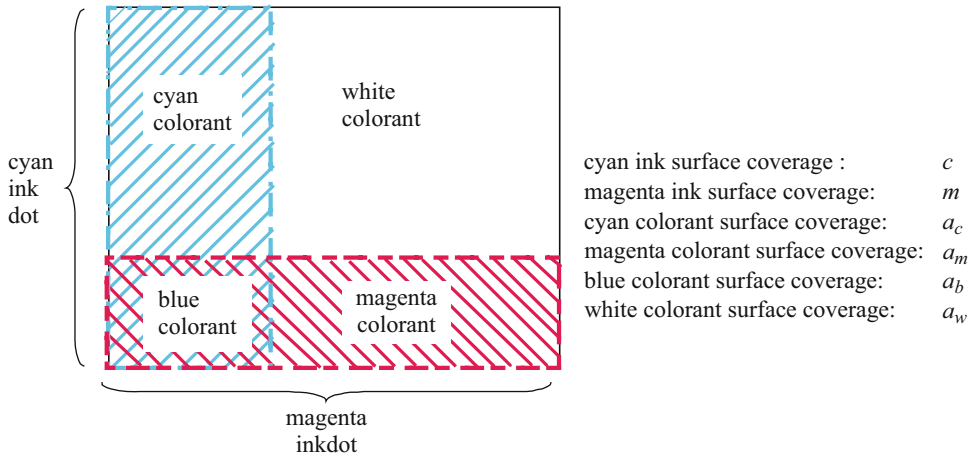
The halftone surface is a “puzzle” of juxtaposed colorants, formed by the paper, the inks, and their superpositions. In Fig. 2, the *colorant cyan* is the region covered by ink cyan only, without superposed magenta, and the *colorant magenta* is the region covered by ink magenta only, without superposed cyan. *Colorant white* is the region covered by no ink, and *colorant blue* is the region covered by both cyan and magenta inks.

If the ink halftone dots are laid out independently, as in Fig. 2, the probability of a light ray to hit the cyan colorant is the probability of hitting the cyan ink multiplied by the probability of not hitting the magenta ink. Similar considerations apply for the other colorants. We obtain

$$\begin{aligned} P(\text{Colorant} = \text{pure cyan}) &= P(c) \cdot (1 - P(m)) \\ P(\text{Colorant} = \text{pure magenta}) &= P(m) \cdot (1 - P(c)) \\ P(\text{Colorant} = \text{blue}) &= P(c) \cdot P(m) \\ P(\text{Colorant} = \text{white}) &= (1 - P(c)) \cdot (1 - P(m)) \end{aligned} \quad (1)$$

Since for an incoming photon the probability to hit a given colorant is proportional to the surface coverage of that colorant, we can deduce the surface coverages of the individual colorants

$$\begin{aligned} a_c &= P(\text{Colorant} = \text{pure cyan}) = P(c) \cdot (1 - P(m)) = c \cdot (1 - m) \\ a_m &= P(\text{Colorant} = \text{pure magenta}) = P(m) \cdot (1 - P(c)) = (1 - c) \cdot m \\ a_b &= P(\text{Colorant} = \text{blue}) = P(c) \cdot P(m) = c \cdot m \\ a_w &= P(\text{Colorant} = \text{white}) = (1 - P(c)) \cdot (1 - P(m)) = (1 - c) \cdot (1 - m) \end{aligned} \quad (2)$$



Demichel Equations, Fig. 2 Example of a halftone with one *cyan* ink half-tone dot (*dashed-dotted*) and one *magenta* ink half-tone dot (*dashed*), yielding four different colorant surfaces, namely *white*, *cyan* only, *magenta* only, and *blue*

This calculation of colorant surfaces has been introduced by Demichel [3] and further elaborated by Neugebauer [4].

Along the same reasoning lines, with 3 inks, we obtain Eq. 3 expressing the 8 colorant surface coverages as a function of the cyan (c), magenta (m), and yellow (y) ink surface coverages. The

colorants are white, cyan, magenta, yellow, red (superposed magenta and yellow), green (superposed cyan and yellow), blue (superposed magenta and cyan), and black (superposed cyan, magenta, and yellow) with respective surface coverages a_w , a_c , a_m , a_y , a_r , a_g , a_b , and a_k .

$$\begin{aligned}
 \text{white : } a_w &= (1 - c) (1 - m) (1 - y); & \text{cyan : } a_c &= c (1 - m) (1 - y) \\
 \text{magenta : } a_m &= (1 - c) m (1 - y); & \text{yellow : } a_y &= (1 - c) (1 - m) y \\
 \text{red : } a_r &= (1 - c) m y; & \text{green : } a_g &= c (1 - m) y \\
 \text{blue : } a_b &= c m (1 - y); & \text{black : } a_k &= c m y
 \end{aligned} \tag{3}$$

With 4 inks we obtain similar expressions (4) for the surface coverages of the 16 colorants, where k is the surface coverage of the black ink:

$$\begin{aligned}
 \text{white : } a_w &= (1 - c) (1 - m) (1 - y) (1 - k); & \text{cyan : } a_c &= c (1 - m) (1 - y) (1 - k) \\
 \text{magenta : } a_m &= (1 - c) m (1 - y) (1 - k); & \text{yellow : } a_y &= (1 - c) (1 - m) y (1 - k) \\
 \text{red : } a_{my} &= (1 - c) m y (1 - k); & \text{green : } a_{cy} &= c (1 - m) y (1 - k) \\
 \text{blue : } a_{cm} &= c m (1 - y) (1 - k); & \text{chromatic black : } a_{cmy} &= c m y (1 - k) \\
 \text{black : } a_w &= (1 - c) (1 - m) (1 - y) k; & \text{cyanBlack : } a_{ck} &= c (1 - m) (1 - y) k \\
 \text{magentaBlack : } a_{mk} &= (1 - c) m (1 - y) k; & \text{yellowBlack : } a_{yk} &= (1 - c) (1 - m) y k \\
 \text{redBlack : } a_{myk} &= (1 - c) m y k; & \text{greenBlack : } a_{cyk} &= c (1 - m) y k \\
 \text{blueBlack : } a_{cmk} &= c m (1 - y) k; & \text{totalBlack : } a_{cmk} &= c m y k
 \end{aligned} \tag{4}$$

These equations are valid in all cases where the ink halftone dots are laid out independently, e.g., in stochastic screening, in error diffusion, and in mutually rotated clustered dot or line screens [5].

References

1. Wyble, D.R., Berns, R.S.: A critical review of spectral models applied to binary color printing. *Color Res. Appl.* **25**, 4–19 (2000)
2. Hébert, M., Hersch, R.D.: Review of spectral reflectance models for halftone prints: principles, calibration, and prediction accuracy. *Color Res. Appl.* **40**(4), 383–397 (2015)
3. Demichel, M.E.: *Procédé* **26**(3), 17–21 (1924)
4. Neugebauer, H.E.J.: “Die theoretischen Grundlagen des Mehrfarbendrucks”, *Zeitschrift fuer wissenschaftliche Photographie* **36**, 36–73, (1937), translated by Wyble, D. and Kraushaar A, in “The theoretical basis of multicolor letterpress printing,” *Color Res. Appl.* **30**, 323–331 (2005)
5. Amidror, I., Hersch, R.D.: Neugebauer and Demichel: dependence and independence in n-screen superpositions for colour printing. *Color Res. Appl.* **25**, 267–277 (2000)

Depth in Color

► [Chromostereopsis](#)

Deuteranopia

Galina V. Paramei¹ and David L. Bimler²

¹Department of Psychology, Liverpool Hope University, Liverpool, UK

²School of Psychology, Massey University, Palmerston North, New Zealand

Synonyms

[Color blindness](#); [Congenital color vision abnormality](#); [Daltonism](#); [Dichromacy](#); [“Green-blindness”](#); [Red-green deficiency](#); [X-linked inherited color vision deficiency](#)

Definition

Deuteranopia (from the Greek *deuteros*, second, + *an*, not, + *opia*, a visual condition) is a congenital form of severe color deficiency (dichromacy) affecting the red-green opponent color system. Deuteranopes do not distinguish between colors along a specific direction (deuteranopic confusion lines) in color space and are able to match all colors using two primaries (unlike trichromats who require three primaries). Deuteranopia is linked to the X-chromosome and arises from loss or alteration of the gene encoding the opsin of the middle-wavelength (M-) photopigment, this malstructure resulting in the absence of functioning M-cones in the retina. Deuteranopia follows a recessive pedigree pattern, with incidence of ca. 1% in the Caucasian male population. It is distinct from deuteranomaly in which some degree of discrimination among such colors remains.

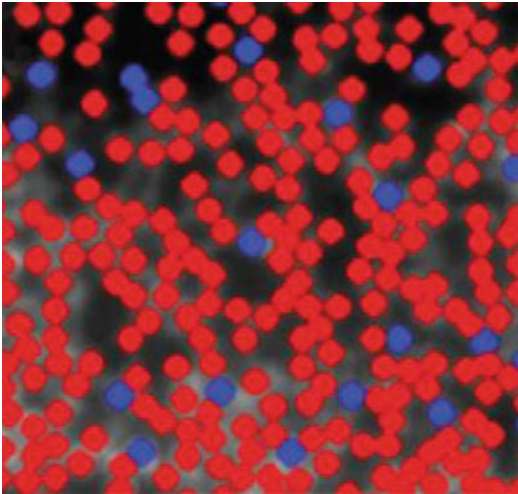
Retinal Mosaic

At the level of retinal function, deuteranopia is the absence of M-cones (Fig. 1) (“► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)” by Stockman), i.e., cone cells expressing a photopigment containing an M-opsin component, tuned for selective sensitivity to medium to longer wavelengths such as green and yellow light [1].

Genetic Background

Functional loss of the L-cone photoreceptor cells is caused by a mutant cone photopigment gene on the X-chromosome (“► [Photoreceptors, Color Vision](#)” by Rodríguez-Carmona). The gene coding for the M-opsin may be missing altogether from a deuteranope’s genome; it may be replaced by a second copy of the L-opsin (sensitive to longer wavelengths); or it may be present but damaged so that the protein is not expressed or cannot be transported to the ciliary segment of cone cells expressing that gene [1].

When present, one or more copies of the M-opsin gene are located on the X-chromosome,



Deutanopia, Fig. 1 Adaptive optics retinal imaging: pseudocolor image of the deuteranopic cone mosaic; all cones are of the L (*red*) or S (*blue*) type (Carroll, J., Neitz, M., Hofer, H., Neitz, J., Williams, D.R.: Functional photoreceptor loss revealed with adaptive optics: An alternate cause of color blindness. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8461–8466 (2004); Fig. 4a, p. 8465. Copyright (2004) National Academy of Sciences, U.S.A.)

preceded by one copy of the L-opsin gene (“► [Cone Fundamentals](#)” by Rodríguez-Carmona). This architecture, and the high similarity of the two genes, facilitates a recombination of DNA during meiotic crossover, leading to deleterious mutations of the M-opsin-encoding gene or replacement by the L-opsin [2].

With two X-chromosomes, females are far less prone to deutanopia. Notably, mothers and daughters of deuteranopes can be heterozygous for deutanopia (i.e., carriers of the condition) with one normal and one aberrant X-chromosome; the former is active in a portion of their cone cells, with normal and abnormal areas interspersed across the retina. Despite having enough M-cones to sustain normal trichromatic vision, deutanopia carriers manifest subtle departures from it, with more frequent Ishihara-test errors (“► [Color Vision Testing](#)” by Paramei, Bimler), wider anomaloscope matching ranges [3], and reduced subjective prominence of red-green dissimilarity [4].

Confusion Lines

L-signal/M-signal comparison in bipolar ganglion color-comparison cells is not present in deuteranopes (“► [Color Vision, Opponent Theory](#)” by Wuerger, Xiao); hence, there is no difference in appearance of any colors which differ only in the L-M signal. That is, although deuteranopes lack the M- or “green” cones, they are not “green-blind” in a literal sense, but rather are red-green blind. Examples of colors confused by deuteranopes are red, yellow, and green or bluish green, gray, and purple. Plotted in a chromaticity diagram (► [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)), the confused colors lie on straight “confusion lines” that converge at the deuteranope copunctal point (Fig. 2).

Luminous Efficiency: Wavelength Discrimination

The absence of M-cones has little effect on deuteranopes’ sensitivity to colors. Their luminosity function has peak sensitivity around 560 nm, compared to the normal trichromatic peak near 555 nm, slightly red shifted toward the λ_{\max} of the L-cone pigment [6, 7].

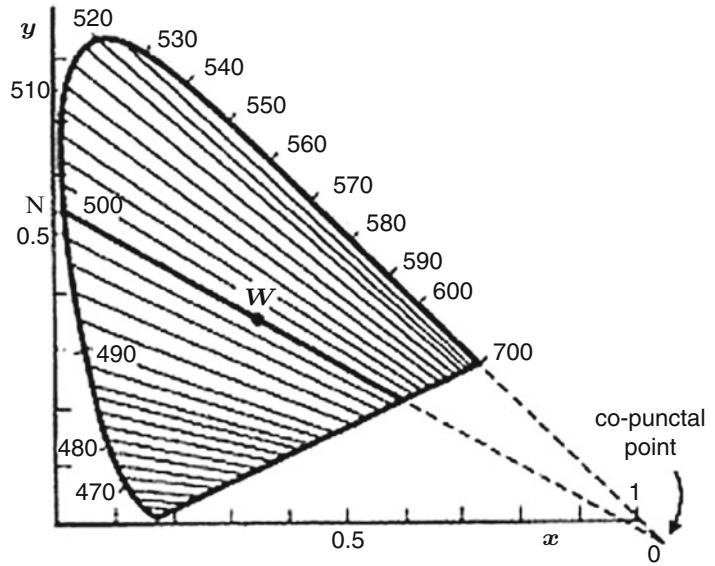
Hue discrimination, as a function of wavelength, is considerably worse for deuteranopes than for normal trichromats. Their lowest discrimination thresholds ($\Delta\lambda$), approaching the trichromatic value, are in the deuteranopic “neutral point,” 495–500 nm (see below). $\Delta\lambda$ rises on both sides of this minimum, increasing rapidly both on the short-wavelength side ($\lambda < 480$ nm) and on the longer-wavelength side ($\lambda > 520$ nm) until wavelengths are not distinguished beyond approximately 530 nm [6].

Color Perception

A deuteranope recognizes far fewer distinct colors than a trichromat (“► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)” by Bonnardel). Deuteranopic naming of spectral colors reveals a “neutral zone” of subjective gray ranging between 495 and 505 nm, with midpoint at ca. 498 nm [8]. For a deuteranope, all equiluminant colors, instead of being arranged in a two-dimensional plane (“► [Psychological Color](#)

Deuteranopia,

Fig. 2 Confusion lines indicate colors that cannot be distinguished by a deuteranope. The line going through the white point (W) indicates neutral (achromatic) colors. The lines converge at the copunctal point defining the chromaticity of the missing M-cone fundamental. Drawn in the CIE 1931 chromaticity diagram [5]



Space and Color Terms” by Bimler), form a single-dimensional line, running between the two extremes of saturated blue and saturated yellow [9]. Deuteranopes distinguish approximately 31 steps along this line, each corresponding to a separate band of interchangeable colors [7]. The usually separate chromatic qualities of “saturation” and “hue” are thus combined along this line. This deuteranopic linear representation of equiluminant colors is part of a broader reduction of a three-dimensional into a two-dimensional color space.

Whether the subjective qualities of the two extremes are identical to the normal trichromatic “blue” and “yellow” remains controversial and rests on reports from a handful of individuals with unilateral dichromacy. The blue-yellow convention is used in dichromacy simulation software, which converts color scenes to show how a deuteranope would see these [10].

Despite this model of a two-hue palette for self-luminous colors, the color percepts of deuteranopes remain in question [11]. Their object-color palette was proved to include all six Hering component colors (red, green, blue, yellow, white, and black) [12]. However, in contrast to normal trichromats, deuteranopes perceive only a weak green component in object colors and do not experience “unique green”; their red and green components are confounded with an achromatic

element (“► [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)” by Bonnardel).

Diagnosis of Deuteranopia

Deuteranopia can be diagnosed using several tests varying in sensitivity, specificity, and ease of use [13] (“► [Color Vision Testing](#)” by Paramei, Bimler). Pseudoisochromatic plates and arrangement/panel tests directly address a deuteranope’s confusions between color pairs along the confusion lines. Both types of tests have the advantage of rapid administration and ease of interpretation, making them suitable for field studies or epidemiological applications.

In the pseudoisochromatic tests (► [Pseudoisochromatic Plates](#), by Marisa Rodriguez-Carmona, Ishihara test, Hardy-Rand-Rittler test), each diagnostic plate shows a normal observer design (digit or curved lines) defined by a color difference which disappears for deuteranopes, so that the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction.

In the arrangement tests (Farnsworth-Munsell 100 hue test, Farnsworth D-15 test), the one dimensionality of a deuteranope’s color “plane” collapses a color circle into a line (“► [Color Circle](#)” by Green-Armytage), causing drastic

departures from a normal observer's sequence. Errors, or cap transposition "distances," are summed to quantify the severity of deuteranopia. Plotted graphically, the transpositions in the D-15 manifest as diametrical circle crossings, indicating the deuteranopic angle of the confusion axis ("► [Color Perception and Environmentally Based Impairments](#)" by Paramei).

The "gold standard" of diagnosis of deuteranopia is the Rayleigh match on a small viewing field (2°) of the Nagel anomaloscope. Conventionally, the levels of 670 nm (red) and 545 nm (yellow) lights are fixed, so that either wavelength at maximum setting stimulates L-cones equally: the so-called deutan mode in which a deuteranope will match *any* red/green ratio to the yellow light of constant luminance [13] (S-cones being unresponsive in this spectral range). Notably, unlike protanopes ("► [Protanopia](#)" by Paramei, Bimler), deuteranopes do not dim the yellow light when adjusting it to the mixed light that is mainly red – the estimate that is diagnostic for differentiating the two red-green deficiency types. Similar to protanopes, when tested with a large viewing field (10°), in setting Rayleigh matches the majority of deuteranopes perform as anomalous trichromats. Various factors related to retinal eccentricity are presumed, e.g., rod intrusion, S-cone intrusion, spatial variation in the macular pigment, etc. [7].

More recently developed computerized tests for color vision diagnosis, the Cambridge Colour Test and Colour Assessment and Diagnosis test ("► [Color Vision Testing](#)" by Paramei, Bimler), interactively quantify any loss of chromatic sensitivity along the deuteranopic confusion line. For a deuteranope, in the frame of the CIE ($u'v'$) 1976 chromaticity diagram ("► [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)" by Schanda), a MacAdam ellipse of indistinguishable colors displays a major axis lengthened to infinity and orientation pointing to the deuteranope copunctal point [14, 15].

Incidences of Deuteranopia

Prevalence of inherited deuteranopia varies between human populations. For Caucasian

populations, the reported incidence of protanopia is 1.27 % in males and 0.01 % in females [7]. Recent surveys of red-green deficiency in populations of different racial origin found the incidence of deuteranopia in Caucasian males to vary between 0.96 % and 1.37 %. As in the case of protanopia ("► [Protanopia](#)" by Paramei, Bimler), in indigenous Native American, Pacific, Asian, and African populations, the incidence of deuteranopia is markedly lower than in Caucasian, about half this rate, and, too, is observed to raise in coastal areas where migrant Europeans settled in historical times [16]. The increasing prevalence of congenital red-green deficiencies, including deuteranopia, has tentatively been explained by relaxation in selection pressures as hunting and gathering cultures evolve toward industrialized societies [7]. More recently, however, it is rather attributed to genetic drift and founder events (variable sampling of the gene pool contingent on the population size) [16].

Cross-References

- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE Chromaticity Coordinates \(xyY\)](#)
- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision Testing](#)
- [Color Vision, Opponent Theory](#)
- [Photoreceptors, Color Vision](#)
- [Protanopia](#)
- [Pseudoisochromatic Plates](#)
- [Psychological Color Space and Color Terms](#)
- [Tritanopia](#)

References

1. Carroll, J., Neitz, M., Hofer, H., Neitz, J., Williams, D. R.: Functional photoreceptor loss revealed with adaptive optics: an alternate cause of color blindness. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8461–8466 (2004)
2. Neitz, M., Neitz, M.: The genetics of normal and defective color vision. *Vision Res.* **51**, 633–651 (2011)
3. Konstantakopoulou, E., Rodríguez-Carmona, M., Barbur, J.L.: Processing of color signals in female carriers of color vision deficiency. *J. Vis.* **12**(2), art. 11 (2012). doi:10.1167/12.2.11
4. Bimler, D., Kirkland, J.: Colour-space distortion in women who are heterozygous for colour deficiency. *Vision Res.* **49**, 536–543 (2009)
5. Pitt, F.H.G.: Characteristics of dichromatic vision. Medical research council special report series, no. 200. His Majesty's Stationery Office, London (1935)
6. Pitt, F.H.G.: The nature of normal trichromatic and dichromatic vision. *Proc. R. Soc. Lond. B* **132**, 101–117 (1944)
7. Sharpe, L.T., Stockman, A., Jägle, H., Nathans, J.: Cone genes, cone pigments, color vision and colorblindness. In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 3–52. Cambridge University Press, Cambridge (1999)
8. Massof, R.W., Bailey, J.E.: Achromatic points in protanopes and deutanopes. *Vision Res.* **16**, 53–58 (1976)
9. Paramei, G.V., Izmailov, C.A., Sokolov, E.N.: Multidimensional scaling of large chromatic differences by normal and color-deficient subjects. *Psychol. Sci.* **2**, 244–248 (1991)
10. Brettel, H., Viénot, F., Mollon, J.D.: Computerized simulation of color appearance for dichromats. *J. Opt. Soc. Am. A* **14**, 2647–2655 (1997)
11. Broackes, J.: What do the colour-blind see? In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, pp. 291–405. MIT Press, Cambridge, MA (2010)
12. Logvinenko, A.: On the colours dichromats see. *Color. Res. Appl.* **39**, 112–124 (2014)
13. Dain, S.J.: Clinical colour vision tests. *Clin. Exp. Optom.* **87**, 276–293 (2004)
14. Regan, B., Reffin, J.P., Mollon, J.D.: Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vision Res.* **34**, 1279–1299 (1994)
15. Barbur, J.L., Rodríguez-Carmona, M., Harlow, J.A.: Establishing the Statistical Limits of “Normal” Chromatic Sensitivity. Publication CIE x030, Ottawa/Ontario (2006)
16. Birch, J.: Worldwide prevalence of red-green colour deficiency. *J. Opt. Soc. Am. A* **29**, 313–320 (2012)

Diagnosis of Defective Color Vision

► Color Vision Testing

Diagnostic Color

► Memory Color

Dichromacy

► Comparative Color Categories

► Deuteranopia

► Protanopia

► Tritanopia

Dichromat Color Naming

► Color Categorization and Naming in Inherited Color Vision Deficiencies

Diffuse Lighting

Wout van Bommel
Nuenen, The Netherlands

Definition

Lighting in which the light on a point, plane, or object in the lit space is incident with equal luminous intensities from all directions.

Daylighting

The sun of course is the primary source of daylighting. When there are clouds in the sky, part of the primary sunlight is absorbed and part reflected from the clouds. The result is a combination of direct sunlight and secondary reflected light from

the clouds. In a situation where the sky is completely covered with clouds, the overcast sky acts as a perfect diffuser, resulting in complete diffuse daylighting. The overcast sky lighting condition is used by architects and lighting designers to calculate the lowest level of daylight contribution in buildings. With a partly overcast or complete clear sky, the contribution will be higher. To standardize these calculations, CIE has introduced the so-called CIE overcast sky with a defined luminance distribution. The luminance of the CIE overcast sky is highest in the zenith, the point straight above the point under consideration, and decreases to one third at the horizon. For simple noncomputerized calculations, sometimes the CIE uniform sky is used in which the luminance distribution of the sky is uniform.

Electric Lighting

Near complete diffuse electric lighting is obtained from indirect lighting in interiors. Here the electric lighting is directed toward the ceiling to create a near uniform ceiling luminance. The light reaches the working plane only after reflection from the ceiling. Since a white ceiling does usually not have a reflectance higher than some 70 %, at least 30 % of the light is absorbed at the ceiling.

In some interiors, artificial skies are created with luminous ceilings existing of large-sized luminaires with opal type of louvers. The lighting of these artificial skies is usually diffuse although with LED or OLED solutions, nonuniform skies and consequently non-diffuse lighting can be created.

Conventional indoor lighting luminaires for general lighting purposes in fact have a diffuse component with, under elevation angles of some 60–70°, a directional component. Dependent on the actual light distribution of the luminaire, the lighting is more diffuse or more directional.

Consequences of Diffuse Lighting

The consequences of complete diffuse lighting are best explained by considering the condition of a

clouded sky. The diffuse lighting of it is not appreciated very much: it gives a dull, monotonous impression in which shadows, depth, shape, and texture are poor or completely absent. This has negative consequences for both visual pleasantness and visual performance. An illustration of the latter is a complete cloudy sky on a ski slope: irregularities and bumps in the snow cannot be seen because of the absence of shadows. Indeed it is generally considered dangerous to make a somewhat difficult descend under these viewing conditions.

Diffuse electric lighting has the same negative consequences for both pleasantness and visual performance. The effect of lack of shadows, here, is often referred to as bad modeling.

Good interior lighting consists of both a diffuse and directional component. This again is analogue to a relative clear sky outdoors with a strong directional component from the sun with some diffuse light from a couple of white clouds.

Cross-References

- [Architectural Lighting](#)
- [Daylighting](#)
- [Interior Lighting](#)

Dry or Mechanical Finishing

- [Finish, Textile](#)

Dye

Jae Hong Choi
Department of Textile System Engineering,
Kyungpook National University, Daegu,
South Korea

Synonyms

[Colorant](#)

Definition

The term *dye* refers to compounds which can impart color to a substrate when applied in solution from either aqueous or organic solvents. The substrates include textiles, plastics, polymers, etc. Both the applications to materials and color constitutions of dyes involve essentially chemical principles. Dyes and pigments are both commonly applied as a coloring material, but they are distinguished on the basis of their solubility either in water or organic solvents; therefore dyes are soluble, whereas pigments are insoluble.

History of the Development of Synthetic Dyes

Until the middle of the nineteenth century, the dyeing of textiles and leather was carried out using dyes derived from animal and vegetable sources. Of such natural dyes, indigo is perhaps the most well-known and most important material [1]. Alizarin is the most important species of anthraquinone-based red dyes obtained from the roots of plants of the *Rubiaceae*. A highly prized plant was madder which produced dyeings on cotton known as Turkey Red by means of Ca-Al mordant. Also of importance in early times was the use of the red dye, Kermes, obtained from insects [1, 2]. In 1856, WH Perkin synthesized a brilliant violet dye, Aniline Purple (which later became known as Mauveine), by the oxidation of a mixture of aniline bases [3–5]. Perkin's (accidental) discovery of Mauveine and his development of methods to apply the dye to silk and cotton stimulated enormous commercial interest in dye synthesis. Further discoveries quickly followed, such as that of the diazo reaction by Griess in 1859 and the introduction of the first azo dye, Bismarck Brown, by Martius in 1863 [6] and the structural elucidation and synthesis of Alizarin that occurred in 1868 and Indigo by von Baeyer in 1883 and Heumann in 1890; shortly afterwards the first anthraquinoid vat dye indanthrone was prepared by Bohn in 1901 [6, 7]. As a result of intensive attempts to devise new dyes for application to the fiber cellulose

acetate, the water-soluble *Ionamin* range of dyes was introduced in 1922 [8] which were the precursors of the current wide range of sparingly soluble disperse dyes. The first reactive dyes for cotton were introduced by ICI in 1956; such dyes are able to attach covalently to cellulosic fiber, thereby offering dyeings of excellent fastness to washing and wet treatments [9]. From the 1990s, non-textile applications for dyes, such as information technology, reprographics, optoelectronics, and solar cells, have contributed to the diversity of modern synthetic dyes [1, 10].

Principles of Color Constitution in Dyes

When ultraviolet or visible radiation is absorbed, an n or π electron is promoted and occupies an empty π^* or σ^* antibonding orbital, the wavelength, λ , of the photon being equal to the energy difference, ΔE , between the two levels, ground state and excited state, involved:

$$\lambda = hc/\Delta E$$

where h is Planck's constant and c the velocity of light.

When the energy difference is equal to the energy of visible light, the compound has an absorption band in the visible region (400~700 nm) and exhibits the color complementary to that of the absorbed light [7].

Early Theories Using Qualitative Method

The elucidation of the relationship between the color of dyes and their chemical constitution has attracted much attention. In 1876, Witt suggested that color arises from the combination of unsaturated groups, such as $-\text{N}=\text{N}-$, $-\text{C}=\text{O}$, and NO_2 , with basic groups, such as $-\text{NH}_2$, $-\text{OH}$, and $-\text{NR}_2$, within a dye molecule. Witt defined the former groups as *chromophores* and the latter as *auxochromes*. The terminology of *bathochrome* and *hypsochrome* was also coined at the same time [6, 11]. Hewitt defined the role of conjugation in a dye molecule as "the longer the conjugated chain, the more bathochromic the dye" and pointed out the importance of a conjugated chain

for the production of color. With the discovery of the electron in 1897 [12], attention was then given to the behavior of electrons within dye structures as providing the origin of color.

Modern Theories

All modern theories involve two distinct approaches, namely, the Valence Bond (VB) and Molecular Orbital (MO) theories, which are developments from quantum mechanics. VB theory is principally based on bonding valence electron pairs localized between specific atoms in a dye molecule [12], whereas MO theory has played a key role in the development of modern color theories for solving the intricate relationships between color and chemical constitution. A satisfactory method to calculate the transition energy and intensity of visual color was developed in terms of the Linear Combination of Atomic Orbitals (LCAO) method and the Hückel Molecular Orbital (HMO) method, which were originally developed for unsaturated hydrocarbons. Hückel assumed that the transition state would correspond to the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

Classification and Nature of the Chromogens

Colored organic molecules can be rationalized by classification into four groups [13]:

1. $n \rightarrow \pi^*$ chromogens
2. Donor–acceptor chromogens
3. Acyclic and cyclic chromogens
4. Cyanine-type chromogens

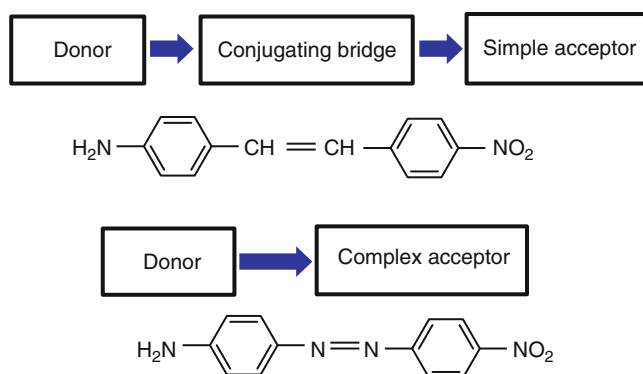
Thus, dyes which exhibit electronic transitions from a lone pair, nonbonding atomic orbital of the molecule belong to the $n \rightarrow \pi^*$ group of chromogens. However, most dyes are regarded as donor–acceptor chromogens, which, characteristically, possess electron donor and electron acceptor groups that are linked by a conjugated π -electron system (the conjugating bridge). Depending on the role of the accepting group, donor–acceptor chromogens can be subdivided into simple acceptors or complex acceptors, as represented in Fig. 1 [13].

For azo dyes, in general, the strength of the electron acceptor correlates closely with the Hammett σ constant of the substituent. As the value of the σ constant of a substituent increases, the electron acceptor group exerts a more powerful effect and produces a larger bathochromic shift (absorption maximum moves to longer wavelength).

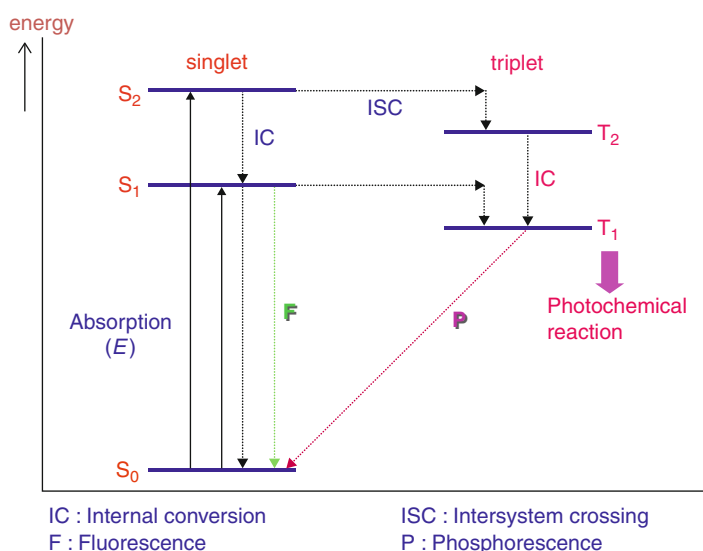
Effect of Substituents on Color in Azo Dyes

In simple azo dyes, a weak long-wavelength absorption is observed due to the overlap of the atomic orbitals of the two nitrogen atoms, giving rise to an $n \rightarrow \pi^*$ transition. The intensities of $n \rightarrow \pi^*$ transitions vary depending on whether the transition is symmetrically allowed or forbidden [14]. A weak absorption band, involving $n \rightarrow \pi^*$ transition, tends to be overshadowed by the more intense $\pi \rightarrow \pi^*$ absorption band. If an electron-donating group, such as an dialkylamino group, is introduced into the cyclic unsaturated residue (**D**) of an azo dye, of general formula $A - N = N - D$, an intense new absorption

Dye, Fig. 1 Schematic representation of simple acceptor and complex acceptor



Dye, Fig. 2 Energy-level scheme of a photoexcited dye molecule (Jablonski diagram)



D

band arises as a result of a $\pi \rightarrow \pi^*$ transition, which is referred to as a charge-transfer (CT) transition. The migration of a π electron delocalized in the nitrogen orbital of the amino group towards the azo dye is responsible for such charge-transfer transitions [13]. A further bathochromic shift can be exerted by the additional incorporation of electron-withdrawing groups, such as NO_2 , CN , and halogen, into the *A* ring. The greatest effect in terms of longer wavelength is achieved by placing the substituents at positions *ortho*- or *para*- to the azo group, which maximizes conjugation. Similarly, electron-donating groups situated *ortho* to the azo linkage in the *D* ring normally give rise to the most effective bathochromic shifts [13, 15].

Principles of Photofading in the Dye Molecule

Jablonski Diagram

Theoretically, once a dye molecule absorbs a photon from incident light, an electron will be excited to different energy levels. As depicted in Fig. 2, the excited molecule tends to lose its energy and returns to the ground state, this process involving the emission of fluorescence or phosphorescence, or can undergo a chemical reaction, leading to decomposition. This type of reaction is referred to as a *photochemical reaction* or photofading.

Not surprisingly, both the first singlet excited state and the first triplet excited state are the crucial energy levels for this reaction. The longer lifetime of the triplet excited state may be linked with a higher likelihood of photodegradation of the dye molecule [6, 13].

Internal conversion of the excited state is expected to be the fastest step for dyes and UV stabilizers that are employed in polymers, whereas fluorescence should be both dominant and fastest for optical brightening agents and laser dyes. Intersystem crossing should be fastest for dyes used as sensitizers [6].

Types of Photoreaction

Photofading may occur in different ways depending on the nature of the substrates, wavelength distribution of the incident light, the presence of moisture, and air and dye structure. In general, two principal types of photochemical reaction are recognized: photooxidation and photoreduction. The influence of substituents on degradation rate has been rationalized by plotting photofading rate against Hammett σ constant for aromatic azo dyes containing various substituents. Thus, if the fading rate increases as electron-withdrawing power increases (higher σ value), a reductive mechanism is favored (positive), whereas a decreased fading rate is observed when oxidative degradation dominates (negative) [11, 16]. As far

as photooxidation is concerned, singlet oxygen plays a very important role.

Solvatochromism of Dyes

Solvatochromism can best be described as the effect of a solvent on the color of a dye. Solvent–solute interactions have various influences on the absorption spectrum, such as λ_{max} value and the intensity of a band; the bandwidth may also vary.

Solvatochromism Types

Generally, in many dye molecules, the ground state is less polar than the excited state so that a polar solvent will tend to stabilize the excited state more than the ground state, leading to a bathochromic shift in the absorption maximum; this effect is termed positive solvatochromism. In contrast to positive solvatochromism, negative solvatochromism is relatively rare and is most often observed in dyes that contain a delocalized charge [6]. The interaction of a solvent with a dye molecule is greatest in polar solvents, for example, ethanol, which possess a strong permanent dipole, and is most pronounced with a solute molecule which contains a permanent dipole. Intermolecular hydrogen bonding can also give rise to shifts in the absorption spectrum, particularly in case of $n \rightarrow \pi^*$ bands. For heteroatoms, such as nitrogen, sulfur, and oxygen, the non-bonding lone pair electrons can form intermolecular hydrogen bonds with suitable protic polar solvents, such as ethanol or water. Accordingly, the excitation energy will be raised by the hydrogen bond, which stabilizes the n -electrons resulting in a hypsochromic shift in the absorption spectrum [13, 17].

Halochromism of Dyes

This term is defined as the color change of a substance accompanying a change in the pH of the solution [13]. Although many types of aminoazo dyes show this effect, it is observed only when the amino group is situated *para* to the azo linkage.

Halochromism Types

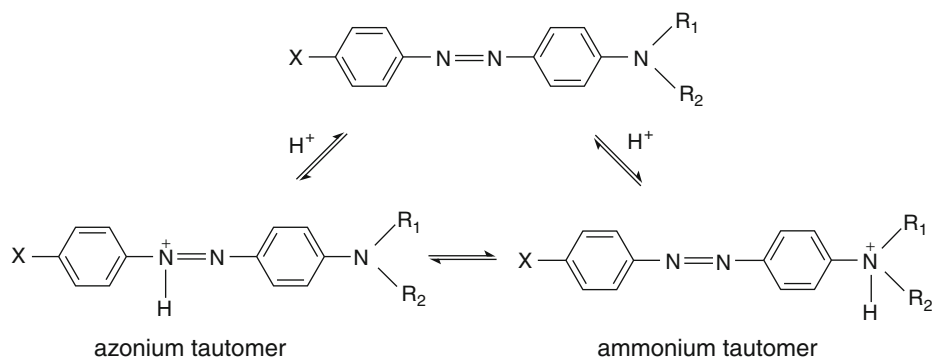
In contrast to hydroxy azo dyes, in acidic solution, many dyes derived from 4-aminoazobenzene are susceptible to protonation either at the β -nitrogen atom of the azo linkage, to give the azonium cation, or at the amino nitrogen atom to produce an ammonium ion. Bathochromic shifts combined with increased color intensity are often observed for the azonium ion. This shift is termed as a positive halochromism, whereas the opposite phenomenon is referred to as a negative halochromism [13, 18, 19].

Halochromism in 4-Aminoazobenzene Dyes

As the ground state of the azonium cation is best represented by a quinonoid structure, it is theoretically presumed that electrons migrate from the β -nitrogen atom towards the terminal amino group. Thus, as the strength of electron donation of the substituent (X) increases (Fig. 3), a more bathochromic shift is observed, because the excited state will be stabilized by the electron donor group, resulting in a smaller energy difference between the ground and excited states [20]. Although negative halochromism is rarely encountered, when several powerful electron-withdrawing groups are present as substituents (X), a hypsochromic shift may occur in acidic solution [13, 20].

Solubility of Dyes

Dyes mainly exhibit absorbing properties, whereas pigments are materials that feature scattering properties. Scattering in a colorant depends on its solubility in the substrate, the size of the colorant molecule, and its tendency to form crystals. Thus, colorants of small molecular size and good solubility usually dissolve monomolecularly in the substrate. Depending on solubility and intermolecular forces operating between the colorant molecules, some colorants may form aggregates; if the solubility of a colorant in a substrate is low, precipitation easily occurs [21]. The principal parameters in selecting a dye are that it can impart the desired color to a substrate, an intense color being essential, so that the substrate can be dyed with as little dye as possible.



Dye, Fig. 3 Protonation equilibria of 4-aminoazobenzene dyes

Future Prospects

Although dyes continue to enjoy widespread use in the coloration of textiles and other substrates, such traditional usage is being extended into other exciting areas such as conducting materials, sensors, dye-sensitized solar cells, light-emitting devices, and color filters.

Cross-References

- [Colorant](#)
- [Colorant, Natural](#)
- [Colorant, Textile](#)
- [Pigment, Ceramic](#)

References

- Freeman, H.S., Peters, A.T. (eds.): *Colorants for Non-textile Applications*. Elsevier Science, Amsterdam (2000)
- Welham, R.D.: The early history of the synthetic dye industry: part 1 the chemical industry. *J. Soc. Dye Col.* **79**(3), 98–105 (1963)
- Rys, P., Zollinger, H.: *Fundamentals of the Chemistry and Application of Dyes*. Wiley-Interscience, London (1990)
- Peter, J.T., Anthony, S.T.: A history of the international dyestuff industry. *Am. Dyestuff Rep.* **81**(11), 59–100 (1992)
- Welham, R.D.: The early history of the synthetic dye industry: part 2 the industrial history (1856–1900). *J. Soc. Dye Col.* **79**(3), 146–152 (1963)
- Zollinger, H.: *Color Chemistry: Synthesis, Properties, and Applications of Organic Dyes and Pigments*. VCH/VCH, Zürich (2003)
- McLaren, K.: *The Colour Science of Dyes and Pigments*. Adam Hilger Ltd, Bristol (1986)
- Green, A.G., Saunders, K.H.: The ionamines: a new class of dyestuffs for acetate silk. *J. Soc. Dye Col.* **39**(1), 10 (1923)
- Shore, J. (ed.): *Colorants and Auxiliaries*, vol. 1. SDC, Bradford (1990)
- Kim, S.H. (ed.): *Functional Dyes*. Elsevier, Amsterdam (2006)
- Griffiths, J. (ed.): *Developments in the Chemistry and Technology of Organic Dyes*. Blackwell Scientific Publications, Oxford (1984)
- Gordon, P.F., Gregory, P.: *Organic Chemistry in Colour*. Springer, Berlin (1987)
- Griffiths, J.: *Colour and Constitution of Organic Molecules*. Academic Press, London (1976)
- Fabian, J., Hartmann, H.: *Light Absorption of Organic Colorants*. Springer, Berlin (1980)
- Griffiths, J., Roozpekar, B.: Synthesis and electronic absorption spectra of dicyano-derivatives of 4-diethylaminoazobenzene. *J. Chem. Soc. Perkin Trans. I*, 42–45 (1976)
- Giles, C.H., Hojiwala, B.J., Shah, C.D.: Quantum efficiency measurements of fading of some disperse dyes in nylon and polyester films and in solution. *J. Soc. Dye Col.* **88**(11), 403–407 (1972)
- Christian, R., Thomas, W.: *Solvents and Solvent Effects in Organic Chemistry*, 4th edn. Wiley-VCH, Weinheim (2010)
- Griffiths, J., Hill, J., Fishwick, B.: The application of PPP-MO theory to the halochromism of 4-aminoazobenzene dyes. *Dyes Pigments* **15**(4), 307–317 (1991)
- Bello, K.A., Griffiths, J.: Violet to cyan azo dyes derived from 4-amino-3-nitrobenzaldehyde as diazo component. *Dyes Pigments* **11**(1), 65–76 (1989)
- Hallas, G.: The effect of terminal groups in 4-aminoazobenzene and disperse dyes. *J. Soc. Dye Col.* **95**(8), 285 (1979)
- Kuehni, R.G.: *Color: an Introduction to Practice and Principles*. Wiley-Interscience, Hoboken (2005)

Dye Functional

► Colorant, Nonlinear Optical

Dye, Functional

Sung-Hoon Kim and Eun-Mi Lee
Department of Textile System Engineering,
Kyungpook National University, Daegu,
Republic of Korea

Definition

Since the introduction of the first commercially available synthetic dye, *mauveine*, in 1856, traditional organic dyes have been used for the coloration of textiles by means of dyeing technology. The term “functional dye” was first coined by Japanese scholars in 1981 and stemmed from interest and activity in the field of dye chemistry interrelated with high-technology applications [1–6]. Nowadays, the term *functional dyes* refers to ultraviolet (UV) and particularly infrared (IR) active molecules which have been specifically designed for high-technology applications in fields such as electronic materials, electronic devices, reprographics, etc., as such “functional dyes” differ to traditional dyes.

Interactions of Functional Dyes

It is pertinent to examine the interactions of functional dyes with various agents before discussing the many and varied applications of functional dyes. Such interactions enable particular phenomena to be associated with each application. Functional dyes are designed to interact with electromagnetic radiation, pH, electricity, heat, pressure, and even frictional forces.

Electromagnetic Radiation

Functional dyes interact with electromagnetic radiation in the near-UV (300–400 nm), visible

(400–700 nm), and near-IR (700–1,500 nm) regions to produce a variety of effects required for high-technology (hi-tech) applications. Selective absorption of nonvisible radiation, such as UV and particularly IR, is also important in applications such as optical data storage, computer to plate, security, and printing. A color change imparted by electromagnetic radiation is called *photochromism*. Photochromism is a chemical process in which a compound undergoes a reversible change between two states having separate absorption spectra, i.e., different colors. Although photochromism is undesirable in traditional textile dyes, it is widely used in, for example, spectacles and optical data storage. Indicator dyes are probably the most familiar example of a color-change effect.

Luminescence occurs when a molecule in an excited state, normally achieved by absorption of a photon, loses some or all of the excess energy as light rather than via the normal relaxation mode of heat loss. Luminescence from the first excited singlet state to the ground singlet state is known as *fluorescence*. Fluorescence is important in biological application, laser dyes, emissive displays, and in providing vivid, bright dyes, particularly for ink-jet printing.

Heat

Most molecules lose energy from the excited state as heat. The most efficient molecules for converting electromagnetic radiation into heat are those that absorb in the near-infrared region, namely, infrared absorbers (IRAs). IRAs have attracted much interest because of their use in laser thermal transfer, optical data storage, computer-to-plate printing, and as solar screens for car windscreens and windows.

Thermochromic dyes change color with temperature (heat). Thermochromic dyes find use in direct thermal printing and as temperature sensors, as well as in clothing and novelties.

Electricity

Functional dyes have been designed to interact with electricity to produce a color change or to fluoresce. *Electrochromic dyes* change their color, generally from colorless to colored, when an

electrical voltage is applied. The main application areas for electrochromic systems are electrically switchable anti-dazzle rearview mirrors, glazing units for temperature and light control, as well as visual displays.

Frictional Force/Pressure

Piezochromic and *tribochromic functional dyes* have been designed to produce an electrostatic charge when mechanical means is applied, the former under applied frictional or grinding forces and the latter by pressure. Such compounds have not as yet enjoyed much commercial exploitation. Known as charge-control agents, *tribochromic* dyes are used in the toners for photocopiers and laser printers to both produce and regulate the triboelectric charge on the toner particles. Pressure-sensitive barochromic dyes find use in imaging (carbonless paper) and for testing pressure points in, for example, airplanes and cars.

Functional Dyes Classified by Application

As mentioned above, development of electronics industry and the affected industries is needed of functional dyes, which are developed, discolored, or achromatized by very small energy such as light, heat, electricity, and so on. Functional dyes can be classified in several ways. This entry will describe specific examples of functional dyes by classified application such as data storage, data display, energy transfer, and so on. A list of fields and application of functional dyes are given in Table 1.

Data Storage

The data storage field can be divided into two main parts, namely, *optical data storage* and *printing*. The technologies related to printing seek to produce visible images on a hard copy, usually paper or special media; ink-jet, laser printing, photocopying, and thermal printing are examples of the printing techniques employed. However, there are other technologies which produce invisible images, usually for the entertainment and publishing fields, using functional dyes;

Dye, Functional, Table 1 Application areas of functional dyes

Fields	Applications	
Data storage	Optical data storage	CD-R, DVD
	Printing	Ink for ink-jet printer, OPC, toner, thermal recording paper, D2T2
	Others	Pressure-sensitive paper, color shifting ink for counterfeit deterrence
Data display	Display	Color filter for LCD, organic EL, near-infrared absorption dyes for PDP
	Others	Digital papers, thermosensitive materials
Energy transfer	Solar cells	DSSC
	Others	Dye laser, nonlinear optical materials, organic semiconductor
Biomedical application		Diagnostic medical devices, laser dyes for photodynamic therapy

optical data storage is such a technology which concerns invisible imaging.

In its broadest sense, optical data storage involves a medium which, by optical means, employs near-infrared radiation or visible light for the recording and retrieval of digital data. Various types of storage media have been developed, such as read-only compact discs (CD-ROM) to rewritable digital versatile discs (DVD-RW); optical data recording systems which use laser addressable organic dyes include WORM (write once read many), CD-R, and DVD-R.

The mechanism of optical recording using dyes involves converting excitation energy, which is generated by a focused laser beam, into thermal energy by non-radiative decay. In simple terms, a dye is coated onto a disc substrate to provide the recording layer. In writing mode, the laser is used near to full power and “blasts” holes in the very thin layer of dye at appropriate parts of the disc, as dictated by the digitally encoded information. When the information is retrieved, the

power of the laser is turned down to avoid damaging the recording layer and is played over the surface of the disc to read the encoded signal with the aid of a detector. Since the mechanism of hole formation relies upon the dye absorbing laser radiation, heating up, and migrating from the hot region during the hole formation stage, the dye must absorb at the wavelength of the laser used. A number of technical criteria are demanded of suitable dyes, in addition to the general commercial requirements of low cost and high availability:

- Good solubility in the organic solvent used for spin coating the discs and the ability to form solid films
- High stability for environmental and processing conditions, i.e., oxidation, hydrolysis, and light
- High reflectivity of the solid film at the recording wavelength
- Low thermal conductivity and hence high sensitivity for heat-mode recording

The requirements for dyes in WORM, CD-R, and DVD-R media are very similar, but there are differences in their absorbance wavelength. Several types of infrared absorbers have received attention for use in optical data storage. Both phthalocyanine and cyanine colorants satisfy the commercially required technical criteria for optical data recording; phthalocyanines have many desirable attributes suitable for use in optical data storage.

Data Display

In recent times, there has been a significant development of display technologies which can interface with an ever-increasing number of electronic devices. The role of the display is simply to render the output from the device in a form that is intelligible to the human reader; examples of devices include computers, calculators, watches, televisions, public signboards, etc.

Cathode Ray Tube

For many decades, the cathode ray tube (CRT) was the most common display used for televisions or visual display monitors. The operating

principle in CRT involves an electron beam that is generated from a cathode located at one end of a glass vessel, which is held under vacuum. The electron beam is accelerated and is then moved horizontally or vertically using electrostatic or electromagnetic means. The inner, front surface of the glass vessel is coated with inorganic phosphors, and, when the beam of highly accelerated electrons hit this surface, the phosphors emit visible light. This process is known as *cathodoluminescence* which is an emissive technology and therefore produces bright images.

Liquid Crystal Display

Because of the large bulk and mass of CRTs, liquid crystal displays (LCDs) have been developed. LCDs employ specific organic molecules, such as *para*-cyanobiphenyl, which can be readily aligned in an electrical field. When the molecular axis of *para*-cyanobiphenyl is orthogonal to a beam of polarized light, the light is absorbed and a dark color results. When its molecular axis is parallel to the direction of the light beam, no absorption occurs and a pale-colored area results. This contrast allows monochrome imaging. Neither CRTs nor LCDs are the most suitable technology for very large area displays, where other display techniques such as plasma panels and electroluminescent devices offer advantage.

Organic Light-Emitting Devices

Organic light-emitting devices (OLEDs) are one of several technologies available for use as emissive flat panel displays. OLEDs possess the advantage that their molecular properties can be tailored, as well as inherent potential low cost, low voltage operation, and relative simplicity of processing and fabrication into devices.

Two types of OLEDs are being developed: one based on multilayers of vapor-deposited, low molecular weight materials and other on polymeric materials, either by precursor polymerization or by casting from solution and subsequent processing. Most research has been done on low molecular weight material OLED.

A wide variety of low molecular mass materials have been designed for high-performance materials. The hole-transport layer (HTL)

materials are normally substituted aromatic amines, which are known to have high hole mobility and which are applied to the surface of a glass substrate which has been coated with the conducting transparent material, indium tin oxide. The electron-transport layer (ETL) materials commonly used are emissive metal complexes, commonly aluminum, beryllium, or lanthanides. Some ETL materials are chosen because they are non-emissive so as to act as combined ET and hole-blocking layers.

Polymeric LED materials are a more recent development and have many attractive properties. One prime advantage is the ease with which it is possible to lay down thin films on large areas of a substrate either by spin coating or dip coating. Typical polymers are the green poly(*para*-phenylene vinylene) (PPV), the orange-red dialkoxyl derivatives, and the blue polyfluorene.

Energy Transfer

Energy transfer via functional dyes involved in electronic materials has been discussed above. In this section, organic semiconductors and solar cells will be considered.

Electroluminescence (EL) involves the conversion of electrical energy into nonthermal light. To achieve this conversion, the materials can be either inorganic or organic. *Low-field* devices comprise conventional light-emitting diodes (LEDs), which are constructed from monocrystalline semiconducting materials.

Organic semiconductors are used in numerous devices, including field-effect transistors, light-emitting diodes, Schottky diodes, photovoltaics (solar cells), light-emitting electroluminescent devices, etc. Organic semiconductor devices have been studied as unipolar devices based on hole-transport materials, because of the lack of suitable electron-transport materials. Conventional semiconductor LEDs are made by growing a region of *n*-type semiconductor material which has been doped with donors and *p*-type semiconductor materials doped with acceptors.

Organic Semiconductor

Organic semiconductors can be small molecules or polymers as employed in the case of

electroluminescent molecules for emissive displays. Hole-transport materials include oligo- and polythiophenes, poly(thienylene vinylenes), and pentacenes. Electron-transport materials include copper phthalocyanine, hexadecafluoro copper phthalocyanine, naphthalenetetracarboxy dianhydride, and perylenetetracarboxy dianhydride.

Solar Cells

Solar cells have enjoyed much interest, especially in recent times. Attempts to mimic photosynthesis using chemicals closely related to chlorophyll such as phthalocyanine constitutes have received interest. However, the efficiency of most solar cells, including those based on silicon, is very low and requires improvement in order for solar cells to make a real impact.

Most recent research is focused on dye-sensitized solar cells (DSSCs) which provide both technically and economically viable alternatives to present-day photovoltaic devices. In DSSCs, the two processes of light absorption and charge carrier transport are separated, contrary to conventional systems. Light is absorbed by the *dye sensitizer* which is anchored to the surface of a wide-bandgap semiconductor. Charge separation takes place at the interface via photo-induced electron injection from the dye into the conduction band of the solid. Carriers are transported in the conduction band of the semiconductor to the *charge collector*. Relatively inexpensive semiconducting materials are used, such as titanium dioxide, zinc oxide, and tin oxide; various dyes are used.

DSSC technology has been used as an energy provider for devices using electron chromic cells in smart windows and display.

Nonlinear Optical Dyes

Nonlinear optics (NLO) is concerned with the interaction of electromagnetic radiation with various media to produce new radiation that is altered in phase, frequency, and amplitude from the incident radiation. Electrooptical activity, applying a field to change the refractive index of a material, means that the incoming light can be manipulated from one path to another.

While there are several types of nonlinear effects, the most important is *second-order frequency doubling*, whereby the incident radiation is converted to radiation of double the frequency; this has many applications in both telecommunications and optical data storage. For example, since in telecommunications the most efficient way to transmit data is via infrared laser radiation, consequently, an important use of nonlinear optical materials would be to convert the infrared radiation to visible radiation by frequency doubling, thus enabling easier detection of the signals. An example in optical data storage is that frequency doubling would enable more data to be stored per unit area because of the smaller spot size of the frequency-doubled radiation. Molecules capable of high polarizability give the best NLO effects; in this context, organic molecules having powerful donor and acceptor groups conjugated through a delocalized π -electron system give the largest second-order effects.

Laser Dyes

Although both solid-state and semiconducting lasers have made many inroads into applications that previously were the domain of dye lasers, nevertheless, dye lasers continue to offer several advantages:

- Tunability over a wide range of wavelengths from <400 to 1,000 nm
- Hyperfine tuning
- High average power in both pulsed and continuous modes
- High feed power
- Ideal light source for the generation of short pulses

Dyes have the important role in dye lasers of allowing a fixed wavelength of laser input to be tuned to a wide range from which a selection can be made appropriate for the desired end use. Modern laser dyes have changed little from those used a decade or so ago; coumarin and xanthene colorants are the two most important colorant types. Recently, dye lasers operating in the spectral region of 600–750 nm have been of particular interest as they have found applications in

photodynamic therapy of cancer, medical diagnostics, surgery, dermatology, urology, fluorescent immunoassays, laser isotope separation, and environmental monitoring.

Biomedical Applications

Dyes have a long history of use in biomedical fields both for diagnostic and therapeutic purposes. Many dyes were used for staining bacteria and azo dyes can selectively detect metal cations. Recently, attention has focused on two interrelated aspects of *photomedicine*, namely, *phototherapy* and *photodiagnosis*. Phototherapy concerns the use of ultraviolet, visible white light, or near-infrared radiation to treat a variety of diseases, while photodiagnosis relates to the use of optical methods, based on fluorescence, for early diagnosis.

A vast range of organic chromophoric ring systems exhibit fluorescence. Fluorescence techniques are commonly used in biological, medical, and drug development. As such, fluorescent labels and probes have been employed in immunoassay, fluorescence microscopy, DNA sequencing, detection on arrays, flow cytometry, gel based quantification of nucleic acids and proteins, capillary electrophoresis, and single-molecule detection.

The *fluorophores* used in these analytical and biological outlets are based on cyanines, coumarins, naphthalimides, anthraquinones, acridines, etc.; however, such colorants have to be modified in order to be suitable for labeling purposes, DNA fluorescence, etc. Cyanines, in particular, are the most popular and extensively researched colorant type for use in biological and medical imaging protocols owing to their excellent synthetic flexibility and fluorescence properties. Indeed, cyanines are the only synthetic, near-infrared fluorophores which are commercially available for use in this area.

Smart Textiles with Photochromic Dyes

A photochromic compound is characterized by its ability to undergo a reversible color change.

Many photochromic organic compounds have been found and widely investigated in various fields. Spiropyrans and the closely related spirooxazines are important classes of photochromic dyes. While these two classes of compound are similar in many respects, the replacement of the benzopyran ring with a naphthoxazine ring results in spirooxazines having the advantage of greatly improved resistance to prolonged UV irradiation, which confers greater commercial importance to them. Applying photochromic dyes to a fabric can impart to the fabric “smart” fashion effects as well as enhanced UV protection. In this context, the application of spirooxazines to textiles has attracted research interest [7–11]. When UV light was applied to spirooxazine dyeings, photochromic reactions were mainly observed at 610 nm in which the color intensity of the dyeings increased with irradiation time. The color-produced photomerocyanine form of the colorant is usually unstable and returns to the original spiro form. The color strength (K/S) of the color-produced spirooxazine dyeings decreased with increasing storage.

Cross-References

- [Colorant, Nonlinear Optical](#)
- [Colorant, Photochromic](#)
- [Colorant, Thermochromic](#)
- [Dye, Photodynamic Therapy](#)

References

1. Hunger, K. (ed.): *Industrial Dyes: Chemistry, Properties, Applications*. Wiley-VCH, Weinheim (2004)
2. Bamfield, P. (ed.): *Chromic Phenomena: Technological Applications of Colour Chemistry*. The Royal Society of Chemistry, Cambridge (2001)
3. Waring, D.R., Hallas, G. (eds.): *The Chemistry and Application of Dyes*. Plenum, New York (1990)
4. Kim, S.H. (ed.): *Functional Dyes*. Elsevier, Amsterdam (2006)
5. Gregory, P.: *High-Technology Applications of Organic Colorants*. Plenum Press, New York (1991)
6. Yoshida, Z., Kitao, T.: *Chemistry of Functional Days*. Mita Press, Tokyo (1989)
7. El-Shishtawy, R.M.: Functional dyes, and some hi-tech applications. *Int. J. Photoenergy Article ID 434897*, **2009**, 21 pp (2009)
8. Durr, H. (ed.): *Photochromism: Molecules and Systems*. Elsevier, Amsterdam (2003)
9. Crano, J.C., Guglielmetti, R.J. (eds.): *Organic Photochromic and Thermochromic Compounds*. Plenum, New York (1999)
10. Lee, S.J., Son, Y.A., Suh, H.J., Lee, D.N., Kim, S.H.: *Dyes Pigments* **69**, 18–21 (2006)
11. Son, Y.A., Park, Y.M., Park, S.Y., Shin, C.J.: *Dyes Pigments* **73**, 76–80 (2007)

Dye, Ink-Jet

Algy Kazlauciusas

Department of Color Science, University of Leeds, Leeds, West Yorkshire, UK

Definition

An inkjet dye is a solvent-soluble colorant present in inkjet ink formulations to produce text and photo-realistic images for drop-on-demand piezo and thermal deskjet printers, signage for continuous industrial inkjet printers, and all manner of designs for piezo and thermal inkjet textile printers.

Principal Applications

Dyes used for inkjet printing can be placed into three categories:

- (a) Deskjet photo printing (A4 and A3 formats) of photo-realistic images, using drop-on-demand piezo and thermal inkjet systems

Paper substrate materials containing ink-specific ink-receivable layers are used, with both dye and pigment colorants being utilized.

Wide-format printing of images onto paper substrates (Giclée prints), using *Iris Graphics* printers, was the dominant method for producing large format photo-realistic fine art prints during the 1980s and early 1990s. These

printers employed dye-based ink formulations to produce images. However, these printers have long been replaced by a range of wide-format printers capable of producing images on a range of substrate materials, including paper, poly(vinyl chloride), and poly(propylene). This new breed of printer type uses pigments as the colorant in all ink formulation types, i.e., water-based inks, solvent-based inks, and UV curing inks. The dyes which previously had been employed for large format images, especially for outdoor purposes, were replaced with pigments because of the inferior light fastness of the dye prints.

(b) Industrial inkjet printing

This area is associated with continuous inkjet printing that is used for high-definition coding on a range of substrates, bar codes, micro printing, logos, special characters and graphics. For this type of inkjet printing, solvent dyes are heavily utilized. In addition, this category also encompasses piezo micro drop-on-demand and large character, valve jet technologies. For these latter types of printing, highly dispersed pigments are favored.

(c) Printing of fabrics

Both dyes and pigments are used for the printing of textiles, with reactive dyes primarily used for cotton, viscose, and other cellulosic fabrics; acid dyes for silk, polyamide, and wool; disperse dyes for polyester; and pigments for a wide range of different fabrics.

Deskjet Photo Printing

In the early days of deskjet printing, prior to the advent of photo-realistic imaging, both thermal and piezo drop-on-demand printers used aqueous dye-based inks only. The first-generation dyes utilized were commercially available direct and acid dyes produced for the textile market as well as food dyes [1–5]. As prints produced using vibrant colors, as opposed to dull colors, are far more attractive to the average observer, it is no surprise that the early dyes used were chosen primarily for their vibrancy. Typical dyes chosen were C.I. Acid Yellow 23 (monoazo), commonly

known as tartrazine, C.I. Acid Red 52 (xanthene), and C.I. Acid Blue 9 (triphenylmethane) with C.I. Food Black 2 (disazo) chosen as the black dye [6]. Unfortunately the vibrancy of these dyes was seriously offset by the poor light fastness of the ensuing prints. Subsequent dye introductions included azo H-acid dyes such as C.I. Reactive Red 180 for the magenta component, the phthalocyanine dye C.I. Direct Blue 199 for the cyan, C.I. Direct Yellow 132 as an alternative yellow, and metalized C.I. Reactive Black 31 for the black component. It should also be noted that for a number of deskjet ink sets, aimed at the photo-realistic imaging market, C.I. Pigment Black 7 (carbon black) was utilized instead of a black dye.

At the time, deficiencies in the inks including limited water solubility and insoluble impurities such as salts were addressed by replacing strongly acidic sulfonic acid groups with weakly acidic carboxylic acid groups so as to give high solubility in alkaline inks and by the use of dialysis/ultrafiltration purification techniques to remove insoluble impurities. At this point in time, the key properties required of inkjet dyes for photo-realistic imaging are brilliant/vibrant colors; high resolution; low color bleed (ideally none); high color strength; 5–20 % solubility; very low salt content (chloride, sulfate, calcium, and iron all less than 50 ppm); good fastness to light, water, and smear; thermal and chemical stability; high surface tension; nontoxicity; and good storage stability at extremes of temperature.

Further key aspects in the drive to rival silver halide images were the introduction of light magenta and light cyan inks to improve skin tones and the gradual development over time of ink-substrate compatibility using ink-receivable layers, namely, swellable polymer (based on polyvinyl alcohol, gelatin, etc.) and pigmented microporous (silica/alumina), coated onto a conventional photographic resin-coated paper base [7–9].

A further development in the late 1990s was the introduction by Epson of the first pigment-based desktop inkjet printer. While this was a further attempt to improve the light fastness of photo-realistic images, other image attributes

were sacrificed, namely, a reduced color gamut, the propensity to bronzing, and metamerism problems.

Currently deskjet, photo-realistic, inkjet printing principally uses anionic water-soluble dyes, incorporating sulfonic/carboxylic acid groups to ensure solvent solubility [10, 11]. With respect to Colour Index dye classification, these include acid dyes, direct dyes, and passivated reactive dyes [12]. The general approach, when used with the more commonly available microporous ink-receivable layer, is that the anionic dyes are attracted to the cationic polymers that modify the surface of the silica/alumina particles.

Table 1 highlights the more commonly used water-soluble dyes that are commercially available for photo-realistic deskjet printers [13, 14]. The table has been assembled with the aid of information gleaned from the Society of

Dyers and Colourist's Colour Index and the web sites www.clariant.com, www.pigments.clariant.com, www.nip.clariant.com, and www.dyes.com.

Although pigment-ink photo-realistic printers now tend to dominate the A4/A3 market, due mostly to their light fastness, many photographers still prefer dye-ink prints rather than equivalent pigment-ink and conventional silver halide prints. When Clariant launched their Claria dye-based ink set in 2007, this revitalized the demand for dye-based photo-realistic images, the dyes [15–17] displaying light fastness ratings of at least 6 [18]. A comprehensive survey in 2008 (Zhang) revealed the preference of dye-based prints in the key performance areas of color gamut, image resolution, and image contrast. Although dye-based inkjet prints were also perceived to be superior to pigment-based inkjet prints, regarding skin tone representation, they

D

Dye, Ink-Jet, Table 1 Inkjet water-soluble dyes

Colour Index reference	Chemical class	Light fastness range	Additional information	C.I constitution number
C.I. Acid Yellow 3	Quinoline	3	Shading	47005
C.I. Acid Yellow 23	Monoazo	4–5	–	19140
C.I. Direct Yellow 86	Disazo	5	–	–
C.I. Direct Yellow 132	Disazo	4–5	High performance images	–
C.I. Direct Yellow 169	Disazo	4–5	–	–
C.I. Reactive Yellow 37	Monoazo	4	–	–
C. I. Reactive Red 23	Azo Cu complex	6	High performance images	16202
C. I. Reactive Red 24:1	Monoazo	3–4	–	–
C.I. Reactive Red 180	Monoazo	4	–	181055
C.I. Acid Red 52	Xanthene	2	Shading	45100
Undisclosed	Azo Cu complex			
Reactive magenta from BASF	6	–	–	
C.I. Acid Blue 9	Triarylamine	2–3	Shading	42090
C.I. Direct Blue 199	Phthalocyanine	7	High performance images	74190
C.I. Acid Black 94	Azo/1:2 Cr-complex	5–6	–	–
C.I. Direct Black 168	Trisazo	5	–	–
Undisclosed	Azo Cu-complex black from Clariant	5–6	High performance images	–
Undisclosed	Polyazo black from Clariant	5	–	–

n.b. Fastness properties according to DIN ISO 12040 using an aqueous-based ink-jet ink containing 4 % black dye or 2.5 % color dye onto plain paper

Clariant Brochure – products for nonimpact printing/colorants for water-based ink-jet inks

Dye, Ink-Jet, Table 2 Inkjet solvent dyes

Colour Index reference	Chemical class	Light fastness range	C.I. constitution number
C.I. Solvent Black 27	Azo Cr-complex	7	–
C.I. Solvent Black 28	Azo Cr-complex	7	–
C.I. Solvent Black 29	Azo Cr-complex	7	–
C.I. Solvent Black 45	Azo 1:2 Cr-complex	7	–
C.I. Solvent Blue 44	Phthalocyanine	6	–
C.I. Solvent Red 91	Azo 1:2 Co-complex	6	–
C.I. Solvent Red 127	Azo 1:2 Cr-complex	4	–
C.I. Solvent Yellow 83:1	Azo 1:2 Cr-complex	7–8	–

n.b. Fastness properties according to DIN ISO 12040 using a print containing 5 % dye in vinyl copolymer, applied at 2–3 g/m² on aluminum foil

Clariant Brochure – products for nonimpact printing/colorants for solvent-based inkjet inks

were perceived to be inferior to conventional silver halide prints [19].

A review of recent patent literature indicates ongoing research activity with respect to improved performance for dye colorants used for printed images, in the areas of improved image permanence on porous photo media [20,21], or dye admixtures to extend chroma [22].

Industrial Inkjet Printing

Industrial inkjet printing is primarily utilized on production lines, or in manufacturing processes. Industrial inkjet printers can be broadly classified as continuous or drop on demand, although there are variants within each of these individual classifications.

Continuous printing technology [23] grew on the back of legislative requirements for product identification, i.e., product and data coding. Thus, this type of inkjet technology is commonly used to print the date, together with 1D and 2D bar codes, onto a range of different substrates, e.g., cartons, bottles, and cans. As this type of printer is used on moving production lines, the inks must be able to dry quickly; this is achieved using volatile solvent-based inks. Although the original inks employed were methyl ethyl ketone (MEK)-based, the drive for the minimization of volatile organic compounds (VOC's) has resulted in the introduction of solvent mixtures such as acetone in combination with lactate esters [24].

Although solvent dyes are predominantly used in solvent ink formulations, a number of black inks using carbon black pigment have been developed, specifically for bar code-OCR readability. A key advantage of using dyes for solvent-based inks is the fact that they are able to provide their own conductivity, provided either by a counter ion or via existing residual salts. In addition, dyes also have the advantage over pigments that they are present as a solution rather than as a dispersion. This means that the stability of the ink generally is superior to an equivalent pigment-based ink, both in storage and when in use in the printhead.

Table 2 shows the most commonly used solvent dyes employed for industrial continuous inkjet printing on a range of different substrate materials. The table has been assembled with the help of information provided by the Society of Dyers and Colourist's Colour Index and a selection of web sites, i.e., www.nip.clariant.com, www.pigments.clariant.com, and www.dyes.com.

Inkjet Textile Printing

Whereas the 1990s were key years for the development of both deskjet and wide-format inkjet printing with respect to photo-quality images onto IRL-coated paper substrates [25], inkjet textile printing did not become a widespread commercial proposition until 2003, when several inkjet industrial printers were launched at ITMA (International Textile Machinery Association), Birmingham, UK. One key aspect of these

printers was the introduction of higher-resolution printheads.

While the market for digital textile printing continues to grow, the pace of growth is not dramatic, with digital printing, accounting for only 1 % of the total output of textile printing technologies [26]. However, inkjet textile printing is suitable for short runs, small batches, and “just in time” production, making it ideal for upmarket textile items such as luxury home furnishings, fashion dress, and fashion accessories. As Italy is acknowledged as the major area for the production of high-end textile printing in Europe, a high percentage of commercial digital inkjet systems are located within this region. Although major technological improvements have been claimed since ITMA 2007 in Munich, further improvements are required, namely, penetration of colorants into the textile substrate, a wider color gamut, the development of a universal ink for all textile substrates, specialty printing, and greater print production speeds [27].

The desired requirements for dye-based inks utilized in digital inkjet printing on textiles are not dissimilar to the requirements needed for inkjet photo-realistic imaging onto ink-receivable layer-coated papers. Key requirements are as follows:

- (a) Good optical density
- (b) Narrow absorption curve
- (c) Minimal color to color bleed
- (d) Good light fastness
- (e) Good water fastness
- (f) High solvent solubility
- (g) Minor levels of inorganic salts (less than 100 ppm chloride and sulfate anions)
- (h) Thermally and chemically stable

It is common practice that inkjet printing onto textile fabrics, using reactive and acid dyes, generally involves pretreatment and posttreatment to ensure dye fixation onto the fabric. There are a number of important reasons as to why the dyes are kept apart from the thickeners and other required chemicals, prior to fabric application:

- (a) Many of the chemicals required for the process could cause corrosion of the inkjet

printhead and/or instigate stability problems in the ink.

- (b) The rheological properties of the thickeners used are incompatible with the printheads
- (c) Reactive dyes may hydrolyze if alkali is present in the ink formulation.
- (d) Large levels of salt in aqueous inks would not only contribute to print-head corrosion but also reduce the level of dye solubility.

Water-based dye inks tend to be specifically used for piezo-electric and thermal printheads and usually contain a glycol solvent as a humectant. As previously mentioned, reactive dyes are used in ink formulations specifically prepared for the printing of cotton, viscose and other cellulosic fabrics. For wool, silk, and nylon fabrics, acid dyes tend to be the norm, whereas for polyester fabrics disperse dyes are used. Huntsman with their Novacron[®] reactive dyes, Lanoset[®] acid dyes, and Terasil[®] disperse dyes are a major supplier of digital inkjet dye-based inks. Other key companies involved in producing these inks are Dupont, with their Artistri range of dye-based inks, along with DyStar, Kiian, Sensient Technologies, and Kolorjet.

To ensure dye fixation, steaming under atmospheric pressure at just over 100 °C is required for both acid and reactive dyes. However, for the fixation of disperse dyes, a higher temperature of 170–180 °C is required, although this can be lowered to between 130 and 150 °C if pressure steaming is employed.

Other less popular methods of dye-based inkjet printing on textiles are still in use, namely, solvent-based, oil-based, and phase change. Hot-melt inks are involved in the phase change method, with the inks being solid at room temperature and fluid when heated to between 60 and 125 °C. Piezo printers tend to be used for this method of printing, and the ink vehicle is a mixture of fatty carboxylic acids and alcohol. Solvent dyes utilized are soluble in the vehicle and hence hydrophobic.

A wide range of research activity within the field of digital inkjet printing onto textile fabrics continues to be published, indicating a healthy future. The research areas involved include the

development of both new and modified reactive dye structures for inkjet printing [28–30], fabric pretreatment [31, 32] and thickeners [33], development of inks using natural dyes [34], dye aggregation/interaction with water-soluble polymers [35], and new azo pyridone disperse dyes [36].

Cross-References

- [Colorant, Textile](#)
- [Coloration, Fastness](#)
- [Dye](#)
- [Dye, Metal Complex](#)

References

1. Kenyon, R.W.: Ink jet printing. In: Gregory, P. (ed.) *Chemistry and Technology of Printing and Imaging Systems*, pp. 113–138. Blackie Academic and Professional, Glasgow (1996)
2. Wnek, W.J., Andreottola, M.A., Doll, P.F., Kelly, S. M.: Ink jet ink technology. In: Diamond, A.S., Weiss, D.S. (eds.) *Handbook of Imaging Materials*, 2nd edn, pp. 531–602. Marcel Dekker, New York (2002)
3. Bamfield, P., Hutchings, M.G.: Phenomena involving the absorption and reflection of light. In: *Chromic Phenomena, Technological Applications of Colour Chemistry*, 2nd edn., pp. 141–233. RSC Publishing, Cambridge (2010)
4. Bauer, W., Ritter, J.: Tailoring dyes for ink-jet applications. *Surf. Coat. Int.* **9**, 407–410 (1996)
5. Gregory, P.: Ink-jet printing. In: *High-Technology Applications of Organic Colorants*, pp. 175–205. Plenum Press, New York (1991)
6. Gregory, P.: Ink jet printing: joining the jet set. *Proc. RMS* **36**, 232–238 (2001)
7. Kazlauciusas, A.: Longevity issues associated with photo-realistic ink jet images. Paper presented at the 3rd international conference on preservation and conservation issues related to digital printing and digital photography, Institute of Physics, London, 24–25 Apr 2006
8. Kazlauciusas, A.: Photorealistic ink-jet digital printing – factors influencing image quality, image stability and print durability. *Color. Technol.* **126**, 135–144 (2010)
9. Frenkel, M.: Tailoring substrates for inkjet printing. In: Magdassa, S. (ed.) *The Chemistry of Inkjet Inks*, pp. 73–122. World Scientific Publishing, Singapore (2010)
10. Gregory, P.: Colouring agents for non-impact printing – a survey. *Surf. Coat. Int. B Coat. Trans.* **85**, 9–17 (2002)
11. McFall, P.: Ink technologies for inkjet printing. *Surf. Coat. Int. A* **10**, 420–426 (2004)
12. Soffel, J.L., Morris, P.C.: US6458195. Hewlett Packard Company (2002)
13. Carr, K.: Dyes for inkjet printing. In: Freeman, H.S., Peters, T. (eds.) *Colorants for on-Textile Applications*, pp. 1–34. Elsevier, Amsterdam (2000)
14. Fryberg, M.: Dyes for ink-jet printing. *Rev. Prog. Color.* **35**, 1–30 (2005)
15. Oki, Y., Kitamura, K.: EP 1700889, Seiko Epson Corporation (2006)
16. Kitamura, K., Fukumota, H., Sao, A.: US20070263055, Seiko Epson Corporation (2007)
17. Fukumota, H., Kitamura, K.: US20070263056, Seiko Epson Corporation (2007)
18. Lei, Z.: Analysis and characterisation of inkjet dye-ink formulations, together with the longevity performance of various dye/ink substrate combinations. BSc Research Dissertation, Leeds University (2008)
19. Kazlauciusas, A., Zhang, M., Ball, J.: Investigation into digital print hard copy quality, longevity and durability. *J. Phys. Conf. Ser.* **231**, 012012 (2010)
20. Holloway, A.P., Sisk, M.W., Assignee, A.K.: US20110014375, Lexmark International (2011)
21. Holloway, A.P., Sisk, M.W., Assignee, A.K.: US20110012968, Lexmark International (2011)
22. Jackson, C., Kung, H.K., Valentine, J.E., Jackson, C., Kung, K.H.: WO 2009076457 A1, E.I. Du Pont de Nemours & Company (2010)
23. Kipphan, H.: *Handbook of Print Media*. Springer, Heidelberg (2001)
24. Samuel, J., Edwards, P.: Solvent-Based Inkjet Inks. In: Magdassi, S. (ed.) *The Chemistry of Inkjet Inks*, pp. 141–159. World Scientific, Singapore (2010)
25. Pond, S.: *Inkjet Technology*. Torrey Pines Research, California (2000)
26. Inkjet textile printing status report 2010. Paper presented at the Digital Inkjet Textile Seminar, Hangzhou (2010)
27. Ujiie, H.: State of art of inkjet textile printing – status report 2012. Paper presented at the 28th international conference on digital printing technologies, Quebec City, 9–13 Sept 2012
28. Clark, M., Yang, K., Lewis, D.M.: Modified 2,4-difluoro-5-chloro-pyrimidine dyes and their application in ink-jet printing on wool fabrics. *Color. Technol.* **125**, 184–190 (2009)
29. Maheshwari, A. C.: IN 2009MU00138 (2010)
30. Roentgen, G., Fekete, L., Nicollet, M.: WO 2012136428. Huntsman advanced materials (2012)
31. Zhu, Q., Cao, J., Wei, W., Zhong, J., Yao, J., Ye, Y., Yang, X.: Effects of the cotton fabric pretreatment on application properties of digital inkjet printing with reactive dyes. *Adv. Mater. Res.* **331**, 398–401 (2011)
32. Liao, S.-K., Chen, H.-Y.: Ink-jet printing of nylon fabric using reactive dyestuff. *Color. Technol.* **127**, 390–395 (2011)

33. Yuen, C.W.M., Ku, S.K.A., Kan, C.W.: Use of a bio-material as a thickener for textile ink-jet printing. *J. Appl. Polym. Sci.* **107**, 1057–1065 (2008)
34. Palacios-Guberti, S.B.: GB 2492641 (2013)
35. Park, J.-Y., Hirata, Y., Hamada, K.: Dye aggregation and interaction of dyes with a water-soluble polymer in ink-jet ink for textiles. *Color. Technol.* **128**, 184–191 (2012)
36. Neubauer, S.: WO 2009118260 (2009)

Dye, Liquid Crystals

Vladimir G. Chigrinov

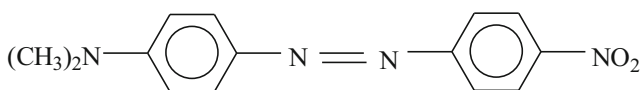
Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong

Definition

A liquid crystal is an anisotropic organic material in which component rodlike molecules assume a preferable orientation.

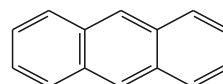
A dichroic dye is an organic molecule that has a rodlike shape and displays a unique anisotropy in which its light absorption properties occur parallel and perpendicular to the molecule, this being characterized by the dichroic ratio.

The guest–host effect is an electro-optical effect that involves adjusting dichroism to suit a specific light polarization by rotating the dichroic dye molecules within a voltage-controllable liquid crystal cell.



DNANAB is a *positive* dichroic dye or *L-type* dye (longitudinal direction of the absorbing oscillator, $\beta = 0$).

In the case of $\beta = 90^\circ$, such as is observed with anthracene molecules *T-type* dichroic dyes are obtained (transverse direction of the absorbing oscillator):



Photoaligning refers to the process by which liquid crystals on a surface are aligned as a result of their interaction with dichroic dye molecules attached to the surface.

Dichroic Dye

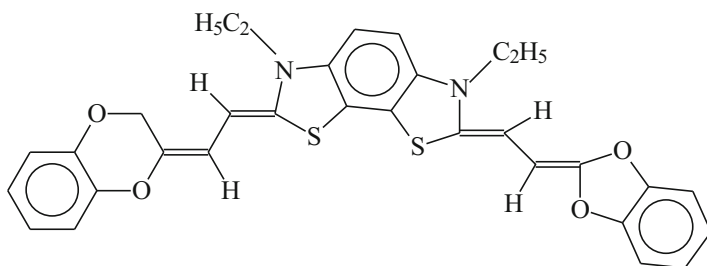
Absorption anisotropy (dichroism) in liquid crystals (LCs) may arise due to a dichroic dye (“guest”) being dissolved in a liquid crystal (“host”). Consider a molecule of the “guest” dye, whose long molecular axis is aligned along the direction of the host LC molecule. If the absorption oscillator is located at an angle β with respect to the long molecular axis of the dye, the order parameter of the dichroic dye, S_{dye} , and the dichroic ratio $= D_{\parallel}/D_{\perp}$ satisfy the following relationship shown in Eq. 1 [1]:

$$S_{\text{dye}} = \frac{(N - 1)}{(N + 2) \left[1 - \left(\frac{3}{2} \right) \sin^2 \beta \right]^{-1}} \quad (1)$$

where $D = -\log T$, where T is the relative light transmission through the LC cell.

The dichroic ratio of the dye, N , in Eq. 1 is defined as the ratio of the absorbances D_{\parallel} and D_{\perp} as determined from the polarization absorption spectra of the dichroic dye solution in the liquid crystal when oriented parallel (D_{\parallel}) and perpendicular (D_{\perp}) to the light polarization vector. An example of a dichroic dye is 4-dimethylamino-4'-nitroazobenzene (DNANAB), where $\beta = 0$ [1]:

If the dichroic dye molecule possesses simultaneously two oscillators with $\beta = 0^\circ$ and $\beta = 90^\circ$, such dichroic dyes are referred to as *L–T type*, as exemplified by bismerocyanine.



If $\beta \neq 0^\circ, 90^\circ$ then according to Eq. 1, the dichroic ratio $N = D_{\parallel}/D_{\perp}$ will decrease monotonously with β , so that

$$\begin{aligned} N = D_{\parallel}/D_{\perp} &> 1, \text{ if } \beta < \beta_m \\ N = D_{\parallel}/D_{\perp} &= 1, \text{ if } \beta = \beta_m \\ N = D_{\parallel}/D_{\perp} &< 1, \text{ if } \beta > \beta_m, \end{aligned} \quad (2)$$

where $\beta_m = \arcsin(2/3)^{1/2} = 54.7^\circ$. Dyes with $\beta \neq \beta_m$ are *L-T* types, respectively. For isotropic dyes, $\beta = \beta_m$ and $D_{\parallel}/D_{\perp} = 1$ (*I-type* dyes).

Although difficulties are encountered in developing negative dichroic dyes in the case of rodlike molecules, it has been proposed [2] that this might be resolved if each dye contains functional molecular blocks of three types:

- (i) An isolated chromophoric system with the visible light absorption oscillator located in a given direction (as exemplified by portions of dyes of different classes such as azo, anthraquinone, etc.)
- (ii) Dye solubility systems derived from alkyl (R) or substituted alkyl groups (RO-, RS-, RCOO-, RHN-, etc.)
- (iii) Order parameter improving systems, which impart a high degree of dye order parameter to the liquid crystal "host" matrix (such as aromatic, heterocyclic, cycloaliphatic, and other ring systems)

Connections between these functional blocks are made by means of linking or bridge groups such as -CH-, -CH₂CH₂-, -COO-, -O-, and others. Using this approach, a variety of *T*- and *L-T*-type azo and anthraquinone dyes have been developed

which display advantages in various practical applications [2].

Commercial dyes are characterized by the following set of parameters (Table 1) [1]:

"Guest-Host Effect"

This effect is observed in liquid crystals which have been doped with dyes. In this case, the liquid crystalline matrix (or *host*) is subjected to an electrical field, the purpose of the dye (or *guest*) being to enable the effect; the dye molecules are reoriented together with the LC molecules. Figure 1a gives a representation of a typical LC cell that has undergone homogeneous orientation. The dye is dissolved in the liquid crystal, the dye molecules are elongated in shape, and the transition dipole moment is parallel to the long molecular axis. In this case, the polarizer *P* is parallel to the initial director *L*, the absorption spectrum of the dye being characterized by the absorbance D_{\parallel} .

When a voltage which exceeds the threshold for the electro-optical effect is applied across the cell, the liquid crystal is reoriented with the director along the electric field, reorienting also the dye molecules (Fig. 1b). If the field is strong enough, a spectral characteristic change occurs in D_{\perp} , the light being polarized along the initial orientation of the director. The second (output) polarizer is not needed. In particular, the maximum value of $\Delta D = D_{\parallel} - D_{\perp}$ is ~ 1.5 and the ratio of the intensity of the light, transmitted when a field is applied (T_E) to that when there is no field (T_0), is $T_E/T_0 = 30$ [1].

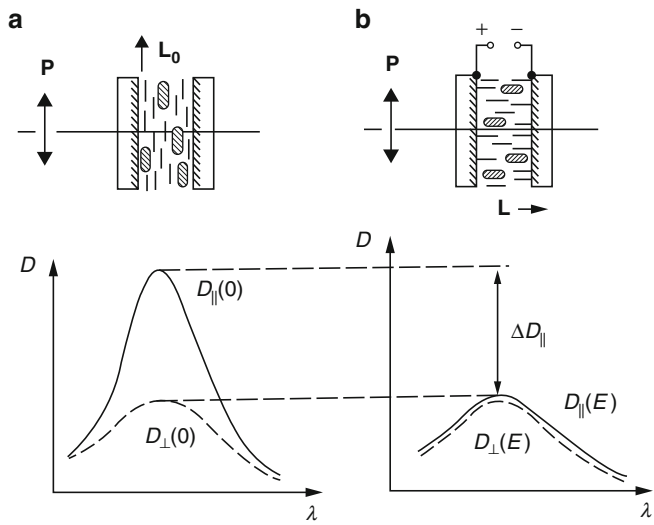
The introduction of dye molecules into the liquid crystalline host does not change the characteristic properties of the host, provided that not too much dye is introduced ($\leq 1-2\%$). Indeed, the

Dye, Liquid Crystals, Table 1 Parameters of commercial dichroic dyes

Dye	Sign of dichroism	Wavelength of maximum absorption, $\lambda_{\text{max}}/\text{nm}$	Extinction coefficient ^a $\varepsilon \times 10^4$	Dichroic ratio ^b , $N = D_{\parallel}/D_{\perp}$	Maximum solubility at room temperature % mass in ZLI-1840
KD-8	+	387	5.1	12.3	2.7
KD-9	+	450	3.5	12.3	1.8
KD-184	+	530	5.4	10.0	2.4
KD-10	+	645	1.6	11.0	2.7
D-5	+	594	1.2	5.5	5.0
D-16	+	596	1.2	6.6	2.2
	+	554	1.53	7.1	1.7
D-77	+	558	1.35	7.4	1.6
KD-208	–	468	1.3	4.2	3.0
KD-261	–	536	1.6	5.3	2.5

^aExtinction coefficient is defined as $\varepsilon = D/cd$ where $D = -\lg T$ is the absorbance (the relative transmission of the incident light through the LC cell), d the cell thickness, and c the concentration of a dye in the liquid crystal
^bDichroic ratio is calculated as D_{\parallel}/D_{\perp} for positive dichroic dyes and D_{\perp}/D_{\parallel} for negative dyes

Dye, Liquid Crystals,
Fig. 1 “Guest-host” effect in a nematic liquid crystal.
(a) No field present. (b) LC director positioned within a strong electric field



working temperature range of the liquid crystal and its viscous and elastic properties, electrical conductivity (provided the dye is nonionic and does not contain ionic impurities), dielectric permittivity (provided the dye molecule does not have a large dipole moment), and refractive indices remain the same. The only properties of the crystal which are changed are the appearance of the absorption bands in the visible region of the spectrum and viscosity, which increases slightly. Thus, the electro-optical properties of “guest–host” cells, including the controlling

voltages and response times, are virtually the same as observed for the pure LC without dye molecules.

We may distinguish the following characteristics of “guest–host” LCDs [1]:

1. The *sign of dye dichroism* – though both positive and negative dichroic dyes can be employed, positive dichroic dyes display superior characteristics and offer advantages from a chemical synthesis perspective [2] and in the development of black dichroic dye mixtures

- [1], owing to their high-order parameter and high solubility (Table 1).
2. *The type of contrast* – positive contrast corresponds to dark symbols on a bright background as well as negative-to-bright symbols on a dark background. Positive contrast is preferable especially under weak illumination [1].
 3. *The nature of the electro-optic effect employed for a “guest–host” LCD.*

Several electro-optic effects can potentially be used in “guest–host” (GH) mode, namely, *electrically controlled birefringence* (ECB), *twisted nematic* (TN) and *supertwisted nematic* (STN) effects, and electrical field-induced transitions in chiral nematic structures, such as *nematic–chiral nematic transition* ($N \Rightarrow CN$) and the corresponding reversed ($CN \Rightarrow N$) effect [1]. The various types of “guest–host” mode differ according to:

- (i) The initial LC orientation (*H*, homogeneous; *HT*, homeotropic; *QH*, quasi-homeotropic as determined using nonzero projection (L_0)_S onto the substrate plane; *T*, twisted; *ST*, supertwisted)
- (ii) The sign of the dielectric anisotropy (+ or –), which defines the final orientation of the LC molecules within the electric field; i.e., parallel or perpendicular to the electric field, accordingly
- (iii) The presence or absence of a chiral dopant in a nematic mixture (+ or –)
- (iv) The twist angle of the LC director alignment
- (v) The polarizer location parallel or perpendicular to the LC director on the input substrates (one of the options is the absence of polarizers)

The proficiency of a “guest–host” LCD system is measured in terms of, for example, the perceived contrast ratio or the ratio of the transmitted luminance in the on and off states, response times, viewing angle, etc.

A single-cell GH mode is not the only possibility insofar as similar effects can be secured when two “guest–host” cells with perpendicular director orientations are positioned together. In

this case, the initial alignment can be homogeneous (*H*), twisted (*T*), or homeotropic (*HT*). Such a device provides a negative (two *H* or *T* cells) or a positive (two *HT* cells) switching contrast for nonpolarized light. Such double-cell GH-LCD shows high brightness, wide viewing angles, and low operating voltages [1]. The superior characteristics of double-cell GH-LCD arise because the GH mode affects both polarization directions of the incident light, so no polarizers are required [1].

The quality criteria of guest–host devices in display systems are based on the following principal characteristics:

1. *The dichroic ratio* $N = D_{\parallel}/D_{\perp}$, which strongly correlates with the dye order parameter and can be improved by careful selection of the portion of the dye molecule.
2. *The perceived contrast ratio between “on” and “off” states* $C = B_{\text{on}}/B_{\text{off}}$, which is the extent of transmission obtained measured relative to the sensitivity of the human eye $y(\lambda)$ and the spectral distribution of the illumination source $H(\lambda)$ as well as the transmission of GH LC cell $T(\lambda)$

$$B = \int_{\lambda} H(\lambda) y(\lambda) T(\lambda) d\lambda / \int_{\lambda} H(\lambda) y(\lambda) d\lambda$$

$$T_{\text{on(off)}}(\lambda) = 10^{-D_{\text{on(off)}}^{(\lambda)} c d}, \quad (3)$$

where $D_{\text{on(off)}}$ is the absorbance in the “on” and “off” state, c the dye concentration, and d the layer thickness. Since $D(\lambda)$ is independent of both c and d , the perceived contrast ratio increases with increasing c and d at the cost of lowering the brightness of the GH-LCD. The perceived contrast ratio can be optimized for twisted GH displays by careful choice of the angle of the twist structure φ , the LC optical path length, and the dye concentration c [1].

3. *The color difference* (ΔE) between that of the GH-LCD in the “on” and “off” states as measured in an appropriate CIE color space [1]. This difference, which is determined

experimentally, is proportional to the number of distinct hues between the two display states [1]. In the case of CIE 1976 color space, the CIE (L^* , a^* , b^*) or CIE (L^* , u^* , v^*) coordinates are calculated according to Eq. 4 and (Fig. 2) [1]:

$$\begin{aligned} L^* &= 116\left(\frac{y}{y_0}\right)^{1/3} - 16, \\ u^* &= 13L^*(u' - u_0'), \\ v^* &= 13L^*(v' - v_0') \end{aligned}$$

(4)

where

$$u' = \frac{4x}{(x + 15y + 3z)}, \quad v' = \frac{9y}{(x + 15y + 3z)}$$

and

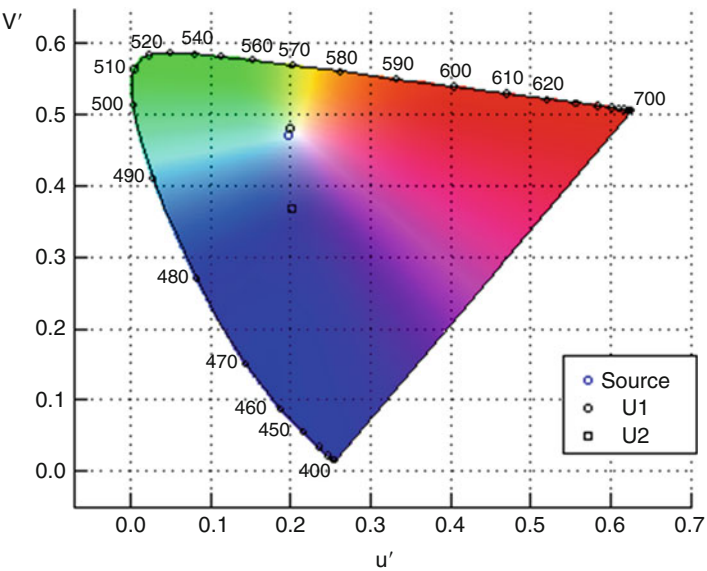
are the tristimulus values; $T(\lambda)$ the transmittance of the sample; $X(\lambda)$, $Y(\lambda)$, $Z(\lambda)$ the color-matching functions for a CIE 1931 standard colorimetric observer; $H(\lambda)$ a characteristic of the illumination source; and the index (0) the spectral power distribution of the illuminant ($T(\lambda) = 1$); the integral is calculated over the visible wavelength range ($\sim 380\text{--}720\text{ nm}$).

From the transmittance of the display in the on and off states, the CIE 1976 color difference (ΔE^*) is calculated using Eq. 5:

$$\Delta E^* = \left[(L_{\text{on}}^* - L_{\text{off}}^*)^2 + (u_{\text{on}}^* - u_{\text{off}}^*)^2 + (v_{\text{on}}^* - v_{\text{off}}^*)^2 \right]^{1/2}$$

(5)

Dye, Liquid Crystals,
Fig. 2 CIE color coordinates and color difference ($\Delta E^* = \text{DE}_{\text{Luv}}$) between two points, corresponding to two different voltages applied to a LC cell (U_1 and U_2). The corresponding image is shown by the letter “M” in the middle of the figure



	U1	U2
L^*	66.16	1.26
u'	0.2001	0.2032
v'	0.4785	0.367
DE Luv	65.67	

The color difference enables not only the luminance contrast to be ascertained but also the particular color purity and the particular hue at which the human eye is most sensitive to. An increase by a factor of three of the number of distinguished rows of an LCD is achieved using colorimetric methods [1]. This effect arises from the fact that the human eye is more sensitive to color contrast than to luminance, as both hue and chroma contribute to the former. Color differences between the “on” and “off” display states of an LCD can be optimized through the judicious choice of, for example, polarizer, concentration of dichroic dye (guest) within the liquid crystal matrix (host), etc. Colorimetric evaluation has been found to be a very useful tool in evaluating LCD quality, including GH mode [1].

Future applications of the GH mode in LCDs are related to finding novel dichroic dyes of high dichroic ratio, with good solubility in the given LC mixture and large extinction coefficient (Table 1). The quality of a particular azo dye mixture is determined by the characteristic absorption spectra of the red (R), green (G), and blue (B) dichroic dye components [1].

Azo Dye-Based Liquid Crystal Photoalignment

Photo-induced reorientation of azo dyes is a promising [3] photoalignment technique which does not involve either photochemical or structural transformations of the molecules. As the

photoaligning films are robust and possess good aligning properties, they are very useful for the generation of novel liquid crystal devices as well as new photovoltaic, optoelectronic, and photonic devices based on highly ordered, thin organic layers. Examples of such applications are light-emitting diodes (OLEDs), solar cells, optical data storage, and holographic memory devices. The novel and highly ordered layer structures of organic molecules may exhibit certain physical properties, which are similar to those of aligned LC layers.

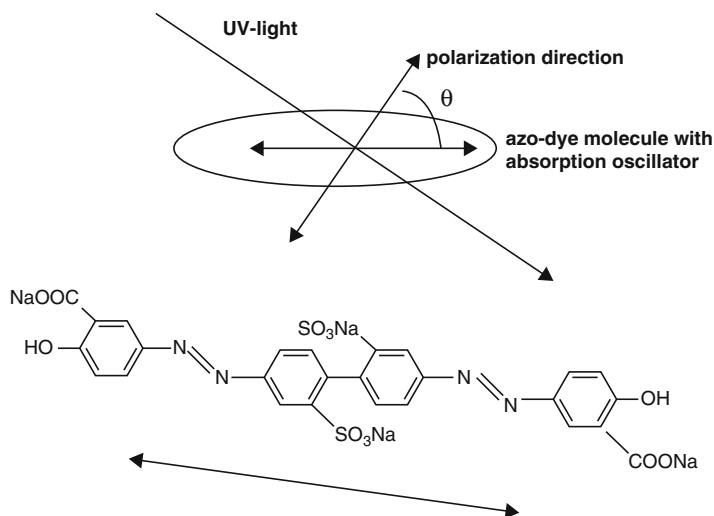
When azo dye molecules are optically pumped using a polarized light beam, the probability of light absorption is proportional to $\cos^2 \theta$, where θ is the angle between the absorption oscillator of the dye molecules and the polarization direction of the light (Fig. 3) [3].

Hence, azo dye molecules which have their absorption oscillators (chromophores) parallel to the light polarization will most probably gain energy, which results in their reorientation. This results in an excess of chromophores that are oriented in a direction in which the absorption oscillator is perpendicular to the polarization of the light.

For rodlike azo dye molecules, as represented by the sulfonated azo dye shown in Fig. 3 which have a cylindrical symmetry, the only coordinate will be the polar angle θ , between the molecular absorption oscillator and the direction of

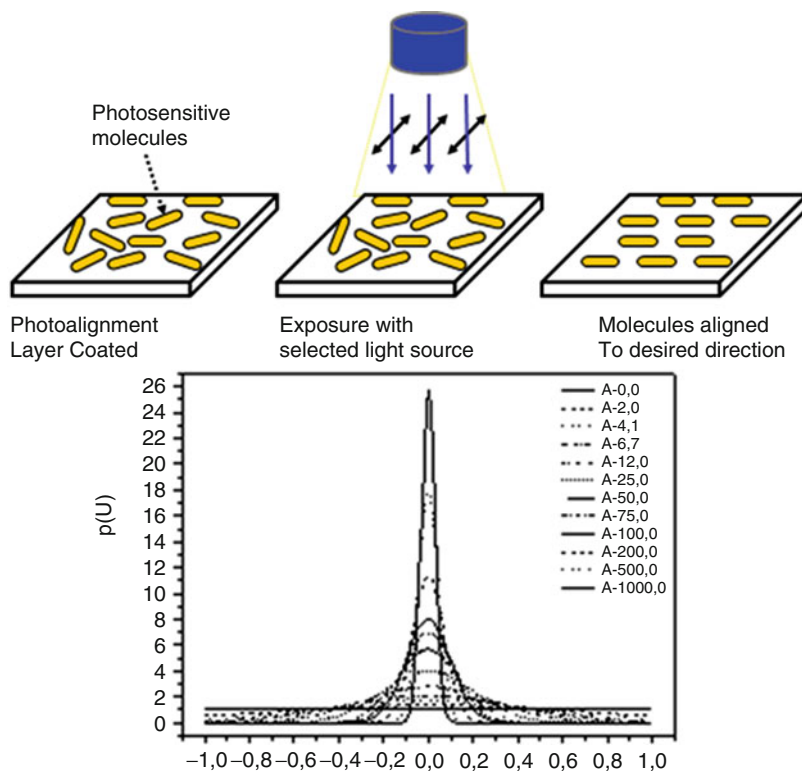
Dye, Liquid Crystals,

Fig. 3 Qualitative interpretation of the photo-induced order in photochemical stable azo dye films. *Upper* diagram shows the geometry of the effect, while the *lower* diagram reveals an azo dye molecule with its absorption oscillator (chromophore) perpendicular to the long molecular axis



Dye, Liquid Crystals,**Fig. 4** *Upper:*

reorientation of azo dye molecules perpendicular to light polarization; *lower:* the distribution function $\rho(u)$, where $u = \cos\theta$ of the azo dye molecules for various values of the parameter A , proportional to the intensity of the activated light



D

polarization of the activating light (Fig. 3). The function $f(\theta)$ represents the statistical distribution of the azo dye molecules which align along the various orientations θ , this being $f = 1/4\pi$ in the initial state to $f = \delta(\theta - \pi/2)$ after a sufficiently high exposure time (Fig. 4).

High-quality LC alignment can be obtained using photoaligned azo dye structures (Fig. 5). Photoalignment possesses obvious advantages in comparison with the usual “rubbing” treatment of the substrates of liquid crystal display (LCD) cells. Possible benefits for using this technique are considered to include [2, 3]:

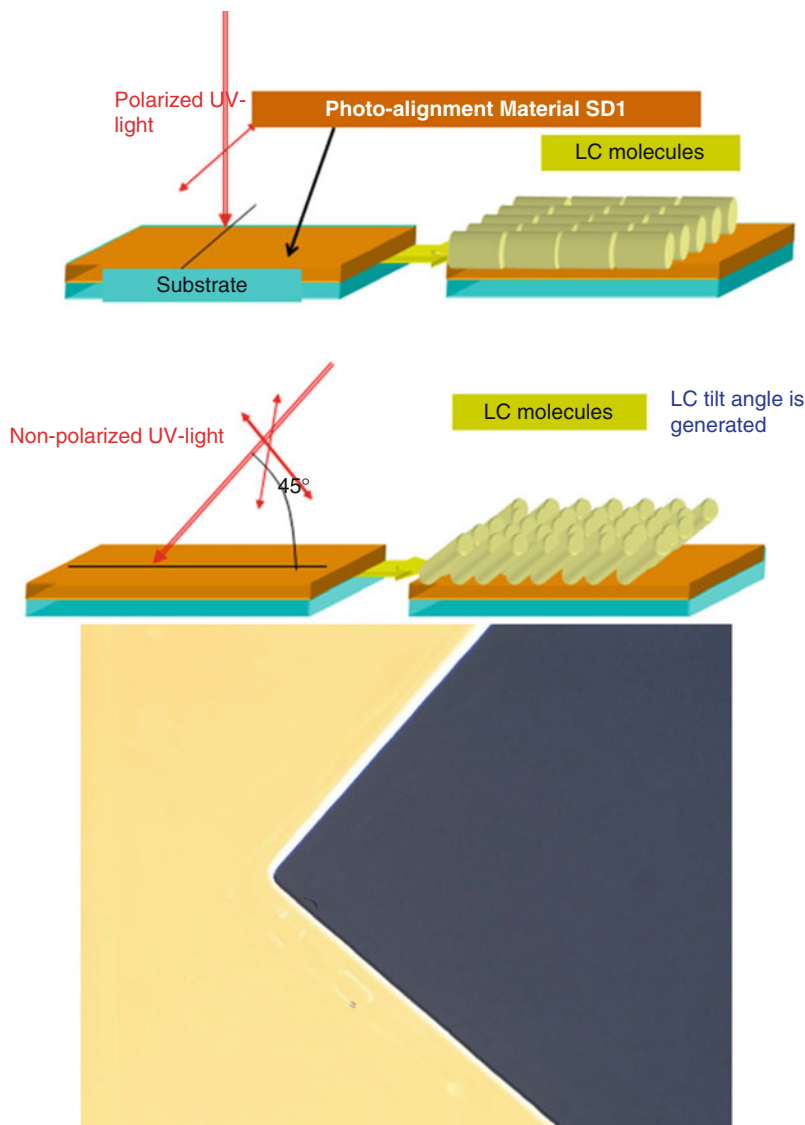
- (i) Elimination of electrostatic charges and impurities as well as mechanical damage to the surface of rubbed polymer;
- (ii) A controllable pretilt angle of the LC cell, as well as high thermal stability, high UV stability, and ionic purity;
- (iii) The possibility of producing structures with the required LC director aligned within selected areas of the cell, thus allowing the

pixel to divide to enable novel LC device configurations for transfective, multi-domain, 3D, and other new display types;

- (iv) A potential increase in manufacturing output, especially of LCDs with active matrix addressing, where the fine pixels of a high-resolution LCD screen are driven by thin-film transistors within a silicon substrate;
- (v) New advanced applications of LCs in fiber communications, optical data processing, holography, and other fields, where traditional rubbing LC alignment is not possible owing to the sophisticated geometry of a LC cell and/or high spatial resolution of the processing system;
- (vi) Efficient LC alignment on curved and flexible substrates;
- (vii) Manufacture of new optical elements for LC technology, such as patterned polarizers and phase retarders, tunable optical filters, polarization nonsensitive optical lenses, with voltage-controllable focal distance, etc.

Dye, Liquid Crystals,

Fig. 5 *Top:* LC photoalignment by azo dye layer. *Bottom:* Polarized microscope photograph of LC alignment on photoaligned film (1 mm²) for the bright and dark states distinguished by the polarizer orientation (TN LC cell)



Perhaps dye photoalignment may replace the currently more common rubbing technology employed for LCD production [3, 4]. A very high uniformity and quality of the LC surface alignment is expected in this case, which considerably improves the optical quality of the LC display.

Cross-References

- CIE 1976 L*u*v* Color Space
- Color Contrast

References

1. Chigrinov, V.G.: Liquid Crystal Devices: Physics and Applications. Boston/London, Artech-House (1999). 357 pp
2. Ivashchenko, A.V.: Dichroic Dyes for Liquid Crystal Displays. CRC Press, Boca Raton (1994). 355 pp
3. Chigrinov, V.G., Kozenkov, V.M., Kwok, H.S.: Photoalignment of Liquid Crystalline Materials: Physics and Applications. Wiley, Chichester (2008). 248 pp
4. <http://sharp-world.com/corporate/news/090916.html>

Dye, Metal Complex

J. N. Chakraborty

Department of Textile Technology, National
Institute of Technology, Jalandhar, Punjab, India

Synonyms

Premetallized acid dyes

Definition

Metal-complex dyes and pigments are the coordination complexes of bi- or polyvalent transition metal ions with various dyes and have versatile application in various fields, which include the dyeing of nylon and protein fibers, paint, toners for photocopiers, laser and ink-jet printers, photoconductors for laser printers, nonlinear optics, singlet oxygen generators, dark oxidation catalysts, and high-density memory storage devices.

Chemistry and Application

Coloration of Textiles

The most prevalent examples are monoazo (C.I. Reactive Violet 1, Fig. 1) and disazo (C.I. Reactive Blue 82, Fig. 2) dyes. While formazan dyes are bi-, tri-, and tetradentate complexes, only tetradentate copper complex formazan dyes (e.g., C.I. Reactive Blue 160, Fig. 3) have found commercial use owing to their intense colors [1].

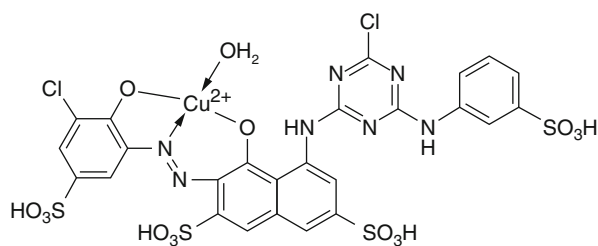
Metallization imparts a bathochromic shift producing deeper but duller shades. The spectra of 1:1 metal-complex azo dyes display narrower curves, while that of 1:2 complex dyes show a broader curve exhibiting blueness and dullness (Fig. 4).

1:1 metal-complex azo dyes are stable in strongly acidic pH and are preferably applied to wool, silk, and nylon fibers at a low pH (<4); they exhibit low to moderate coverage of fiber irregularities and poor compatibility in mixture. In contrast, 1:2 complexes are stable in weakly acidic pH and hence are applied from weakly acidic or neutral dyebaths (pH ~ 5–7); they display good

Dye, Metal Complex,

Fig. 1 C.I. Reactive

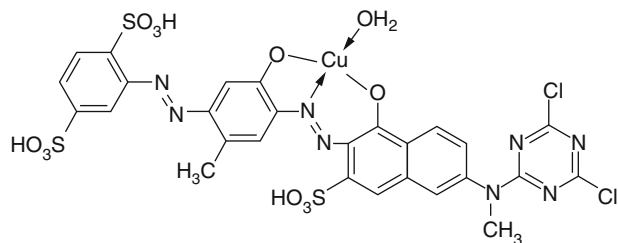
Violet 1



Dye, Metal Complex,

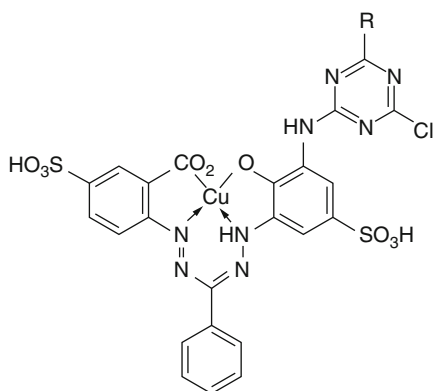
Fig. 2 C.I. Reactive

Blue 82



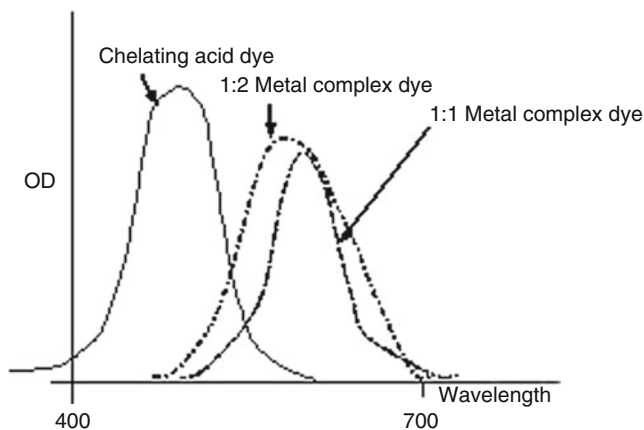
buildup on wool, silk, and nylon fibers, exhibiting high compatibility in mixtures [2]. Dyeings of chromium complex dyes display high overall fastness compared to those obtained using copper, cobalt, and nickel complexes [3]. The metal remains bonded to the lone pair with one of the azo nitrogen atoms and not to the p-bond of the azo group in 1:2 copper complex dye (Fig. 5) [4]; however, in 1:1 copper complex dyes, the symmetrical dihydroxyazo ligand (Fig. 6) forms the 1:1 square planar complex (Fig. 7) [5].

Copper phthalocyanine (Fig. 8) is a bright blue pigment of λ_{max} 660nm characteristically displaying very high all-round fastness and it exists in several polymorphic forms [6, 7]. The redder, α -metastable form is preferred for use in paint, whereas the greener, stable β -polymorph is employed in printing inks. Complete chlorination



Dye, Metal Complex, Fig. 3 C.I. Reactive Blue 160

Dye, Metal Complex, Fig. 4 Spectral change in 1:1 and 1:2 complex dye with that of the parent dye



of azo phthalocyanine blue pigment (Fig. 9) produces the high fastness green pigment, C.I. Pigment Green 7 (Fig. 10).

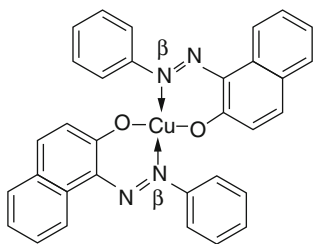
Water-soluble phthalocyanine dyes possess either sulfonic, sulfonamido, or both types of solubilizing group and are useful for the ink-jet printing of textiles. C.I. Direct Blue 199 (Fig. 11) and C.I. Reactive Blue 71 (Fig. 12) are used in the coloration of paper and cotton, respectively.

Ink-Jet Printing

Ink-jet inks are water-soluble dyes in an aqueous vehicle. Metal-complex dyes and pigments possessing high absorbance, high chroma, and excellent lightfastness are employed. The copper complex dye C.I. Reactive Black 31 (Fig. 13) is very widely used as black in almost all ink-jet printers. For ink-jet printing of black on textiles, the 1:2 complex of a chelating azo dye with a mixture of metals "M" ($M = \text{Cr}^{3+}/\text{Co}^{3+}$, 70:30) is extensively used [8], as exemplified by C.I. Reactive Black 8 (Fig. 14).

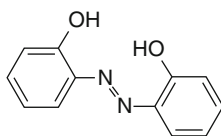
The black phthalocyanine dye (Fig. 15) was developed through reaction of amino phthalocyanine with acrylic or methacrylic acid, where "M" represents metal or hydrogen and "R" hydrogen or methyl. Solvent-soluble chromium-complexed azo black dyes, a mixture of 4- and 5-nitro, such as C.I. Solvent Black 35 (Fig. 16), is used in hot melt ink-jet printing [9].

Though yellow and magenta azo dyes are metal-free, the copper complex azo dye, C.I. Reactive Red 23 (Fig. 17), and the nickel

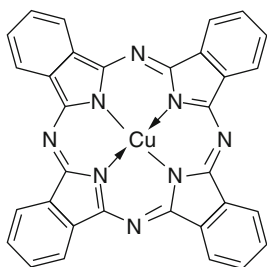
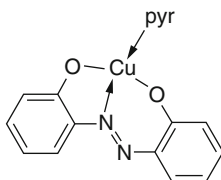


Dye, Metal Complex, Fig. 5 1:2 Copper complex azo dye

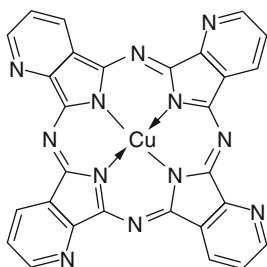
Dye, Metal Complex, Fig. 6 2-2'-Dihydroxyazobenzene (H₂DAB)



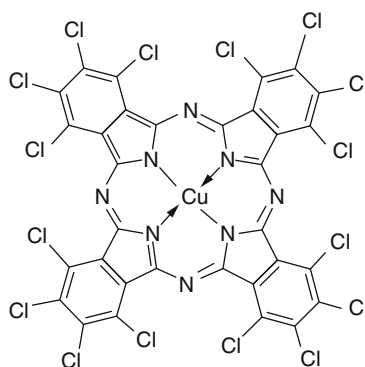
Dye, Metal Complex, Fig. 7 Copper (DAB). Pyr



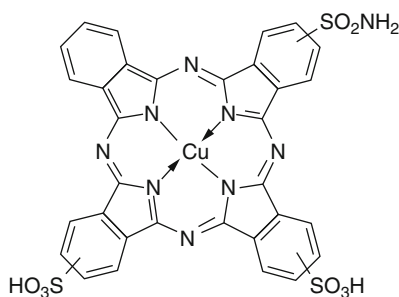
Dye, Metal Complex, Fig. 8 Copper phthalocyanine (C.I. Pigment Blue 15:4)



Dye, Metal Complex, Fig. 9 Azo phthalocyanine blue pigment



Dye, Metal Complex, Fig. 10 C.I. Pigment Green 7



Dye, Metal Complex, Fig. 11 C.I. Direct Blue 199

complex PAQ dye (Fig. 18), where R = H, Cl, or COOH and R' = SO₂NHalkyl, display dull and bright colors, respectively, each with similar high light fastness (Fig. 17) for ink-jet printers [10].

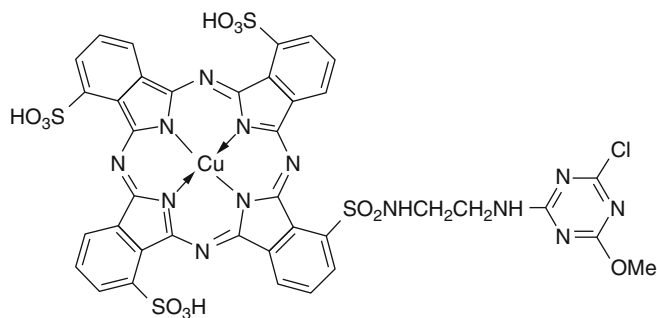
Copper phthalocyanine is extensively used as cyan for ink-jet printing; for textile ink-jet printing, metallized reactive dyes are used [8], such as C.I. Reactive Blue 71 (Fig. 12).

Photoconductors for Laser Printer

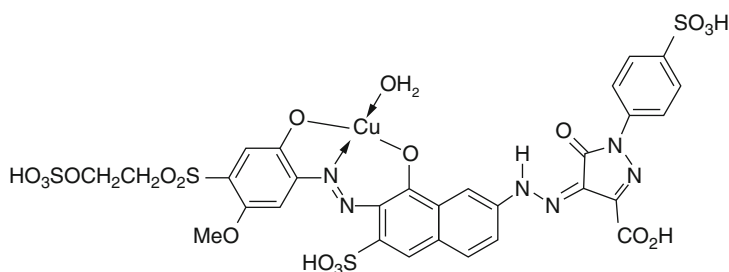
Photoconductors are dual-layer devices, and are exclusively used in laser printers. Metal-complex dyes, especially titanyloxy-phthalocyanine (Fig. 19), can act exclusively as charge generation materials (CGMs) in photoconductors [11]. All modern laser printers use titanyloxy-phthalocyanine, type IV polymorph, as the CGM possessing a “shallow shuttlecock” shape has a subtle effect on the packing of the molecules [12]. The pigment used as the photoconductor must be extremely pure and possess the correct morphology.

Dye, Metal Complex,**Fig. 12** C.I. Reactive

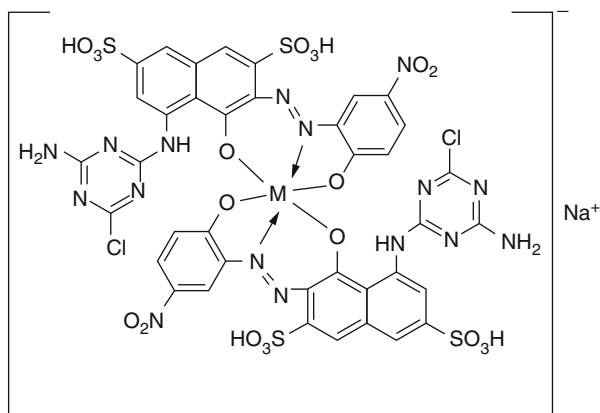
Blue 71

**Dye, Metal Complex,****Fig. 13** C.I. Reactive

Black 31

**Dye, Metal Complex,****Fig. 14** C.I. Reactive

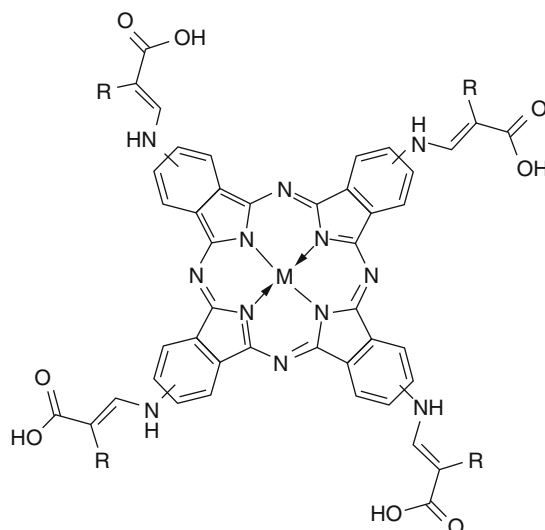
Black 8

**Toners for Photocopiers and Laser Printers**

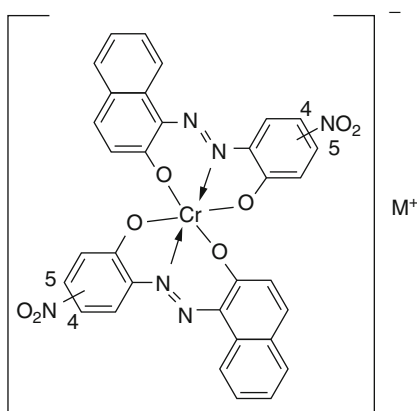
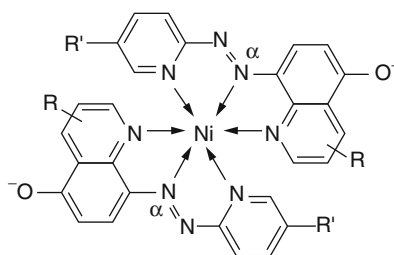
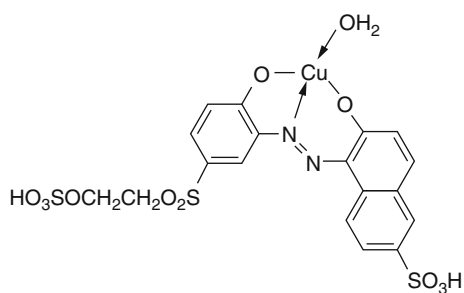
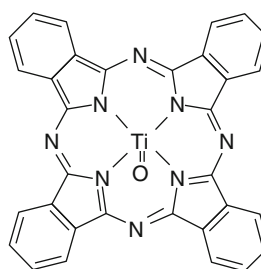
Laser printers, photocopiers, and especially ink-jet printers employ metal-complex dyes as toner. Laser printers and photocopiers use electricity and light to form an image with the help of a photoconductor; the latent electrostatic image on the photoconductor is thermally fixed using a toner [11, 12]. Metal-complex dyes and pigments produce the positive hole in the organic photoconductor and also perform key functions

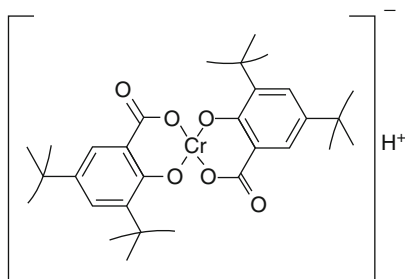
in both the image generation stage and the toner development stage of the reproduction process.

Two main types of colored molecules are used in a toner: a colorant is employed to impart color to the toner, and the charge control agent (CCA) imparts color as well as control the triboelectric charge on the toner particles. Copper phthalocyanine, C.I. Pigment Blue 15:4 (Fig. 8), is the cyan colorant; however, pigments used for the yellow and magenta colorants are metal-free. CCAs can

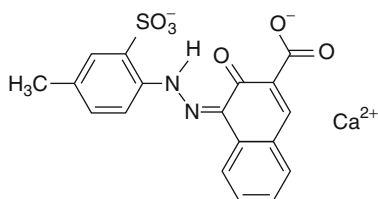
Dye, Metal Complex,**Fig. 15** Black
phthalocyanine dye

D

**Dye, Metal Complex, Fig. 16** C.I. Solvent Black 35**Dye, Metal Complex, Fig. 18** 1:2 Nickel complex PAQ
dye (PAQ: pyridyl azo Quinoline)**Dye, Metal Complex, Fig. 17** C.I. Reactive Red 23**Dye, Metal Complex, Fig. 19** Titanyloxy
phthalocyanine



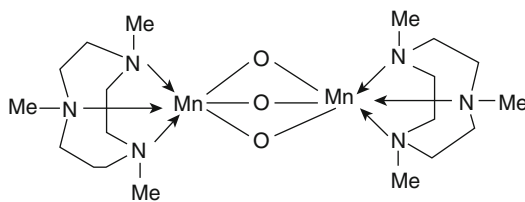
Dye, Metal Complex, Fig. 20 Bontron E-8136



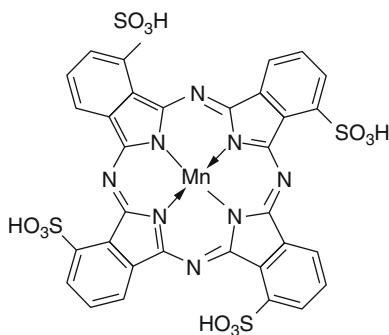
Dye, Metal Complex, Fig. 21 C.I. Pigment Red 57:1 (calcium 4B toner)

be either colored or colorless and are mainly metal-complex dyes. Colorless CCAs are useful for yellow, magenta, and cyan toners, while colored CCAs are more effective than colorless agents but are restricted to black toners. CCAs that impart a negative charge to the toner are invariably metal complexes, particularly 1:2 chromium-complexed azo dye with a mixture of 4- and 5-nitro (Fig. 16) and another one with a mixture of 4- and 5-chloro [12]. Efficient, colorless, negative-charging CCAs were developed by producing colorless analogues of 1:2 chromium complex of an aromatic *ortho*-hydroxy carboxylic acid, such as the complex of di-*tert*-butyl salicylic acid, e.g., *Bontron E-8136* (Fig. 20).

Another important class of supramolecular metal-complexed pigments is the so-called toner pigments which are water-soluble azo dyes containing sulfonic acid groups and are insolubilized by complexing with a divalent cation, such as calcium and barium. These compounds provide the standard magenta colorants for printing inks, such as C.I. Pigment Red 57:1, known as *Calcium 4B toner* (Fig. 21); chlorination of this product develops the red pigment which is suitable for characterization by synchrotron radiation.



Dye, Metal Complex, Fig. 22 Organo-manganese complex (Unilever)



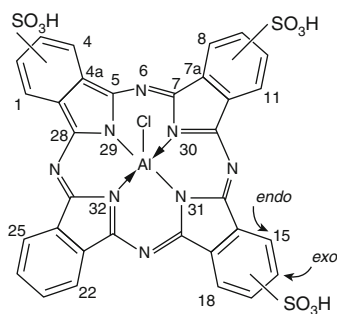
Dye, Metal Complex, Fig. 23 Manganese phthalocyanine tetra-sulfonic acid (Proctor and Gamble)

Dark Oxidation Catalysts

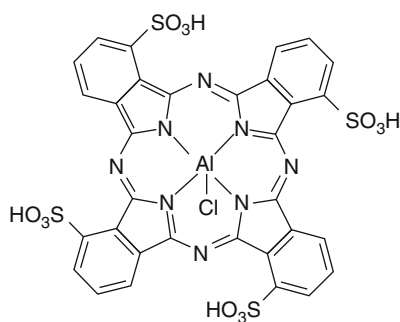
These are organometallic complexes based on transition metals possessing variable oxidation states and are used as **chemical transformers and in laundry powders** [13]. Manganese complexes were developed and marketed, respectively, by Proctor and Gamble and Unilever, as washing powders [14–16]. Unilever's manganese complex (Fig. 22) was effective in removing dirt and stains from soiled fabrics, but this was accompanied by serious damage of the fabric. Proctor and Gamble developed a similar dark oxidation catalyst, manganese phthalocyanine tetra-sulfonic acid (Fig. 23), but could not achieve commercial success because of the “manganese factor” which was probably at the root of disintegration of the fabric.

Singlet Oxygen Generators

These have multipurpose applications, such as in skin, breast, and oropharyngeal cancer treatment through photodynamic therapy (PDT) using a photosensitizer [17, 18]. In photodynamic



Dye, Metal Complex, Fig. 24 Sulfonated of chloroaluminum (III) phthalocyanine (Photosens)



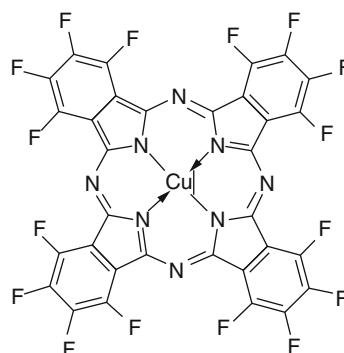
Dye, Metal Complex, Fig. 25 Tinolux BBS

therapy, electromagnetic radiation in the visible and adjacent ultraviolet and infrared regions is used for photochemical treatment of (i) jaundice (neonatal hyperbilirubinemia) in the newborn, and especially in the prematurely born, (ii) psoriasis using light in the UV-A range and an administered photosensitizer, (iii) wet form of age-related macular degeneration with a photosensitizer as well as a laser light source, and (iv) certain cancer, such as skin, breast, and oropharyngeal cancer with a photosensitizer and red light [19–24].

The phthalocyanine-based commercial product *Photosens* (Fig. 24) is a widely used photosensitizer prepared by the direct sulfonation of chloroaluminum (III) phthalocyanine. Singlet oxygen generators are also used as hard-surface disinfectants, soaps, washing powders [25], and insecticides [26]. These eco-friendly photobleaches convert triplet oxygen ($^3\text{O}_2$) into highly reactive singlet oxygen ($^1\text{O}_2$) in the presence of light, thus acting swiftly to bleach stains. The blue aluminum phthalocyanine dye, *Tinolux BBS* (Fig. 25), is a commercial singlet oxygen generator [27]; zinc phthalocyanines, which absorb in the near-infrared region at ~ 725 nm, are also very good singlet oxygen generators [28].

Organic Semiconductors

Organic semiconductors are used to fabricate electronic devices. Metal-complex pigments, such as hexa-deca-fluoro copper phthalocyanine (Fig. 26), are used for electron transport [28].



Dye, Metal Complex, Fig. 26 Hexa-deca-fluoro copper phthalocyanine

Infrared Absorbers

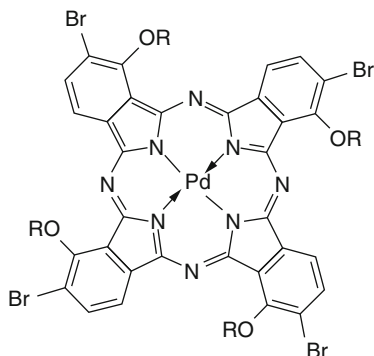
Upon complete chlorination, blue copper phthalocyanine is converted into its green-colored, fast pigment which also is an efficient infrared absorber, possessing high solubility in organic solvents and displaying narrow absorption curves at around 780 nm. Phthalocyanines containing amino donors, in combination with either thiols [29] or halogens [30], provide broader curves absorbing at around 900 nm.

Copper phthalocyanine loses its planar structure upon conversion to infrared absorber, with eight of the sterically demanding thiophenol groups lying above the plane of the phthalocyanine ring and eight lying below. Taking advantage of this feature, non-planar copper phthalocyanines are used as high-level security components in currency notes [31, 32], in computer-to-plate technology to convert near-infrared radiation into heat [33, 34] and as solar screens on car windscreens

and windows of buildings which allow daylight through but screen out infrared radiation component which generates heat [35]. However, as the durability of non-planar copper phthalocyanine is not high, palladium phthalocyanine (Fig. 27) is the leading phthalocyanine infrared absorber for CD-R applications [35, 36].

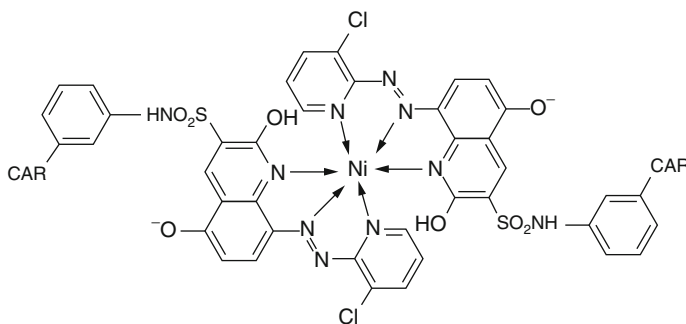
Silver Halide Photography

Metal-complex azo dyes are used in color diffusion transfer photography, employing nondiffusible, magenta dye-releasing colorants which, upon development of the silver halide layer, release a diffusible magenta dye; the nickel complex PAQ dye (Fig. 28), where “CAR” denotes Ballast-Carrier-Link, is a typical example [37]. Other metal-complex azo dyes include pyridylazonaphthols [38] and pyridylazoaminophenols [39].



Dye, Metal Complex, Fig. 27 Brominated tetra (α -2,4-dimethyl-3-pentoxo) palladium phthalocyanine

Dye, Metal Complex, Fig. 28 Nickel complex PAQ dye retaining CAR (Ballasted carrier moiety)



Nonlinear Optics

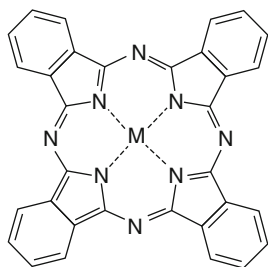
Metal complexes, mostly of δ -transition metals, can exhibit nonlinear optical (NLO) behavior which can be used to manipulate the fundamental properties of laser light beams, thereby facilitating optical data processing and storage [40–42], and offering possibilities for the combination of quadratic (second-order) and cubic (third-order) NLO properties along with other molecular electronic properties. Examples include bimetallic complexes containing pyridyl carbonyl centers with ruthenium σ -acetylide; ferrocenyl (Fc) electron donor groups can also exhibit NLO properties [43–45].

In metalloporphyrin derivatives, such as metallophthalocyanines and metallo-2,3-naphthalocyanines, metallation greatly enhances the cubic responses of the compound; such materials often exhibit strongly absorbing, long-lived triplet excited states. Asymmetrically substituted metalloporphyrin derivatives can exhibit extremely large quadratic NLO responses; the β -response of the Cu(II) complex shows a marked wavelength dependence, while its Zn(II) analogue displays essentially identical β 830 and β 1064 values [46].

Some porphyrin-like metal complexes, mostly of Cd(II), in which one of the pyrrole rings is replaced with an aryl or vinyl group, exhibit very large cubic NLO responses [47, 48].

High-Density Memory Storage

The electrochromism exhibited by metal-complex dyes in the near-infrared region (NIR) of the spectrum (ca. 800–2,000 nm), obtained by the generation of different electronic absorption bands, enjoys applications such as optical data storage



Dye, Metal Complex, Fig. 29 Metal Phthalocyanine

(ODS) on various disk media, i.e., read-only, write-once, and rewriteable media [49, 50]. In such applications, a layer of a metal-complex dye based on highly conjugated ligands displays a significant absorption maximum which is either at or very close to the wavelength of the laser light and also has relatively low optical absorbance and high reflectance at the recording wavelength and is interposed between the substrate and the reflecting layer of a CD. With appropriate optimization of the optical properties, this composition will record and reproduce [35, 51, 52].

Though polymethine, porphin, triphenylmethane, and azoic metal-complex dyes possess good photostability and absorption in the range 750–850 nm exhibiting electrochromism, phthalocyanine (Fig. 29), naphthalocyanines, and cyanine dyes based on indole and benzindole are the most widely used for the mass production of on-disk-structure (ODS) media because of the simplicity of modification, excellent photostability, and cost-effectiveness. However, selective metal-complex dyes should meet certain requirements; partial bromination allows fine-tuning of the film absorbance and reflectivity [35, 49, 53], commercially known as *Supergreen*. Phthalocyanine dyes are not suitable for DVD-R media, since the main chromophore cannot readily be modified to produce a sufficiently large hypsochromic shift; dyes potentially suitable for DVD-R include metal azo complexes, quinophthalones, and diphenylmethanes. The cyanine dyes are very useful for all current optical disk recording applications [35].

A novel rewriteable optical disk comprising a Si-naphthalocyanine dye and the conductive polymer polythiophene as the recording layer has been

reported which could be played back on conventional CD players, even after ten cycles of rewriting [54]. The phthalocyanine derivative has recently been reported to be capable of recording high-quality signals at high speeds [55].

A series of metallized thiazolyl-azo dyes with a hydroxy substituent ortho to the azo linkage shows apparently promising performance [56, 57]. Zn(II) complex dyes possess solubility, thermal stability, and suitable absorption at 635 nm and are used as DVD-R dyes [58].

Cross-References

- [Coloration, Textile](#)
- [Colorant, Nonlinear Optical](#)
- [Coloration, Fastness](#)
- [Dye, Photodynamic Therapy](#)
- [Dye, Ink-Jet](#)

References

1. Gregory, P.: Chapter 9.12. In: McCleverty, J.A., Meyer, T.J., (eds.) *Comprehensive Coordination Chemistry II: From Biology to Nanotechnology*, 2nd edn., vol. 9. Elsevier, Oxford, (2005)
2. Sekar, N.: *Colourage* **47**(4), 27–28 (2000). 35
3. Hannemann, K.: *Int. Dyer* **186**(3), 16–19 (2001)
4. Price, R.: Dyes and pigments. In: Wilkinson, G. (ed.) *Comprehensive Coordination Chemistry*, pp. 35–94. Pergamon, Oxford (1987)
5. Adams, H., Bucknall, R.M., Fenton, D.E., Garcia, M., Oaks, J.: *Polyhedron* **17**(23–24), 4169–4177 (1998)
6. Lenoir, J.: Organic pigments. In: Venkataraman, K. (ed.) *The Chemistry of Synthetic Dyes*, vol. V, pp. 313–474. Academic, New York (1971)
7. Gordon, P.F., Gregory, P.: *Organic Chemistry in Color*, pp. 219–226. Springer, Berlin (1983). Chapter 5
8. Provost, J., Gregory, P. (Avecia Ltd.): Multicolor Ink Jet Printing. US Patent 6,336,721, 2002
9. Kayaku, N.: World Patent Application, 200064901A, 1999
10. Evans, S., Weber, H.: Metal Complex for Ink Jet. US Patent 6,001,161, 1999
11. Gairns, R.S.: Electrophotography. In: Gregory, P. (ed.) *Chemistry and Technology of Printing and Imaging Systems*, pp. 76–112. Blackie, London (1996)
12. Gregory, P.: *High-Technology Applications of Organic Colorants*, pp. 59–122x. Plenum, New York (1991). Chapter 7

13. Moser, F.H., Thomas, A.L.: Phthalocyanine Compounds. Rheinhold, New York (1963). pp 62, 109, 133, 134
14. Hage, R., Iburg, J.E., Kerschener, J., Koek, J.H., Lempers, E.L.M., Martens, R.J., Racherla, U.S., Russell, S.W., Swarthoff, T., Van Vliet, M.R.: *Nature* **369**(6482), 637–639 (1994)
15. Martens, R. J., Swarthoff, T. (Unilever): Detergent compositions containing dinuclear manganese complex as bleach. Canadian Patent 2,083,661, 1993
16. Gregory, P., Reynolds, S., White, R.L. (Zeneca): Sulfonated manganese phthalocyanines and compositions thereof. US Patent 5484,915, 1996
17. Dougherty, T.J., Gomer, C.J., Henderson, B.W., Jori, G., Kessel, D., Korbelik, M., Moan, J., Peng, Q.J.: *Natl. Cancer Inst.* **90**, 889–905 (1998)
18. Stranadko, E.F., Skobelkin, O.K., Vorozhtsov, G.N., Mironov, A.F., Beshleul, S.E., Markitchev, N.A., Riabov, M.V.: *Proc. SPIE* **3191**, 253–262 (1997)
19. Bonnett, R.: Chapter 9.22. Metal complexes for photodynamic therapy. In: McCleverty, J.A., Meyer, T.J. (eds.) *Comprehensive Coordination Chemistry II: From Biology to Nanotechnology*, 2nd edn., vol. 9. Ward, M.D., (ed.) *Applications of Coordination Chemistry*. Elsevier, Oxford, (2005)
20. Finsen, N.R.: Phototherapy; Sequeira, J.H., Transl. Arnold, London (1901)
21. Regan, J.D., Parrish, J.A. (eds.): *The Science of Photomedicine*. Plenum, New York (1982)
22. Lim, H.W., Soter, N.A. (eds.): *Clinical Photomedicine*. Marcel Dekker, New York (1993)
23. Bonnett, R.: *Rev. Contemp. Pharmacother.* **10**, 1–17 (1999)
24. Boehncke, W.-H.: *Curr. Opin. Anti-inflamm. Immunomodulatory Invest. Drugs* **1**, 338–352 (1999)
25. Gregory, P., Thetford, D. (Zeneca): Poly-substituted phthalocyanines. US Patent 5486,274, 1996
26. Service, R.F.: *Science* **268**, 806 (1995)
27. Tinolux BBS Trade Literature, Edition; Clba-Geigy, April, 1986. 578 Metal Complexes as Speciality Dyes and Pigments (1986)
28. Bao, Z.N., Rogers, J.A., Katz, H.E.J.: *Mater. Chem.* **9**, 1895–1904 (1999)
29. Stark, W.M.(ICI): Substituted Phthalocyanine. US Patent 4,824,947, 1989
30. Kaeida, O., et al. (Nippon Shokubai): Novel phthalocyanine compounds: production method thereof: near infrared ray absorption materials containing the same. European Patent 523,959, July 14, 1992
31. Gregory, P.J.: *Porphyrins Phthalocyanines* **3**, 468–476 (1999)
32. Gregory, P.: *High-Technology Applications of Organic Colorants*, pp. 215–254. Plenum, New York (1991). Chapter 11
33. Haley, N.F., Corbiere, S.L.: Radiation-sensitive composition containing a resole resin and a Novolac resin and use thereof in lithographic printing plates. US Patent 5,372,907, 1994
34. Haley, N.F., Corbiere, S.L.: Method of making a lithographic printing plate containing a resole resin and a Novolac resin in the radiation sensitive layer. US Patent 5,372,915, 1994
35. Hurditch, R.: Dyes for optical storage disc media. Chemichromics '99, New Orleans, (1999)
36. Wolleb, H. (C.I.ba-Geigy): Process for the preparation of brominated, alkoxy-substituted metal phthalocyanines. US Patent 5,594,128, 1997
37. Evans, S., Elwood, J.K. (Eastman Kodak Co.): Photographic products and processes employing novel non-diffusible magenta dye-releasing compounds and precursors thereof. US Patent 4,420,550, 1983
38. Chapman, D., E-Ming, W. (Eastman Kodak Co.): Photographic products and processes employing novel nondiffusible heterocyclazonaphthol dye-releasing compounds. US Patent 4,207,104, 1980
39. Anderson, R.B., et al. (Eastman Kodak Co.): U.S. Patent Application Serial No. 282,613, 1981
40. Harper, P., Wherrett, B. (eds.): *Nonlinear Optics*. Academic, New York (1977)
41. Shen, Y.R.: *The Principles of Nonlinear Optics*. Wiley, New York (1984)
42. Saleh, B.E.A., Teich, M.C.: *Fundamentals of Photonics*. Wiley, New York (1991)
43. Calabrese, J.C., Tam, W.: *Chem. Phys. Lett.* **133**, 244–245 (1987)
44. Mata, J., Uriel, S., Peris, E., Llusar, R., Houbrechts, S., Persoons, A.J.: *Organomet. Chem.* **562**, 197–202 (1998)
45. Senechal, K., Maury, O., Le Bozec, H., Ledoux, I., Zyss, J.J.: *Am. Chem. Soc.* **124**, 4560–4561 (2002)
46. LeCours, S.M., Guan, H.W., DiMagno, S.G., Wang, C.H., Therien, M.J.J.: *Am. Chem. Soc.* **118**, 1497–1503 (1996)
47. Gong, Q.H., Wang, Y., Yang, S.C., Xia, Z.-J., Zou, Y. H., Sun, W.F., Dong, S.M., Wang, D.Y.: *J. v Phys. D: Appl. Phys.* **27**, 911–913 (1994)
48. Sun, W.F., Byeon, C.C., Lawson, C.M., Gray, G.M., Wang, D.-Y.: *Appl. Phys. Lett.* **77**, 1759–1761 (2000)
49. Hurditch, R.: *Adv. Color Sci. Technol.* **4**, 33–40 (2001)
50. Hunt, P.A.: Optical data-storage systems. In: Gregory, P. (ed.) *Chemistry and Technology of Printing and Imaging Systems*. Blackie, London (1996)
51. Hamada, E., Shin, Y., Ishiguro, T.: *Proc. SPIE* **1078**, 80 (1989)
52. Mortimer, R.J., Rowley, N.M.: Chapter 9.13. In: McCleverty, J.A., Meyer, T.J. (eds.) *Comprehensive Coordination Chemistry II: From Biology to Nanotechnology*, 2nd edn., vol. 9. Ward M.D. (ed.) *Applications of Coordination Chemistry*. Elsevier, Oxford, (2005)
53. Clba, Supergreen Product Dossier, Issue 2.0 (1999)
54. Tomiyama, T., Watanabe, I., Kuwano, A., Habiro, K., Takane, N., Yamada, M.: *Appl. Opt.* **34**, 8201–8208 (1995)
55. Verhoeven, J.A.Th., Mischke, W.S.: Developments in CD-R. In: Gan, F., Hou, L. (eds.) *Proc SPIE* 2001.

- International Society for optics and photonix (SPIE). Fifth International Symposium on Optical Storage (ISOS 2000), vol. 4085, Bellingham, US, pp. 298–305
56. Wang, S., Shen, S., Xu, H., Gu, D., Yin, J., Tang, X.: *Dyes Pigments* **42**, 173–177 (1999)
 57. Wang, S., Shen, S., Xu, H., Gu, D., Yin, J., Dong, X.: *Mater. Sci. Eng.* **B79**, 45–48 (2001)
 58. Park, H.Y., Lee, N.H., Je, J.T., Min, K.S., Young, J.H., Kim, E.R., Lee, H.: *Mol. Cryst. Liq. Cryst.* **371**, 305–308 (2001)

Dye, Photodynamic Therapy

Mark Wainwright

Pharmacy and Biomolecular Sciences, Liverpool
John Moores University, Liverpool, UK

Definition

Photodynamic therapy may be thought of as a directed and controlled application of the cellular damage possible from light energy. There are two main areas of clinical interest, these being the *photodynamic therapy* of cancer (shortened to PDT) and the parallel application to microbial disease and infection control, the latter having several names, including *antimicrobial photodynamic therapy* (APDT), *photodynamic antimicrobial chemotherapy* (PACT, used in this review), and *photodynamic disinfection* (PDD). In the case of both areas clinical interest, in order to target and mediate cellular damage, a light-absorbing chemical, called a *photosensitizer*, is used.

In both anticancer and antimicrobial guises, the underlying science is the same, namely, (i) uptake by or association of the target cell with a photosensitizer, (ii) illumination of the target cell and the concomitant electronic excitation of the photosensitizer, and (iii) chemical reaction of the photosensitizer with its biomolecular environment or energy transfer from the photosensitizer to a local oxygen molecule. In both of these final scenarios, reactive oxygen species result which are highly toxic in the cellular environment (Fig. 1). The combination of locally administered

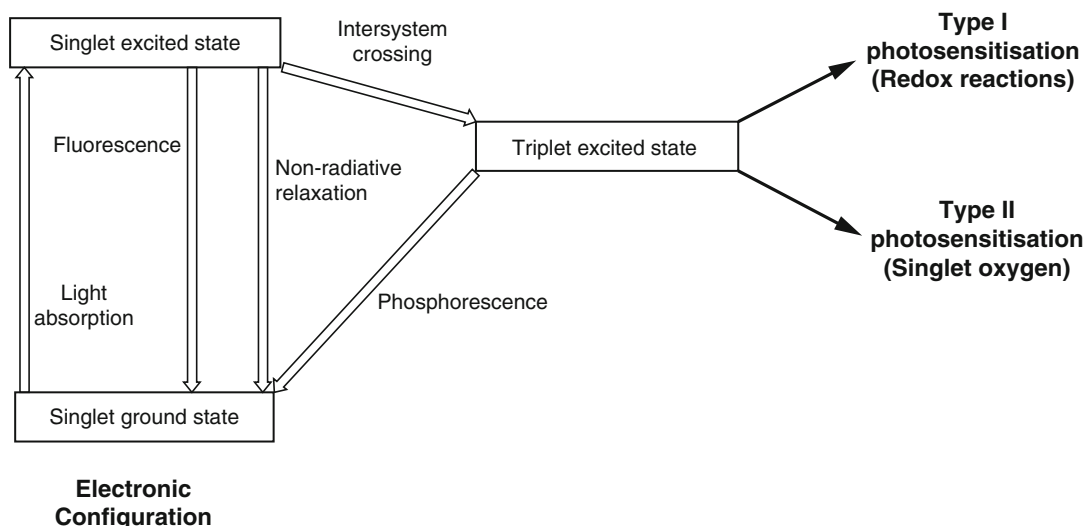
photosensitizer and targeted delivery of illumination results in the production of a rapid, destructive local effect by means of which local therapeutic control can be maintained that is absent of side effects in parts of the body remote from the treatment site.

Introduction

The interaction of light and matter is an ordinary, everyday occurrence, so much so, that it is ignored by the majority. However, without the effects caused by solar radiation on plant species, there would be no life on planet Earth. Despite the perceived simplicity of plant life, photosynthesis, the process by which plants absorb light and store the energy in the form of carbohydrates, is a complex sequence of chemical interaction and energy transfer. This process is, of course, a positive use of light energy. The converse may be seen in the increasing incidence of skin cancers, usually in those individuals who, unwittingly or not, allow parts of their skin structure to be damaged by the higher energy portion of sunlight or by the radiation provided by sunbeds.

Photosensitizers and Photodynamic Therapy (PDT)

Ideally, PDT involves the use of a nontoxic photosensitizer which produces a toxic effect in the target cells upon illumination. In practice, this has proved elusive, either because of inherent or “dark” toxicity or lack of selectivity for target cells. However, in comparison to conventional chemotherapy or radiotherapy, PDT offers considerable advantages, mainly due to any side effects experienced being localized. Compared to surgical excision, the demonstrable advantage of the PDT approach lies in much improved cosmesis, i.e., a lack of tissue damage, disfigurement, and scarring. This can be particularly important, for example, in head and neck cancers, from an aesthetic point of view, and also in external gynecological cancers due to loss of functionality and impaired sex drive.



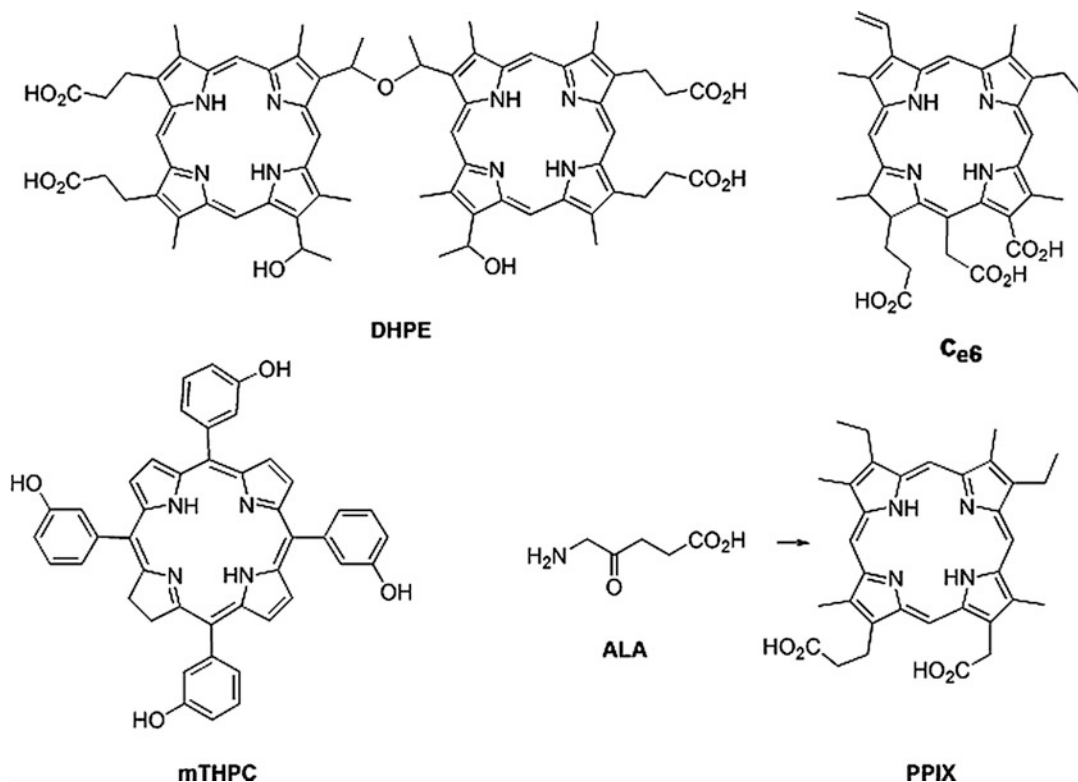
Dye, Photodynamic Therapy, Fig. 1 Photosensitization pathways

Photosensitizers Used in PDT

Much of the original work which suggested photodynamic therapy as an approach to malignant disease was due to clinical findings, for example, Policard's observation of the fluorescence of porphyrins in the blood supply of tumors under fluorescent light during surgery [1]. The porphyrin species involved were naturally occurring species, normally present in the blood. Later work, by Lipson and others, utilized porphyrin compounds derived from bovine blood and thus known as hematoporphyrins or hematoporphyrin derivatives (HpD, see Fig. 2). These were mixtures of short chains (oligomers) of hematoporphyrins (Fig. 2) and their constituency was variable by means of chemical derivation. However, despite the lack of purity entailed, HpD was shown to be reasonably effective in "shutting down" tumors via oxygen starvation, and purer forms, such as "porfimer sodium," have been licensed clinically for a range of presentations. The main side effect associated with such photosensitizers derives from a lack of selectivity when administered systemically, which has often lead to generalized patient photosensitization – i.e., a strong reaction to normal daylight, posttreatment.

So-called "second-generation" photosensitizers related to porphyrins for use in PDT include the chlorins mTHPC and chlorin e_6 (Fig. 2). Both of these agents offer the advantage of longer-wavelength light absorption than their predecessors, due to the marginally altered ring system (one carbon-carbon double bond is missing). However, non-optimal tumor selectivity was again identified as a problem (with consequent generalized photosensitization as recounted above), and the search for improved photosensitizers has continued, producing new compounds related to the original porphyrins, but with variations on the size and constitution of the ring system.

Understanding of the heme biosynthetic pathway has led to the use of a natural porphyrin precursor, 5-aminolaevulinic acid (ALA, Fig. 2), and its derivatives in PDT. ALA is normally utilized in mammalian cell heme (protoporphyrin IX, PPIX, Fig. 2) synthesis, but the increased metabolism exhibited by tumor cells means that reasonable ratios of the resulting photosensitizer are possible between normal and tumor cell populations. Local application of ALA-containing creams and optimized incubation times have produced excellent therapeutic



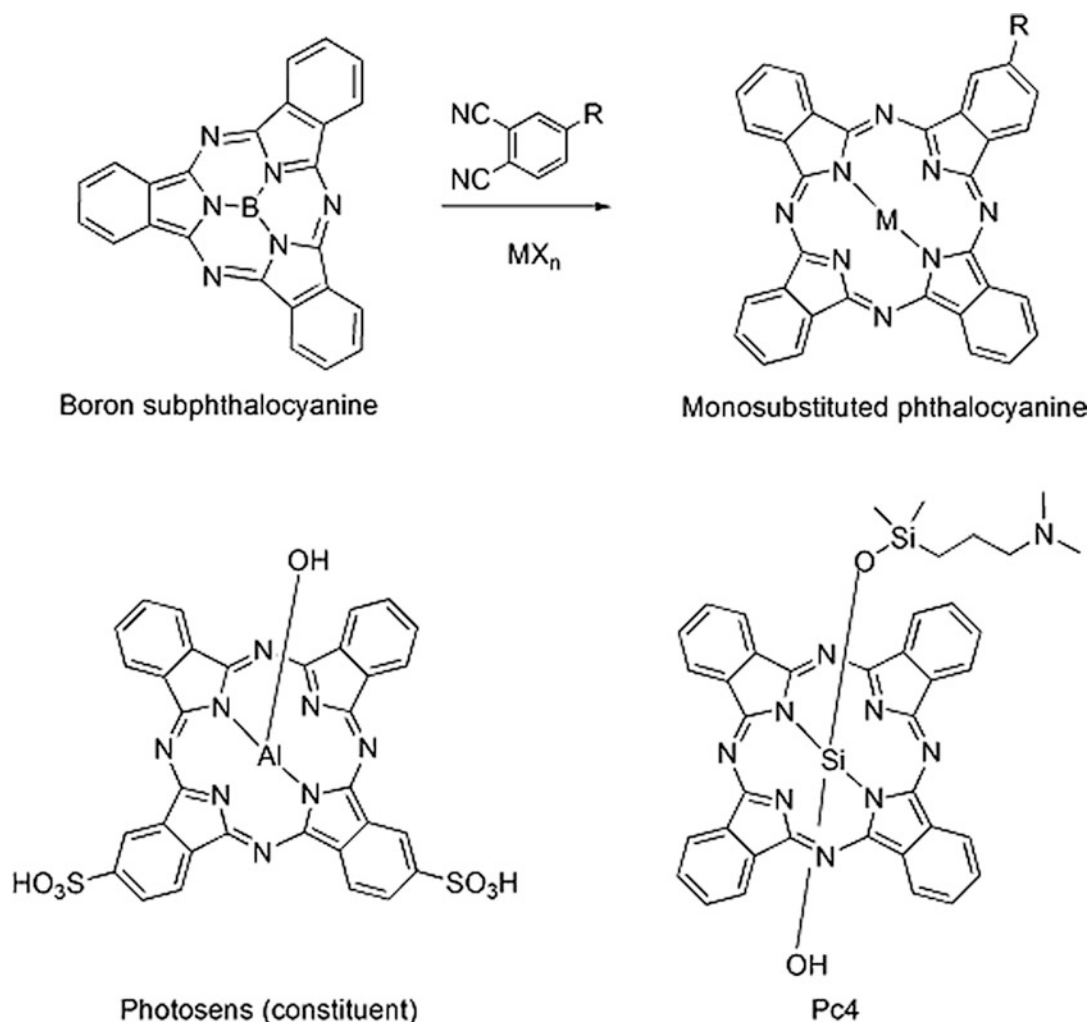
Dye, Photodynamic Therapy, Fig. 2 Porphyrin photosensitizers used in PDT. *DHPE* dihematoporphyrin ether, a constituent of HpD and porfimer sodium, *c₆₆* chlorin *e₆*,

mTHPC meso-tetra (*m*-hydroxyphenyl)chlorin, *ALA* 5-aminolaevulinic acid, a precursor of protoporphyrin IX (PPIX)

outcomes in several forms of skin and soft tissue malignancy. As might be imagined, local application of photosensitizers in PDT does not result in the generalized photosensitivity reported with systemic administration as mentioned above.

Apart from porphyrin-type photosensitizers, the main other chemical class employed for clinical PDT is based on the phthalocyanine group. As may be seen from Fig. 3, this class of photosensitizer is structurally related to the porphyrins covered previously, but the phthalocyanines have no naturally occurring lead compounds and are an entirely synthetic class of photosensitizer. Originally discovered during industrial aromatic dye synthesis in the 1930s, phthalocyanines were accidentally produced via the interaction of phthalic acid, ammonia, and copper from the reaction vessels, copper phthalocyanine being the first example of this type of pigment.

Phthalocyanine derivatives offer a considerable advantage over many previously used porphyrins in exhibiting significantly increased intensity of light absorption coupled with extension of absorption into both the red and near infrared regions of the visible spectrum. Although the dye chemistry required for phthalocyanine production is well established, this is based industrially on the original copper phthalocyanine, which is an extremely stable blue pigment – thermally, chemically, and photochemically. The characteristically high photochemical stability of this class of photosensitizer is attributable to the presence of the central copper(II) ion, which has a paramagnetic electronic configuration. This provides a rapid deactivation route for the photoexcited singlet state (see Fig. 1), and consequently, insignificant amounts of reactive oxygen are formed which might otherwise



Dye, Photodynamic Therapy, Fig. 3 Phthalocyanine derivatives for PDT

decolorize the pigment. Phthalocyanine research for PDT purposes has therefore been concentrated on derivatives having diamagnetic central metal ions such as zinc or aluminum.

In most cases, phthalocyanines for colorant end use are produced as pigments. Since solution formulations are the norm in clinical use, the inclusion of suitable solubilizing chemical groups has been a significant area of development for PDT-useful derivatives. A small amount of liposomal formulation has been also used for tumor delivery of the non-functionalized material.

To produce readily aqueous-soluble phthalocyanine dyes, metallophthalocyanines may be

directly sulfonated. Conversely phthalic acid derivatives bearing water-solubilizing groups (e.g., sulfonic or carboxylic acids) may be employed. In both of these approaches, the derivatives produced are mixtures by degree of functionality and isomeric constituency. While this is acceptable in commercial dyes, it is not in pharmaceutical development – significantly, a major argument against the continued use of hematoporphyrin derivatives remains their lack of chemical purity. However, a sulfonated mixture based on chloroaluminum phthalocyanine (*Photosens*, Fig. 3) has been in use in the former Soviet Union.

The problem of isomeric functionality has been overcome via the use of subphthalocyanine chemistry (Fig. 3), allowing the introduction of monofunctionality in the ring system, and via the use of a reactive central atom such as silica, which may be functionalized both above and below the plane of the ring (e.g., Pc4, Fig. 3).

Necrosis/Apoptosis

The majority of photosensitizers employed in PDT are either negatively charged or neutral molecules which are generally oil soluble to a significant degree (lipophilic). In many cases this means that they are extensively bound to blood lipoproteins and, consequently, concentrate in the tumor vasculature. Illumination of the tumor site at this stage thus leads to vascular damage and occlusion, starving the tumor of oxygen.

Direct damage to the tumor itself requires uptake into tumor cells. Illumination then either causes cell death by general, nonspecific oxidation, leading to tumor necrosis, or by programmed cell death (apoptosis). The latter route is more complex and involves an initial damaging event, e.g., to the lysosome or mitochondrion which then triggers intracellular messengers such as the caspases and leads to a cascade of internal processes and cell death.

Treatable Cancers

Initially, the main – and logical – target for PDT was skin cancer, given the ease of application of both photosensitizer and light to such presentations. However, the adoption of fiber optics allowed a rapid expansion of the approach into deeper-seated disease. Similarly, advances in photosensitizer design produced derivatives, as noted above, with extended light absorption characteristics. These provided a more efficient use of light since the longer wavelengths required (≥ 660 nm) lie outside the absorption spectra of endogenous absorbers such as heme and melanin. Table 1 provides an overview of the range of treatment available for malignancy and premalignancy.

Dye, Photodynamic Therapy, Table 1 Photosensitizers clinically trialed in oncology. ^aClinically approved

Photosensitizer	Malignancy indication
Hematoporphyrin derivative (HpD)	Barrett's esophagus
	Bladder ^a
	Cervical ^a
	Gastric (superficial) ^a
	Lung ^a
	Esophageal ^a
	Pancreas
	Peritoneum
5-Aminolaevulinic acid (ALA)	Actinic keratoses ^a
	Barrett's esophagus
	Basal cell ^a
	Bowen's disease
	Gynecological
<i>meso</i> -Tetra(3-hydroxyphenyl) chlorin (mTHPC)	Barrett's esophagus
	Brain
	Head and neck ^a
	Lung
	Pancreas
	Prostate

PACT (Photodynamic Antimicrobial Chemotherapy)

The negative effect of illumination on a stained microorganism was first demonstrated by Raab in 1900, thus pre-dating PDT [2]. However, whereas porphyrin-based PDT has become a significant clinical field, the application to microbial pathogens has taken over a century even to become a clinical reality on a small scale. The main reason for this is the straightforward access to effective “conventional” antimicrobial agents enjoyed by the affluent countries of the world over the past 60 years. Only in the twenty-first century, with the waning efficacy of many drugs, is there a requirement for alternative approaches to infection control, particularly in view of bacterial resistance to conventional drugs, such as penicillins and tetracyclines.

Logically, as PACT uses the same approach as PDT, but applied to microbial pathogens, its limitations are similar also. Thus, the photosensitizers employed in or proposed for PACT, known collectively as “photoantimicrobials” due to their

extremely broad spectrum of activity, require local or topical application due to the light activation factor. While this might, at first glance, appear to trivialize the approach, as it cannot thus be used for systemic infection, this is not the case. Following the PDT paradigm, localized infection may be targeted anywhere in the body via the use of endoscopy and fiber optics. Indeed, initial forays into the clinic have been made in oral and lung infection, as well as the more obvious dermatological areas, such as acne, ulcers, and nail infection.

Photoantimicrobials

One area of dissimilarity between PACT and PDT lies in the chemical types involved. Whereas PDT has remained firmly based on the porphyrin class from which it was first developed, photosensitizers for use against microbial species require different makeup for broad-spectrum activity. However, it should be noted that, since the metabolism of many bacterial species includes porphyrin synthesis, ALA may be of use here also, used in the same way as detailed above for anticancer PDT.

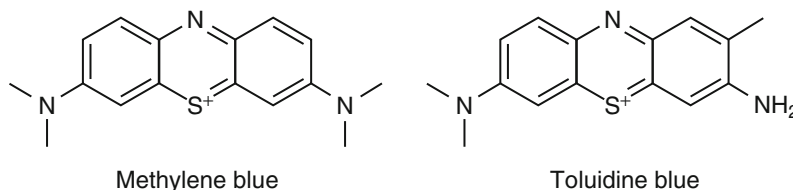
The use of dyes as chemotherapeutic agents in the first half of the twenty-first century was a result of scientific development of the heavily dye-oriented organic chemistry begun by Perkin, Hoffmann, and others. The first dye to be used clinically was the phenothiazine *Methylene Blue* (C.I. Basic Blue 9), which was shown to be antimalarial in 1890. The chemical structures of many modern antimalarials, as well as drugs for the central nervous system, have much in common with *Methylene Blue* as a consequence, the phenothiazine dye being tested as a treatment for many different illnesses mainly due to its very low toxicity in humans. *Methylene Blue* has also been the lead compound in photoantimicrobial research since it was shown to be active against bacteria, viruses, fungi, and protozoa by various groups before the mid-1930s. Unfortunately for the field of photoantimicrobial endeavor, 1935 saw the initial clinical release of the first conventional antibacterial agent, sulfanilamide

(as *Prontosil*) by IG Farben, beginning what has been known to many as the “antibiotic era.” Clearly, effective and easily administered antibacterial agents are more appealing than an approach requiring light activation, and photoantimicrobial research languished as a result.

Unsurprisingly, given the flagrant overuse of our valuable antimicrobial arsenal, drug resistance – particularly in bacteria – has become a very serious clinical problem. Indeed the problem is increasing to such an extent that many strains of bacteria in our hospitals are approaching indestructibility, as far as conventional antibacterial drugs are concerned. This resistance is merely the outcome of evolution directed by overuse, and misuse, of drugs which have only a single site/mode of action, i.e., which may be circumvented by a single alteration as a result of natural bacterial mutation. One of the advantages of the photoantimicrobial approach is that there are multiple sites and modes of action, making microbial resistance development far less likely.

Broad spectrum, in terms of photoantimicrobial agents, means activity against bacteria, viruses, fungi, and protozoa, and this is generally associated with positively charged (cationic) photosensitizers, such as *Methylene Blue* and its derivatives. One of the reasons that the porphyrin photosensitizers utilized in anticancer PDT are not useful leads for photoantimicrobial development is that, as anions, they have little effect on Gram-negative bacteria and so cannot be considered to be broad-spectrum agents. However, even this level of activity is massively different to that of conventional antimicrobial drugs, which are *either* antibacterial *or* antiviral *or* antifungal *or* antiprotozoal.

Methylene Blue itself is a tricyclic heteroaromatic molecule which absorbs red light in the region of 660 nm. The related dye *Toluidine Blue* (C.I. Basic Blue 17) absorbs at a shorter wavelength (c. 625 nm), mainly due to the non-alkylation of one of its amino groups (Fig. 4). Both of these photosensitizers produce significant yields of singlet oxygen, both in vitro and in vivo. While there have been considerable strides made in synthesizing improved agents

Dye, Photodynamic Therapy,**Fig. 4** Phenothiazinium photoantimicrobials

with these dyes as lead compounds, the amount of regulatory testing required has so far hindered progress toward the clinic, and Methylene Blue appears to be the chosen derivative, having been licensed for oral disinfection and blood plasma decontamination in several countries. The extension of such applications to bacterial decolonization and other local anti-infection measures is proposed in order to lessen the burden on conventional antimicrobial delivery, particularly in view of the resistance problem and the dwindling number of agents effective against serious diseases such as septicemia and bacterial meningitis.

Light Sources

Early clinical PDT was carried out normally using in situ hospital lasers, which were expensive both to purchase and to maintain. While the advent of much cheaper diode lasers improved the situation financially, both sources were tied to single-wavelength output and were thus restricted in the number of photosensitizers which might be used. However, both tunable dye lasers and light-emitting diode (LED) sources have revolutionized the area, in terms of adaptability and cost, such that financial considerations present no bar to the adoption of the technique by healthcare concerns. It should be noted that there are still only three photosensitizers generally licensed for clinical use (Table 1).

Treatment of Age-Related Macular Degeneration (ARMD)

Photosensitizers can be used in the treatment of wet age-related macular degeneration – a condition, normally in the older patient, where new blood vessels (neovasculation) grow from the

back of the eye and break through into the region of the retina containing the rods and cones necessary for the perception of color and shading. The new vessels leak blood and fluid, causing scarring which is associated with a gradual loss of central, but not peripheral, vision.

The concentration of lipophilic, anionic (typically porphyrin-type) photosensitizers in tumor vasculature has been mentioned. This propensity has a positive spin-off in the treatment of wet, age-related macular degeneration. The use of lipophilic anionic photosensitizers such as chlorin *e*₆ (verteporfin) allows selective targeting and ultimate destruction of the neovasculation.

Cross-References

► [Dye, Functional](#)

References

1. Brown, S.B., Brown, E.A., Walker, I.: The present and future role of photodynamic therapy in cancer treatment. *Lancet Oncol.* **5**, 497–508 (2004)
2. Wainwright, M.: Photodynamic antimicrobial chemotherapy (PACT). *J. Antimicrob. Chemother.* **42**, 13–28 (1998)

Dyeing

► [Coloration, Textile](#)

Dyes

► [Pigment, Inorganic](#)

Dynamic Color Design

► Environmental Color Design

Dynamic Color Spreading

► Motion and Color Cognition

Dynamics of Color Category Formation and Boundaries

Stephanie Huetten
Department of Psychology, University of
Memphis, Memphis, TN, USA

Definition

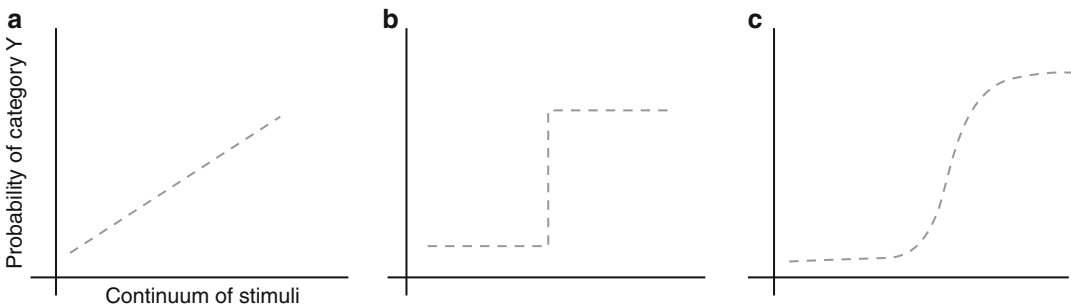
Dynamics of color boundaries is broadly the area that characterizes color categorization as a process that occurs over time.

Color Boundaries

A color category boundary is the area in perceptual color space where the labeling of one color

transitions into another label. Color is a continuous wavelength dimension but distorted into nonlinearly perceived perceptual categories. A category boundary is fundamentally a distortion of continuous color space during perceptual and cognitive processing of color. There is no external signal in the wavelength dimension to tell the perceptual system to make this abrupt transition. This is called *the problem of invariance* [1]. The demarcation of color space has led to empirical work seeking to characterize color boundaries and discover their underlying dynamics.

There are various tasks that can be used to define a color boundary. One way researchers can do this is simply by presenting a color swatch and asking participants for the appropriate color label. Boundaries can also be measured by presenting two color swatches and asking whether they are the same or different, which is called *discrimination* (see “► Color Categorical Perception” for more details). The categorization graphs show a curve that can be classified as one of three types: step, sigmoid, or linear (Fig. 1). The step function is a theoretical ideal not found in empirical data. The discrepancy between the theoretical ideal and obtained data may be due to some meaningless noise in the perceptual system and appears in the data as a sigmoid, even though the actual underlying transition may be a discrete transition. Alternatively, the sigmoid may accurately reflect the boundary: The slope of the sigmoid is indicative of how gradual the transition is between



Dynamics of Color Category Formation and Boundaries, Fig. 1 Panel (a) illustrates a linear or identity function, panel (b) the step function, and panel (c) the sigmoid. A linear function represents sensitivity to continuity and a lack of categorization. The step function is the most extreme categorization, where there is an

instantaneous transition from one category into another category. The sigmoid in panel (c) represents the more gradual transition from one category into another. In each panel, the *x-axis* represents a continuum ranging from category X to category Y, and the *y-axis* represents the probability or activation value for category Y

categories. Finally, there could be a linear perception of the stimuli, which would indicate not having categories.

Human experimental data displays sigmoidal categorization boundaries, where stimuli near the boundary are not always reliably categorized into the same category. This area near the boundary is unique, because it represents a place in perceptual space where there is a distortion. This makes between-category distinctions a larger perceptual distance apart, effectively compressing within-category distances between stimuli and expanding between-category distinctions. One account states that categorical distortions are independent from and occur in parallel to continuous hue processing [2]. This issue bears upon whether categorical distortions have an effect on the perception of color or if they are a cognitive phenomenon induced by labeling rather than a warping of perceptual space [3].

Further support for the role of labels in the formation of category boundaries comes from comparative work on color perception. Color category boundaries do not arise from precortical processing, as there is no invariant signal coming from opponent process cells or the spectral sensitivities of cones themselves. Thus, it is unlikely that boundaries arise from early low-level visual processes likely to be found in other primates. Baboons do not have color boundaries; rather, they perceive hue linearly, even though their retinal and opponent cell processing are similar to humans ([4]; also see “► [Comparative Color Categories](#)”). This evidence supports the role of higher level cortical processing and lends support to the Whorfian hypothesis (also called linguistic relativity). There is also evidence that color boundaries are learned and not innate further emphasizing a strong role for labels [5].

A dynamic view of color boundary formation is well supported by many already existing theories, including linguistic relativity (see “► [Comparative Color Categories](#)” for more detail). This hypothesis acknowledges that the influence of language labels exerts some influence over the perception and formation of boundaries [6]. This influence could modify a boundary in many ways: Two labels could sharpen the boundary between

them, making the sigmoidal slope steeper and expanding perceptual distances across the category boundary. One label could make the slope shallower or even linearly perceived (non-categorical). Finally, a label could translate the boundary, acting as an attractor or repeller for the boundary.

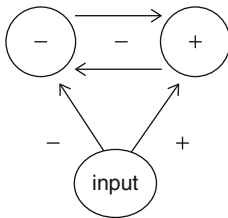
These kinds of dynamics are easily conceptualized as linguistic. However, many other factors could exert an influence on boundaries, such as biology, environment, or context. These alterations of the boundary could also happen at various timescales, on the order of minutes (adaptation; [7]), years (lifespan; [8]), or much longer spans (evolution; [9]). As opposed to thinking of color as something static and invariant, a dynamic view instead emphasizes the role of change and how those occur over time. The typical measurement of boundaries does not reflect influences of context that can alter the boundary [10].

Within a category hues are labeled the same. This could be because they are processed equivalently or because the task forces participants to place the stimuli into one of two categories. If they are processed equivalently, the underlying activation of the category should be all or none: For example, on a given trial, a given hue would be processed as entirely “green.” Alternatively, both color categories may be active to various degrees. For example, the hue would be processed probabilistically as 0.8 “green” and 0.2 “blue.” Over hundreds of milliseconds, the pressure of being required to put the hue in one of two categories would create competition between the two, finally resolving into a discrete answer of either “green” or “blue.” This latter process is called gradient categorization and was shown to apply to color perception [11].

One study demonstrated gradient categorization in color perception using a computer mouse tracking paradigm [11]. A color categorization task in which participants would begin moving vertically up the screen with a computer mouse and a color from a continuum of green to blue would appear briefly on the screen. The participants would continue moving up the screen and indicate their response by moving over a

prototypically blue or green swatch in one of the top corners. The trajectories of computer mouse movements show subtle deviations toward the competing category response option – for example, veering toward green when the final response is blue. These deviations reflect the maintenance of continuous hue detail used to help make a discrete, categorical response. The categorical end goal and the gradation in processing coexist in the competition model and parallel human performance.

To demonstrate the maintenance of some form of continuous detail, whether that is in terms of raw hue value or in terms of distance to the category boundary, there must be evidence of a monotonic relation to the category boundary, as measured by some behavioral channel. This kind of signal is not seen in measurements taken only



Dynamics of Color Category Formation and Boundaries, Fig. 2 The inputs range from -1 to $+1$ which represent a range of hues to be categorized. The second layer represents the two category options, and the lateral connections between them perform the competition

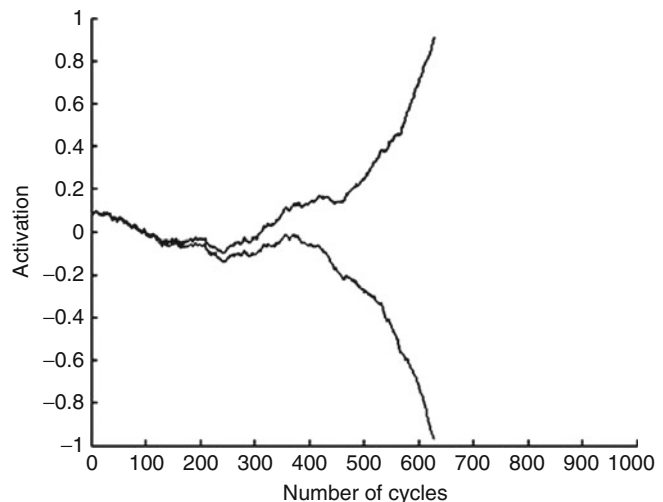
Dynamics of Color Category Formation and Boundaries,

Fig. 3 Competition over time: the two nodes activations are plotted over a number of cycles. You can see the activations begin very similarly, and then gradually one begins to win, and it rapidly takes the activation away from the other node. This simulation contains a random noise term, which is responsible for the smaller fluctuations in each line, and can produce variability in the final response

once per trial, as in typical naming tasks. A category boundary of some sort is mathematically guaranteed in a task requiring a decision between one of two items: Thus, within-category subtleties in processing are often neglected by measuring the boundary, which is informative about what is happening near the boundary (e.g., the steepness of the slope indicating the amount of distortion near the boundary).

Graded categorization can be computationally modeled as competition or even with an evolutionary algorithm emphasizing the role of motor movements [12]. Competition can be implemented with lateral inhibition, meaning that two representations are connected laterally and mutually inhibit one another (Fig. 2). This kind of a network shows gradations in performance during perceptual choices [13]. This network will cycle through the lateral connections, gradually taking the category with the highest activation and increasing its activation by stealing from the other node (Fig. 3). A mechanism like this is hypothesized to underlie the formation of discrete responses, and these gradations in initial activation can be seen in motor responses that are sampled at a very high density [11]. The competition model serves as a mathematical proof that one could accomplish a task such as categorical perception with minimally complex circuitry.

Category boundary formation is a dynamic phenomenon susceptible to context such as labels,



objects, and the task being used. The typical color boundary is a relatively steep sigmoidal shape, and the steepness of this slope indicates the degree to which the perceptual or cognitive system has warped the continuous hue space. This could be due to competing labels or prototypes. The dynamic view emphasizes the characterization of these boundaries over many different timescales, and many different areas of color perception address these kinds of dynamic influences (Linguistic relativity, top-down influences, evolution of color categories, etc.) Future directions in this area may look more closely at these variables, and seek to characterize the process in real time, using dense-sampling methods appropriate to the timescale of the phenomenon.

Cross-References

- [Color Categorical Perception](#)
- [Comparative Color Categories](#)

References

1. Hamad, S. (ed.): *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, New York (1987)
2. Bornstein, M.H., Korda, N.O.: Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychol. Res.* **46**, 207–222 (1984)
3. Pilling, M., Wiggett, A., Özgen, E., Davies, I.R.L.: Is color “categorical perception” really perceptual? *Mem. Cogn.* **31**, 538–551 (2003)
4. Fagot, J., Goldstein, J., Davidoff, J., Pickering, A.: Cross-species differences in color categorization. *Psychon. Bull. Rev.* **13**, 275–280 (2006)
5. Özgen, E., Davies, I.R.L.: Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *J. Exp. Psychol. Gen.* **131**, 477–493 (2002)
6. Roberson, D., Davidoff, J., Davies, I.R.L., Shapiro, L. R.: The development of color categories in two languages: A longitudinal study. *J. Exp. Psychol. Gen.* **133**, 554–571 (2004)
7. Bornstein, M.H., Korda, N.O.: Identification and adaptation of hue: parallels in the operation of mechanisms that underlie categorical perception in vision and audition. *Psychol. Res.* **47**, 1–17 (1985)
8. Raskin, L.H., Maital, S., Bornstein, M.H.: Perceptual categorization of color: a life-span study. *Psychol. Res.* **45**, 135–145 (1983)
9. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. *J. Math. Psychol.* **51**, 359–382 (2007)
10. Goldstone, R.L.: Effects of categorization on color perception. *Psychol. Sci.* **6**, 298–304 (1995)
11. Huette, S., McMurray, B.: Continuous dynamics of color categorization. *Psychon. Bull. Rev.* **17**, 348–354 (2010)
12. Beer, R.D.: The dynamics of active categorical perception in an evolved model agent. *Adapt. Behav.* **11**, 209–243 (2003)
13. Usher, M., McClelland, J.L.: The time course of perceptual choice: the leaky competitive accumulator model. *Psychol. Rev.* **108**, 550–592 (2001)

Dyschromatopsia

- [Color Perception and Environmentally Based Impairments](#)

E

Early Color Lexicons

- [Ancient Color Categories](#)

Early Expression of Color in Language

- [Ancient Color Categories](#)

Eco-/Genotoxicity of Synthetic Dyes

- [Colorant, Environmental Aspects](#)

Effect of Color Terms on Color Perception

Jonathan Winawer and Nathan Witthoft
Department of Psychology, Stanford University,
Stanford, CA, USA

Synonyms

[Categorical perception](#); [Linguistic influences on color perception](#); [Sapir-Whorf hypothesis](#)

Definition

The possibility that naming colors, either in a single instance or habitually over a lifetime, alters color perception.

Color Perception and Color Communication

When we communicate about the colors of scenes and objects comprising our visual experience, what we see informs our choice of words. A question that has interested many cognitive psychologists is whether the color words we use affect how we see. One view is that naming is for communication and has no effect on how we see or experience the world. In this view, color appearance, and therefore performance on tasks which strictly depend on color appearance, is determined entirely by the visual system and not at all by the language that one speaks. Color vision is informationally encapsulated, its output is automatically produced, and while its output is available for cognitive processes like decision making and speaking, higher cognitive processes cannot alter its process [1]. This can be contrasted with the idea that the very act of naming a color, either in a single instance or habitually over a lifetime, can change one's perceptual experience with the result that colors assigned with the same label (i.e., belonging to the same color category) are more similar than colors assigned with

different labels (belonging to different color categories).

The two views have persisted in the psychology literature for at least two reasons. First, the empirical evidence has been mixed, with some researchers finding effects of color terms on perceptual or cognitive tasks and others failing to find such effects. Second, when positive evidence has been found, it has often been unclear at what level the effects are manifested. For example, the same experimental results might be interpreted by some researchers as evidence that color terms affect how one sees colors and by other researchers as evidence that color terms affect how one remembers or makes judgments about colors. In this entry, we examine the question of how to design experiments to probe the effects of color terms on color perception and cognition and how to interpret the effects.

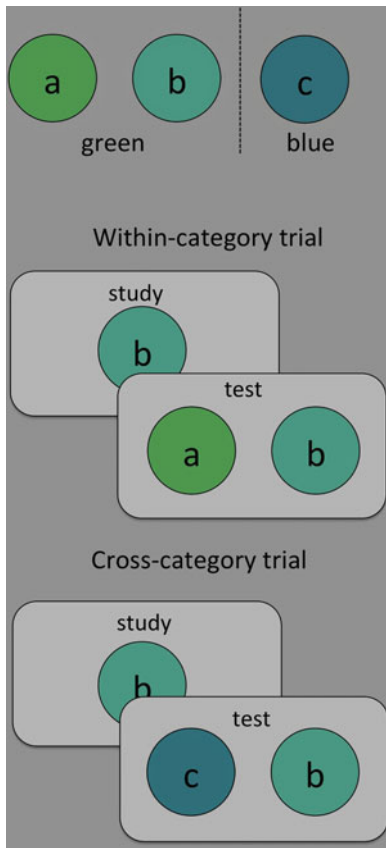
The study of categorical effects in color perception has its roots in two historical traditions. One is a concept that originated in anthropology, called the Sapir-Whorf hypothesis [2]. Whorf and Sapir formulated this hypothesis in different ways in their writing, but researchers have generally taken their view to be something like the way in which experiences are mapped onto words affects the experiences themselves, so that speakers of different languages perceive and conceive of the world differently. A second and related historical tradition is the study of categorical perception [3, 4] and in particular the study of phoneme perception in language. Distinctions between certain phonemes, such as /r/ and /l/, are necessary for speaking some languages, such as English, but not other languages, such as Japanese. Adult English speakers are better at the /r/ versus /l/ discrimination than Japanese speakers (though Japanese speakers are above chance, meaning that the effects are not completely categorical). In phoneme learning, it appears that the learning is primarily a loss rather than a gain; the /r/ and /l/ sounds become more similar as a child learns Japanese rather than more different as a child learns English [5]. Psychologists studying color have similarly asked whether making color distinctions improves sensitivity to distinguish shades that straddle a

category border or worsens one's sensitivity to distinguish shades that fall within a single color category.

No color scientist takes seriously the idea that color perception is as categorical as color naming. It is trivial to find two colors that are both called "green" that can be distinguished from one another. A more nuanced question, then, is whether colors that are named alike show any degree at all of increased perceptual similarity (or decreased discriminability) compared to colors that are named differently.

Breaking the Circularity: The Logic of Measuring Category Effects

To assess the effects of color categories on perception and cognition, one has to sort out the direction of causation: a pair of colors may belong to different categories because they look very different, rather than looking different because one learns to put them in different categories. In fact, it is certain that this happens; colors that look alike are more likely to belong to the same color category, and language groups across the world tend to assign colors to categories in similar ways, though some differences exist [6]. To know whether there is *also* an effect in the reverse direction, that is, an effect of category on perception or cognition, one needs to have a prediction of how similar the colors should be in the absence of a category effect. Given such a prediction, one could ask the following: are two colors from the same category (Fig. 1a, b) less discriminable (or more similar) than two colors from different categories (Fig. 1b, c), even when the two pairs are identically spaced? That is, are "within-category" trials harder than "between-category" trials? This leads to a potentially circular problem. If the stimulus pairs are chosen such that they are equally spaced, then are they not guaranteed to be equally similar? Or, conversely, if one pair is shown to be more discriminable than the other, does this mean that the color space which placed them at equal distances from one another was flawed?



Effect of Color Terms on Color Perception, Fig. 1 Testing effects of color terms on a color task. The *upper* panel shows three stimuli that might be used in an experiment to investigate whether color terms affect a particular task. If the task involves memory, the subject may be instructed to first remember a color (“study”) and then identify the remembered color (“test”). If the incorrect stimulus comes from the same category, the trial is considered “within-category.” If the incorrect stimulus comes from a different category, it is considered a “cross-category” trial. The colors were chosen so that *a* and *b* look *green* and *c* looks *blue*, but the observer may not agree depending on the display and individual differences between observers

This entry summarizes the answer to two questions pertaining to whether and how color naming affects color cognition:

1. How are experiments designed to break the problem of circularity posed above?
2. If the problem of circularity can be broken, then what aspects of color perception and cognition show effects of color names?

Model-Based Approaches

One way to break the circularity problem in studying the effect of color names on color cognition and perception is to compare the results to a null model, meaning a model which does not use color categories as part of its machinery even as it tries to predict them. The choice of color space for a null model is an essential consideration. If the space is designed to reflect perceptual similarity, such as the Munsell color space, then it may already incorporate category effects. And if such a space fails to predict performance on a color judgment task, then it may simply indicate that the space did not perfectly achieve its design goal, rather than something about color categories. Hence, a null model is of limited value for the study of effects of language on color processing if the model is purposely shaped to account for the pattern of human behavior, such as the Munsell system or the CIELab space.

In contrast, a model that sticks more closely to peripheral aspects of color processing can be more informative. For example, Brown and colleagues used the MacLeod-Boynton space to predict reaction time in a visual search experiment for an oddball color [7]. The MacLeod-Boynton space represents color stimuli as a linear transform of cone excitations. It contains no information about color names or categories or other properties of color cognition. In the visual search experiments, one color (the target) differed from the other colors (the distractors), and the time to find the target was modeled as a function of the difference between the target and distractors in MacLeod-Boynton space. Trials in which there was a large difference between the targets and distractors in the space were predicted to have a faster reaction time. Some trials were cross-category and some within-category, analogous to the design in Fig. 1. The authors found that the null model provided a good fit to the empirical results, with no systematic deviations for cross-category trials. For this color task, color discrimination time could be explained by a model based only on cone absorptions, and hence there was no evidence that discrimination was affected by color categories. The null model succeeded.

It is important to consider what happens when a null model fails. Suppose, for example, a model predicts that stimuli a and b and stimuli b and c in Fig. 1 are equally discriminable, and yet experiments show that cross-category pairs are discriminated faster or recalled more accurately than within-category pairs. While additional mechanisms or a different model may fit the data better, it does not follow that color categories made the stimuli more discriminable. It may simply be that the model was incorrect or inappropriate for the task.

Equating Color Stimuli but Varying the Subject Population

Another way to break the circularity problem in studying the effect of color names on color cognition and perception is to test the identical sets of color stimuli with speakers of languages with differing color terms.

For example, for English speakers there is no common color word that refers to all three stimuli in Fig. 1. Speakers of some languages (sometimes called “grue” languages), however, could refer to all three colors by the same name. An experimenter can therefore avoid the difficulty of perfectly equating differences between the stimulus pairs: for speakers of languages like English that place a boundary between stimuli b and c, the b/c pair should be more different than the a/b pair *relative to* speakers of “grue” languages. The colors are controlled not by the exact localization of colors in a color space but rather by showing exactly the same sets of colors to different groups of observers and seeing if their pattern of responses differs. One of the challenges in such experiments is finding appropriate subject populations. Color naming patterns are quite similar across many languages. Nonetheless, large differences in naming patterns can be found by comparing populations from an industrialized western society such as the UK or the USA with speakers from more remote cultures which have had minimal contact with western societies, such as the Dani and Berinmo of Papua New Guinea or the Himba of Namibia. However, even among modern industrial societies, differences can exist. For example, Russian speakers use different

words for the colors light blue (“goluboy”) and dark blue (“sinij”), while English speakers may refer to both as “blue.”

The results of cross-linguistic studies have not been uniform. Some have found effects of language on a color task, such as a color memory task comparing English speakers with Berinmo speakers [8]. Others have found little to no effect. One reason is that the effects, even when found, tend to be small compared to the shared properties of color processing found in all observers. For example, consider the effect of metamers: two colors with quite different spectral power distributions will be indistinguishable to an observer if the two colors result in the same pattern of cone excitations. This is a very large effect and is quite consistent across observers. A metameric pair for one person is likely to be a metameric pair for another person, or nearly so, unless one of the observers has a color vision deficit. Put another way, three-channel color displays such as television do not need to be tailor-made according to the color naming patterns within a language community. Color metamers do not differ with language because the information lost at the level of cone absorptions cannot be regained. The cross-linguistic differences that have been observed tend to be quite modest in magnitude relative to the much larger effects shared across all observers.

Equating Color Stimuli but Varying the Hemifield: Lateralized Whorf

Rather than testing for an effect of language by holding the stimuli constant and varying the language of the subjects, some experimenters have compared the same speakers but on different sides of visual space. In these experiments, ostensibly the same set of color stimuli are shown on either the left or the right side of a fixation point. The logic of this kind of experiment is that verbal processes are computed primarily in the brain’s left hemisphere, and visual inputs are routed to the contralateral hemisphere, so that stimuli to the left of fixation are processed more in the right hemisphere, and vice versa. The separation by hemisphere is not complete for either vision or language; nevertheless, researchers hypothesized

that if effects of color terms on color tasks were to be found, they would be larger for stimuli presented to the right of fixation than the left of fixation. A number of reports have supported this hypothesis, known as the lateralized Whorf effect [9]. Other reports have failed to confirm the finding, however, and the reasons for the discrepancies have not yet been resolved [7, 10].

Equating Color Stimuli but Varying the Role of Language: Verbal Interference

Another strategy to break the circularity of measuring a category effect is to use the same set of stimuli under different conditions designed to make it easier or harder to use language in the task. For example, a color task can be conducted under conditions in which the subjects' verbal faculties are engaged with a secondary task, such as reading a list of words aloud ("verbal interference"). If a secondary, language-related task reduces a putative category effect, then the problem of circularity is solved for this experiment. There is no need to precisely equate the distances between color pairs; rather, the role of language is implicated because performance changes when the availability of language changes. Experiments have shown that manipulating the availability of language during a color task can reduce the effect of color terms on the task.

The classic experiment of this type was done by Kay and Kempton [11], who compared English speakers to Tarahumara speakers from northern Mexico. The Tarahumara use the word "siydney" to name both green and blue. Subjects from both groups were shown triads of color chips which straddled the blue/green boundary and asked to select the chip which was least similar to the other two. They found that English speakers tended to exaggerate the distance between chips which straddled the boundary, while the Tarahumara did not. This finding is consistent with the idea that the availability of the terms "blue" and "green" influenced the English speakers' decisions about which color chip was the most dissimilar when one of the chips came from a different category. Kay and Kempton proposed that rather than resulting from some difference in color appearance, the English speakers were

relying on the linguistic distinction to make decisions when the perceptual information was insufficient. To test this idea, they repeated the task, but only allowed the subjects to compare two of the chips at a time. The pairs were set up so that the chip in the middle of the color range could be alternately compared with one of the two to either side. This meant that if the middle chip was at a boundary, it would seem bluer than one chip and greener than the other, thus eliminating the usefulness of having a verbal code. Under these conditions the English speakers showed the same performance as the Tarahumara.

If a category effect goes away when labels become unavailable or not useful, then it is unlikely that the effect is due to color terms affecting early perceptual processes. While such an account is logically possible, it would require color appearance to be altered only during those moments when one is accessing the labels. A more parsimonious explanation is that the *decision* process is affected by language. Verbal labels may be used to help keep track of the various stimuli in an experiment, either over a memory delay or when comparing stimuli spread over space. If, on a particular trial, all the stimuli come from the same verbal category (e.g., they are all blue), then labels are unlikely to help accomplish the task (and might even hinder performance). In contrast, if stimuli in a trial can easily be assigned with different labels (e.g., one blue and one green), then access to the labels may facilitate memory or the comparison process. If a verbal dual task interferes with the ability to label stimuli, even implicitly, then this may eliminate one strategy or source of information for accomplishing the task and hence may change performance. Thus, verbal interference effects are more likely to reflect a role of color terms on decisions, strategy, and memory, rather than perception.

Cognition and Perception: What to Measure

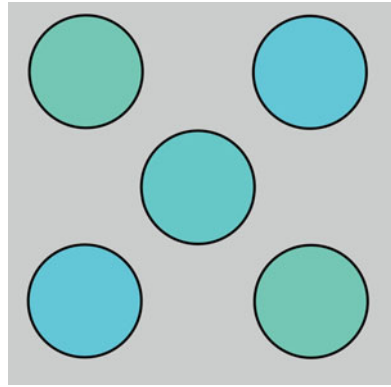
Thus far we have primarily been focused on how to break the problem of circularity when

measuring the effect of color terms on color tasks. Here we focus on the tasks themselves. What kinds of tasks are used to test for the effects of color terms, and what do the results show?

Subjective Judgments

Perhaps the most natural way to ask whether color terms affect color perception is to ask subjects to make direct, overt judgments about color stimuli. For example, one could present subjects with a variety of color samples and ask them to place into groups those that are most similar. Experiments like these have been conducted with observers across languages. Grouping patterns tend to be similar across language groups, but some differences are found consistent with grouping together colors that are given the same name in the language. The difficulty with this type of result is that subjects may choose to group together colors because they have the same name, rather than because they look most alike. Hence, it is difficult to identify whether the effect of color terms is on perception, or on the strategy of how to form groups.

Perceptual Grouping. A more indirect approach can make it less likely that subjects use a naming strategy to make a decision. For example, Webster and Kay took advantage of the tendency of human observers to perceptually group together objects that have a similar color [12]. By arranging disks into two diagonals with an ambiguous central disk, participants could indirectly report perceptual similarity by indicating whether the disks grouped into a line sloping up to the right or up to the left (Fig. 2). The question they then asked was whether the decision to group dots into a triplet was influenced by color categories, specifically whether there was a bias to group the dot with the diagonals when the three came from the same color category, regardless of the proximity in color space. These effects were compared to a null model based on a linear transform of cone excitations, similar to the approach used by Brown and colleagues [7]. The null model did a good job explaining the data, and hence there was no evidence that color categories modulated the task.



Effect of Color Terms on Color Perception, Fig. 2 Gestalt grouping configuration used to study the effect of color categories on color appearance [12]

Hue Scaling. Interestingly, a different measure of color perception using the same null model did show a significant deviation from the model. This measure is called “hue scaling,” in which subjects view a color sample and rate the perceived amount of red, green, blue, and yellow in a stimulus. Compared to a null model, colors that fell on the blue side of the blue/green border were effectively rated as containing more blue than would be expected purely from the model. That is, the results were biased away from the null model in the direction predicted by an effect of color terms on the task.

Subjective judgments are a desirable way to measure perception because they are direct: if one wants to know what something looks like to a subject, then asking for a subjective judgment is a sensible way to find out. On the other hand, because the judgments are subjective (there is no right answer) and there is no time pressure, these tasks are also amenable to strategic decisions that have little to do with perception. For example, when subjects are explicitly asked to rate the amount of blue in a stimulus, the internal mapping from perception to a judgment may be influenced by the fact that the subject knows that the stimulus falls in the category “blue.”

Memory

Memory tasks designed to probe the effect of color terms on perception or cognition often take

the form of remembering a color over a delay and then trying to choose between the same stimulus and a second stimulus, where on some trials the second stimulus comes from the same category and sometimes a different category. Memory tasks have sometimes been used in cross-linguistic studies and sometimes in studies that included dual tasks (verbal interference). Several memory studies have found effects of color category on performance (e.g., [8, 13]). These studies are consistent with the idea that color terms are used as a dual code in memory; if the subject remembers the term in addition to the appearance, then the subject will be more accurate when the memory test includes only one option from the expected color category.

Reaction Time on Suprathreshold Discrimination

In a suprathreshold discrimination task, subjects typically judge which color patch matches a test color. There is only one right answer, and the discrimination is usually well above threshold so that subjects are expected to make the right answer most of the time. The measurement of interest is reaction time, and the typical experimental question is whether reaction time is faster for cross-category trials than within-category trials. In order to avoid the circularity problem, the experimenter can use one of the strategies described above, such as cross-linguistic comparisons, dual tasks, or varying the visual hemifield to which the stimuli are presented. Alternatively or in combination with these strategies, the experimenter can provide a null model to predict the data.

Suprathreshold experiments have had mixed results. Some have reported an effect of color categories on performance and some have not. When an effect of color categories is found in a speeded suprathreshold discrimination task, there are typically two alternative explanations the experimenter is confronted with. The participant may be faster on cross-category trials because stimuli that come from different categories look more different or the participant may be faster on cross-category trials because a decision is easier to make when color labels support the decision. As

discussed above, if a verbal dual task eliminates a category effect, then the effect probably is in the decision stage rather than in the percept.

Discrimination Thresholds

Threshold discrimination experiments are among the least ambiguous experiments in psychology. If an observer can discriminate two stimuli, then we can be certain that the observer's perceptual system has encoded the two stimuli differently. If the stimuli are indistinguishable (below threshold), then information distinguishing the stimuli was either not encoded or was lost in subsequent processing. If discrimination thresholds were altered by the color terms in one's language, this would provide the most direct evidence that color terms affect perception of colors. Experiments that have compared threshold discrimination among speakers of languages that divide the color spectrum differently have found no effect of color terms [14].

Conclusion and Future Directions

We began with the question of whether the words we use to describe colors affect how we see. A major challenge in addressing this question is sorting out the direction of causality: since similar colors are likely to be given the same name, and dissimilar colors are likely to be given different names, how does one measure whether naming has any effect on how the colors appear? Several methods have been used to address this problem, including the use of null models, cross-linguistic experiments, verbal interference, and hemifield manipulations.

Despite some discrepancies in the literature, some general patterns are emerging. Results are consistent with the idea that color terms can serve as a dual code in conjunction with sensory representations for certain kinds of judgments. For example, when one remembers the color of a red car, say, one might remember both the appearance (image-based memory), as well as the fact that the car was red (declarative memory). Similarly, when one makes a judgment or decision about color appearance, the knowledge that a color

belongs to a particular category might affect the speed of the response or the content of the response without affecting the appearance of the color. A variety of tasks and experiments have reported effects of color terms on color cognition in the domains of judgment and decision making and memory. There is no firm evidence that color appearance or color threshold discrimination is affected by color terms.

The pattern of results in this way differs from phoneme discrimination. This may reflect the fact that phonemes have a categorical function: words are combinations of phonemes, and if a phoneme is mis-categorized, then the word may be misheard. In contrast, objects and scenes are not combinations of color categories. There is generally no need to categorize or label a color in order to recognize an object or scene. The fact that we *can* categorize colors by labeling them may have implications for how we remember or reason about colors, but because labeling is not an inherent part of a visual process, perhaps we should not expect it to have a significant effect on visual appearance or discrimination.

Future work on the topic would benefit from explicitly modeling the psychological processes involved in specific color tasks, in order to better understand how color terms affect the various components of cognition such as memory, decisions, and preferences.

Cross-References

- Color Categorical Perception
- Dynamics of Color Category Formation and Boundaries
- Psychological Color Space and Color Terms
- World Color Survey

References

1. Pylyshyn, Z.: Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* **22**(3), 341–365 (1999)
2. Whorf, B.L.: *Language, Thought and Reality*. MIT Press, Cambridge (1956)

3. Harnad, S. (ed.): *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge (1987)
4. Bornstein, M.H., Korda, N.O.: Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychol. Res.* **46**(3), 207–222 (1984)
5. Kuhl, P.K.: Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**(11), 831–843 (2004)
6. The World Color Survey. <http://www1.icsi.berkeley.edu/wcs/>
7. Brown, A.M., Lindsey, D.T., Guckes, K.M.: Color names, color categories, and color-cued visual search: sometimes, color perception is not categorical. *J. Vis.* **11**(12), 1–21 (2011)
8. Davidoff, J., Davies, I., Roberson, D.: Colour categories in a stone-age tribe. *Nature* **402**(6724), 604–604 (1999)
9. Kay, P., Regier, T., Gilbert, A., Ivry, R.B.: Lateralized Whorf: language influences perceptual decision in the right visual field. In: Minett, J., Wang, W. (eds.) *Language, Evolution, and the Brain*, pp. 261–284. The City University of Hong Kong Press, Hong Kong (2009)
10. Witzel, C., Gegenfurtner, K.R.: Is there a lateralized category effect for color? *J. Vis.* **11**(12), 1–25 (2011)
11. Kay, P., Kempton, W.M.: What is the Sapir–Whorf hypothesis? *Am. Anthropol.* **86**, 65–79 (1984)
12. Webster, M.A., Kay, P.: Color categories and color appearance. *Cognition* **122**, 375–392 (2012)
13. Roberson, D., Davidoff, J.: The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cogn.* **28**, 977–986 (2000)
14. Roberson, D., Hanley, J.R., Pak, H.: Thresholds for color discrimination in English and Korean speakers. *Cognition* **112**, 482–487 (2009)

Efficient Coding

- Environmental Influences on Color Vision

Elastic Light Scattering

- Rayleigh and Mie Scattering

Elastic Scattering

- Rayleigh and Mie Scattering

Electrical Control Gear for Lamps

Wout van Bommel
Nuenen, The Netherlands

Definition

Device(s) in the electric lamp circuit of gas discharge and solid-state lamps which limits the lamp current to the required value and ensures that the lamp can be started and once started can be operated stably over a longer period.

The devices needed are called igniters, ballasts, and, for LEDs, drivers.

Igniters

Most of the gas discharge lamps have such a high internal resistance that they need a voltage peak, higher than the normal operating voltage, to initiate the discharge.

Igniters for Fluorescent Lamps

In most fluorescent lamp circuits, the electrodes are preheated for a few seconds before a high-voltage peak is applied across the lamp to initiate the discharge. Preheating of the electrodes facilitates the emission of electrons. Glow-switch starters were often used for this purpose. However, their harsh peak voltages have a negative effect on lamp life. Electronic starters with exactly the same functions as the glow-switch starter, but with a better-controlled voltage peak, are more and more used. They always ignite after one ignition pulse, so eliminating flickering during ignition. Because of the better control of the ignition peak, electrodes suffer less damage so that lamp life increases.

Igniters for HID Lamps

Ignition of HID lamps is initiated by a high peak voltage without preheating of the lamp electrodes. Igniters for HID lamps are normally started using an electronic igniter that generates a series of high-voltage pulses of the required magnitude

(varying for the different lamps between some 200 and 5000 V). The electronic circuit is so designed that these pulses cease after ignition has taken place. Because of differing requirements for ignition voltage, shape of voltage peak, and number of voltage pulses within a certain period, each type of HID lamp (and often also each different wattage) needs its own type of igniter.

Ballasts

If mains voltage is supplied to electrical devices, the current has to be limited; otherwise, it will keep increasing until the device or circuit in which it is flowing breaks down. In many electrical devices, the internal electrical resistance of the device limits the current. These are devices with a so-called positive-resistance characteristic in which the current is automatically limited to a small range. Incandescent lamps have a positive-resistance characteristic and thus need no separate current-limiting device. Gas discharge lamps have a negative-resistance characteristic, meaning that the unlimited current available will increase until the lamp breaks down.

Electromagnetic Ballasts

By introducing an external resistor in the lamp circuit, the current is stabilized. Simple resistors can thus be used as current-limiting ballasts for gas discharge lamps. They do, however, dissipate a lot of power. Instead, inductive coils are used, viz., copper wire wound around an iron core. They have the same effect as a resistor, but at much lower power losses in the ballast. The ballast also ensures that the lamp continues to operate despite the fact that twice during each frequency cycle of the mains voltage (in Europe 50 Hz and in the USA 60 Hz) the current is zero, the lamp is off and thus has to be reignited. An inductive ballast system helps to reduce the time the lamp is “off” during each zero passage of the current. This is because an inductive system shifts the phase between mains voltage and lamp current.

Until the 1980s, all ballasts consisted of inductance type of ballasts also called

“electromagnetic ballasts” because of the electric and magnetic fields their coils generate. Since then, more efficient electronic ballasts have been developed.

Electronic Ballasts for Fluorescent Lamps

Electronic ballasts for fluorescent lamps electronically transform the sinusoidal 50 Hz mains frequency into a square-wave voltage of higher frequency, hence their name HF electronic ballasts. The high-frequency voltage lies between 25,000 and 105,000 Hz (viz, 25–105 kHz). HF electronic ballasts work on the same inductive principle as conventional electromagnetic ballasts. At the higher frequencies, much smaller coils with correspondingly smaller losses can be used than in conventional electromagnetic ballasts. When a fluorescent lamp operates on a frequency higher than some 10 kHz, the efficiency with which the radiation is created increases. The more efficient light production combined with the lower ballast losses means that the system efficacy of fluorescent lamps operated at high frequency increases by some 20–25 % compared to that of the same lamps operated on electromagnetic ballasts.

The electronic ignition function is incorporated in the ballast. With electronic operation the exactly required starting voltage pulse can be accurately supplied to the lamp. This minimizes electrode damage and so considerably increases lamp life. Relative to lamps operated on electromagnetic ballasts, the lamp life is increased by 30–50 %.

Electronic Ballasts for HID Lamps

Electronic ballasts for HID lamps have become available for some types of HID lamps. In contrast with HF electronically operated fluorescent lamps, there is hardly anything to be gained in terms of improved efficiency in the light-creation process by operating HID lamps on an electronic ballast. Nevertheless, energy savings are obtained, because the power losses in electronic HID ballasts are lower. The advantages of electronically operated HID lamps often lie in the better and easier control of certain lamp properties.

Drivers for LEDs

Instant starting is no problem with solid-state lamps, but they are low-voltage rectifiers that allow current to pass in one direction only. This means that the AC mains supply has to be transformed to low voltage and then rectified into a DC supply. Although a solid-state light source has a positive resistance characteristic, the voltage–current dependency is exponential in the area of operation. Small fluctuations in supply voltage therefore cause large variations in current that can damage the light source. A simple series resistor in the electrical circuit stabilizes the current to create, in fact, a “constant” current supply. In practice only miniature indicator LEDs use such a resistor. Because so much energy is lost in resistor-type drivers, all high-power, high-brightness LEDs employ electronic drivers that provide for the transformation from high voltage to low voltage, for rectification, and for the constant current supply.

Power losses in electronic drivers vary, according to their quality, between something like 10 and 30 %, with really -poor-quality drivers, as much as 50 % of the nominal wattage of the LED itself.

Cross-References

- ▶ [High- and Low-Pressure Sodium Lamp](#)
- ▶ [High-Pressure Mercury Lamp](#)
- ▶ [Metal Halide Lamp](#)
- ▶ [Tubular and Compact Fluorescent Lamp](#)

Electrodeless Low-Pressure Mercury Lamps

- ▶ [Induction Lamp](#)

Electromagnetic Radiation

- ▶ [Light, Electromagnetic Spectrum](#)

Elementary Colors

► Primary Colors

Emotional Color Effects

► Psychological Color Effects

Environmental Chemistry of Synthetic Dyes

► Colorant, Environmental Aspects

Environmental Color Design

Galen Minah¹ and Antal Nemcsics²

¹College of Built Environments, Department of Architecture, University of Washington, Seattle, WA, USA

²Department of Architecture, Budapest University of Technology and Economics, Budapest, Hungary

Synonyms

Color dynamics; Contextual color design; Dynamic color design

Definition

Environmental color design involves the use of color in order to configure it as a more beautiful, usable, and informative component of the environment that allows theoretical and practical activities. Color design is a relevant aspect of environmental design, which includes such disciplines and practices as landscape design, urban planning, architecture, interior design, industrial design, engineering, graphic design, textile and

fashion design, etc., all of them aimed at a transformation of the natural environment with the purpose of adapting it to human life.

The Function of Color in Space Experience

Man living in a built environment, the user of a built space, is protected from the tribulations of nature; he utilizes the services of his surroundings and enjoys comfort from these services. In addition to its actual measurable properties required for the physical and biological existence of man, built environment has other qualities too.

Built space acts upon man in several ways: by the proportions, the relationship, and shape of its elements, the order of forms, by surface ► [appearance](#) and colors of the elements, by the relation between space proportions, by the expression of function, by the relation between the expressivity of function and the function proper, and by the shape and color associations expressing function. This effect materializes as an emotional experience of the actual space, and space sensation. Space sensation is an experience about a given space, an accomplishment of one's own personality.

The Function of Color in Space Perception

Space perception is a complex process to which several sensory organs contribute. Among them, visual and auditory stimuli and those arising from motion in space are the most important. All these add up to a space stimulus eliciting the perception of space.

Space stimulus is elicited by measurable and tangible real space, composed of space elements as well as of correlations between shapes and surface ► [appearances](#), all describable by physical magnitudes. Most of the information about the objective correlations, shapes, and surface appearances of space elements is obtained by reflection, absorption, or transmission of light by the surface

of the element that delivers visual stimuli to observers.

Assuming that the surface ► **appearance** of space elements is of the same finish, texture, and color, and that the elements are illuminated from the same direction, with the same intensity and spectral distribution, then due to visual and motional parallaxes, overlapping, line and air perspective, and light-shadow effects, a space perception with a linearity directly proportional to that of the change of the real space is elicited.

Color identity as a condition means that light incident from the surfaces into the eyes has the same wavelength; that is, color sensation is the same throughout, and also, that for the same angle of incidence, the ratio of the quantity of light incident on and reflected by surfaces is the same everywhere, and in the reflected light incident on the eyes, the ratio of complementary radiations, hence saturation, is felt to be the same throughout.

To examine the function of color, let's assume that stimuli arriving into the eyes come from the elements of such an objective space where dimensions, proportions, and relations of the elements do not permit overlapping and the interpretation of line perspective relations; further the onlooker does not move in space, missing the help from laws of motion parallax in space perception. If these conditions are met, and in addition direction, intensity, and spectral energy distribution of light within the space are constant and the former restriction of equal hues, saturations, and lightnesses of surface colors holds, the objective space can be judged only by evaluating the perceived color sensation differences.

Intensity differences of the stimuli emitted by space element surfaces and reaching the eyes permit, first of all, to decide on the spatial position of the light source, then, from hue, saturation, and lightness differences of space element surfaces, on the distance of space elements from the onlooker, hence on the space itself.

It is known by experience that the more remote an object, the more hue component of the color sensation generated by its surface is shifted to hues of shorter wavelengths, its saturation toward achromatic colors, and that its lightness component varies as a function of the two other

components and of the position of the light source. This experience helps the space sensation although its significance can really be perceived only if the former condition of color identity is abandoned. In reality, this is always the case. With space elements painted different colors, it cannot be decided anymore which element is the closer and which is the farther away. Orange and red, even if in reality more distant, are felt to be nearer than blue or green. Saturated colors are felt to be nearer than are unsaturated ones. Very dark surfaces emit very little or no stimuli to the eye, so that these are not sensed, rendering space perception impossible.

The Role of Color in Expressing the Function of Space

Color contributes to space sensation also by expressing the function of space. Function of the built environment is a demand raised to social level. Structural relations in a system composed of man and the elements of his environment are defined by a complex function having three components: utility function, aesthetic function, and informative function. This section explains how color – color stimulus and color perception – contributes to the realization or expression of these functions.

Environment is the scene of human activities, serving human demands. Much of these demands refer to the utility function of environment. Built environment is required to protect from the rigors of weather, to endure dynamic forces generated by human machinery, to protect from such factors as excessive temperature fluctuations, intense noise, and other factors from working processes detrimental to health. A recent requirement is feeling of comfort in one's milieu so as to stimulate the development of mental and physical abilities.

Color has a significant function in meeting these demands. Due to its psychophysical and psychosomatic effects, it may raise blood pressure or change the composition of blood and gastric juices. Color can make one feel healthy or ill. A person in an environment of preferred colors feels better; his/her ambition to work increases.

Some colors favor concentration; others cause deconcentration [1].

Just as anything else, built space and all its elements are separable unities of content and form. Environment fulfills its aesthetic function if it expresses its utility function in conformity with the unity of content and form, where the utility function is the content, while form is expressed by shape and color of environmental elements. Since the content of objects in the environment, let alone in the built environment, is its function, the built space and objects within its content can only be grasped and fully expressed by means of their proper functioning and operation. Practical and spiritual components of the function are interdependent. Even the remark may be risked that aesthetic design of an object or built space is impossible when ignoring its functions. As a conclusion, there are no aesthetic prescriptions of general validity.

In designing color relations for a built environment as a human creation, it is also a question of what importance is attributed to practical functions of the environment for human life in general. Every work and activity is linked to emotions, thoughts, and ideas, therefore every object, tool, or built space demands its share of these mental, emotional, ideological threads, in conformity with its role, significance, and function in one's life. Colors of the built space as elements of form in the couple content and form are made necessary by the sensation of function, giving rise to a harmonious sensation of the indissoluble unity between content and form in human consciousness. Of course, the sight of some color complex may cause aesthetic pleasure, but detached from the content of space, i.e., from its function, this pleasure lacks the effect of complete space experience.

Those who wish to express the message of built space have to know about relations between environmental structures, i.e., about the so-called compositional relationships, in order to be able to create proper relations between forms and color perceptions. These relations comprise color harmonies. Thus, the design of space sensation also exploits color harmonies in this space.

Informative functions of space are features that interpret the functions of the environment and its

elements and explain how to use and operate these elements. A significant part of the informative functions of the environment are borne by chromatic information. According to their message, chromatic information may be interpreted either as logic or as aesthetic information. Both kinds of information are borne by the same elements, but every form of message has its own structure. Their characteristics are determined partly by differences in the visual system, complexity, and structure and partly as psychic differences between their communication content. Information content is transmitted by highlighting, contracting, and grouping some visual symbol elements in the informative surface or space, while omitting others. A color group draws attention when it is clear cut and its structure is easily intelligible.

Chromatic information of a logic nature, i.e., the various standardized color codes, are practical tools that appeal to a logical mind. They transmit messages and serve also to influence observers in their decisions and control their attitudes and behavior.

Aesthetic chromatic information is primarily emotional expression of inner conditions and is expected to have mental and emotional effects by commonly accepted semantics. By their operative and recording functions, visual codes are not only bearers of the meaning of the content of built space and its social concept but also expressions of the approach and culture typical of the creative subject. Chromatic in-built space information necessarily and conveniently takes the form of ► **color harmony** relations. This is why it is of prime importance to examine color harmony relations.

The Color Dynamics of Color Theory and Practice of Environmental Planning

Nowadays color dynamics is regarded as a dual activity. One side is the disclosure of man-to-color of complex man-to-colored environment relations and the elaboration of methodologies for the design of colored environment. The other side is the utilization of these findings in environment design practice. These activities involve the

collection and systematization of knowledge on relations of man, color, and built space offered by different disciplines as well as to devise and realize research to fill the gaps. This activity has taken momentum worldwide, whereby the science of color dynamics came into being.

Color dynamics as a new science is concerned with the relations between the surface ► [appearance](#) of environment and environmental elements and man living in this environment. It studies the interrelations of color, man, and environment. Thus, color dynamics as a science is a complex of theoretical and practical activities directed toward the disclosure of objective relations between man and colored environment, as well as toward a conscious environment color design.

Rather than a collection of everyday experience, color dynamics is a science. Although it investigates and processes the spontaneous, intuitive transformation of the environment by individuals, it handles its information by scientific methods, applying scientific methods in environment color design. It has been proven both theoretically and practically that it is possible to conduct these activities by exact, scientific methods.

Environment design – including any architectural activity – has increasing access to results of this new science. Its practical application helps the built environment to cope better with its function, to be more beautiful and more sophisticated. It helps humans to expand mental and bodily abilities, to compensate for harmful effects of the overwhelming industrialized environment, to develop an adequate space perception, and to understand spatial relations and correlations between spatial processes. Conscious application of colors is expected to direct and orient man between multiplying and often depressive environmental hazards.

The science of color dynamics has five different but inseparable branches; the achievements of each are interdependent. Knowledge and relations amassed by other sciences are collected and purposefully systematized, and its special research problems are based on this foundation.

The fundamental problem of color dynamics is to find relations between color sensations, to develop an aesthetically uniform color space and a color system fairly approximating to it, and to

introduce a new system of color coding suitable for practical color design.

The second group of problems is concerned with man to color relation independent of the environment. This involves color composition problems in connection with the processes of color vision. Such problems are, for instance, stimulus thresholds and difference thresholds, color adaptation, ► [color constancy](#), ► [color contrast](#), ► [color preference](#), color association, and the psychosomatic effects of color.

The third group of problems includes the complex relation between color, man, and built space, including problems of color and space, color and mass, color and form, color and texture, color and function, color and illumination, offsetting harmful environmental effects by color, and finally the social functions of color.

The fourth group of problems is related to ► [color harmony](#) research, the establishment of color composition relationships for use in practical color design: the determination of levels and parts of the concept of color harmony, of the fundamental and accessory conditions of eliciting color harmony sensations.

The last, fifth, group of problems is the development of the most effective methods of color design, the best way of incorporating the finding of color dynamics into practice. Statements are made exploiting practical observations obtained from realized color designs.

Color environment design activities, as all other design activities, constitute a continuous battle between prevailing conditions, demands of different trends, and artistic intentions. This battle may express itself on many different ways depending on the knowledge base and personality of the designer. Color dynamics has no intent to provide design recipes; rather it shows a method to provide possibility for considering different conditions and enforcing different demands [2].

Color Expression in the Built Environment

Finding a basis for using color as a concept in design is possible through the perceptual

experience of color and form. Figure/ground recognition in a visual field is based upon perception. Figure/ground also implies a hierarchical relationship.

Color and form are codependent in built form. This form can be holistic as it relates to an urban environment, or it can be an individual building. Color plays an expressive role in both. In the city, color has the power to create unity in built form by the use of a dominant color in the building materials. This visual unity can be a characteristic of a part, or district, within an urban environment, or it can be a characteristic of the whole. Unity defines a sense of place and gives identity. It also characterizes an urban fabric or background that can define the public spaces and serve as a backdrop for significant figural structures. The architecture of this background is critical to the designer's goal of creating spatial definition and unity in the plan of a city.

Another role for color expression in the built environment is the color choice for individual buildings that become prominent visually in a physical setting. These are defined as figural colors, and they will achieve this visibility through figure/ground contrasts. The atmospheric conditions and the distance from the viewer will also be a determinant in how effective the color choices are in achieving high visibility.

A third tool for color expression in the physical environment is a strategy for defining space not with walls or edges but with a repetitive use of figural color throughout the space. This strategy uses color constellations formed by elements of similar color in the environment that create a spatial matrix when seen as a whole. Constellations can define urban space in two ways: by experiencing these elements of color as a constellation when viewed as a group and through memory when one views a single color element from a pedestrian view and constructs the whole conceptually.

Background Architecture

In urban contexts, background is that median color/form of the built environment which most conforms to the overall urban fabric and contributes to its visual unity. Background is also the

structure or mass that forms and defines the spaces of importance within the city. The color of this background fabric is a component in the overall visual field. The color of building surfaces comprises a large portion of this field, and these colors can be a powerful means for defining unity and consistency. A variety of figural form and color can exist as background in a city through repetition of similar elements distributed throughout the urban fabric. In some cities, such as San Francisco, the background is relatively uniform in hue and saturation. Guanajuato, Mexico, however, has many buildings with highly saturated figural colors throughout the city and maintains its visual unity despite these foreground colors. There are two important roles for color/form in this background architecture. The first is the provision of a context for contrast to structures that are figural and highly visible in the city. The second is one of spatial definition. Background architecture becomes the edges and walls that define public spaces such as streets, avenues, civic plazas, and parks, and their organization represents the hierarchies and communal goals inherent in the plan of a city.

Figural Color

Three perceptual experiences of color have direct relationship to questions of architectural color and form. These are ► [color contrasts](#) that create figure/ground juxtapositions, the spatial effect of color, and the atmospheric color effect. Of these, figure/ground is important conceptually because it represents a hierarchical structure to which expressive values can be assigned. Figural structures in an urban environment are defined by their mass and height, their form, their position or siting within the city, and their color. In an urban context there are colors that make some buildings figural in contrast to their background. Background colors in many modern cities tend toward low blackness ($s = 10\text{--}30$) and low chromaticness ($c = 10\text{--}40$). Building colors that approach maximum whiteness, blackness, or chromaticness will usually be seen as figural in an urban environment [3]. The climate and atmosphere will affect how figural these colors appear. In most weather conditions, white and saturated colors are prominent.



Environmental Color Design, Fig. 1 Dark towers as figural color in the central business district, Seattle, Washington

Dark colors will tend to be less figural in overcast or humid conditions and appear lighter [4]. Figural color in cities creates visibility for buildings, and as a tectonic tool this visibility can denote status and provide meaning to a civic community (Figs. 1, 2, and 3).

Color Constellations

In an urban context, there are examples where figural colors and background colors become joined as a means for defining space within the city. In a two-dimensional color field where many colors are visible, color constellations are groups of similar colors recognized as a pattern or cluster. In a three-dimensional color field such as an urban environment, colors that are similar in hue and saturation can also be perceived as a cluster of repetitive color elements and form color constellations [5]. The area within this array of similar colors can be designated as an urban space, defined not by edges or walls but by a repetitive matrix of similar colors. An example would be the Parc la Villette in Paris [6], where large red architectural sculptures are placed on a grid throughout the precinct of the park. From an aerial

perspective, the spatial matrix formed by these red structures is apparent. From a pedestrian viewpoint, only a few of the structures are visible at any one time, although a memory of the precinct as a whole initiated by the individual red sculptures is always present.

Many city maps are organized by color sections to show important landmarks and institutions, shopping/retail districts, public open spaces, nightlife areas, etc. Many of these nodes and districts are repetitive and dispersed throughout the city. The map represents patterns of color clusters, like a constellation, and a large amount of information about the organization of the parts of the city is made visually accessible and comprehensible. Color in the physical environment has the potential for providing the same visual comprehension. Many cities have zoning in these districts that specify heights of buildings, density, signage and lighting, etc. that help in achieving the goal of comprehensibility from a pedestrian's view of the city, but color is rarely a consideration. One example, however, is in Washington, D.C., where the important governmental and institutional buildings are



E

Environmental Color Design, Fig. 2 Saturated *yellow* as figural color on Queen Anne Hill, Seattle, Washington

Environmental Color Design, Fig. 3 *White* as figural color, denoting status, Schiers, Switzerland



white with the background architecture in darker hues. These important buildings and landmarks form a constellation of white figural structures located in the vicinity of the Mall in the center of the city. As in Parc La Villette, the view of any of

the white buildings recreates an image of the group of figural buildings around the Mall, and any white building near the area becomes a conceptual part of this collection of significant structures (Fig. 4).



Environmental Color Design, Fig. 4 Color constellation at Parc La Villette, Paris, France

Cross-References

- [Appearance](#)
- [Color Constancy](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Color Preference](#)
- [Functionality of Color](#)

References

1. Frieling, H.: *Farbe in Kultur und Leben*. Battenberg, Munich (1963)
2. Nemcsics, A.: *Colour Dynamics. Environmental Colour Design*. Ellis Horwood, New York (1993)
3. Minah, G.: Blackness, whiteness, chromaticness: formulas for high visibility in the modern city. In: Hansuebsai, A. (ed.) *AIC 2003 Bangkok, Proceedings*, pp. 26–30. The Color Group of Thailand, Bangkok (2003)
4. Minah, G.: Figural color in the Seattle cityscape. In: *AIC Color 97, Proceedings of the 8th Congress of the International Colour Association*, pp. 883–887. The Color Association of Japan, Tokyo (1997)
5. Minah, G.: Color constellations in the Seattle cityscape. In: Chung, R., Rodrigues, A. (eds.) *AIC 2001, The 9th Congress of the International Colour Association*, pp. 146–149. SPIE, Bellingham (2001)
6. Tschumi, B.: *Cin gramme folie: le Parc de la Villette, Paris nineteenth arrondissement*. Princeton Architectural Press, Princeton (1987)

Environmental Influences on Color Vision

Mike Webster

Department of Psychology, University of Nevada, Reno, NV, USA

Synonyms

[Color adaptation](#); [Color and culture](#); [Color constancy](#); [Color signals](#); [Efficient coding](#); [Evolution](#); [Individual differences](#)

Definition

It is evident that color vision evolved to inform organisms about their environment and thus must be shaped by the environment the organism is in. The color properties of the world are a potent factor affecting not only the basic mechanisms of color vision within a species but also how they are calibrated and fine-tuned within the individual. As a result, many aspects of the physiology of color coding and the phenomenology of color experience can be predicted by looking outside the observer – to analyze the signals available in the color environment. The following provides a number of specific examples of how knowledge of the color characteristics of the environment has helped to account for some of the known properties of our color vision.

Environmental Signals and the Evolution of Color Vision

A central principle in understanding the design of sensory systems has been coding efficiency – or how to transmit the most information about the stimulus using the least resources. The assumption that vision maximizes coding efficiency provides a powerful link between the properties of the stimulus and the properties of the observer [1] and has led to a surprising number of insights about the early stages of color vision and how they are optimized for the color characteristics of the environment. For example, natural color spectra (of both illuminants and surfaces) typically vary gradually with wavelength and thus can be captured with a small number of spectral samples [2]. This may in part be why color vision is typically low dimensional (e.g., based on three classes of photoreceptors in trichromatic observers). The fact that spectra can be represented by a small number of dimensions has also been a central insight into understanding the capacity and limits of color constancy (e.g., the ability to perceive that the color of a surface has not changed even when the light falling on it has).

Scenes are composed of collections of many spectra, and these color distributions have

characteristic properties. For example, most luminances and chromaticities tend to be near the average, or gray. This property has been used to predict how neurons respond to different intensity or color levels in the image. To code these efficiently, responses should change most rapidly near gray, where signals are more frequent and thus require better discrimination, and asymptote at the extremes (e.g., highly saturated colors), where signals are rare. Thus the color statistics of images can directly account for the sigmoidal response functions of visual neurons and for the fact that we are best at discriminating small color differences near gray [3].

Color vision requires comparing the differences between the responses of receptors with different spectral sensitivities. Because the sensitivities are broad and overlapping, and because as noted spectra vary gradually, the signals within the different cone classes are highly correlated. This redundancy can be removed by recoding the cone signals into “luminance” mechanisms (e.g., in cells that receive excitatory inputs from different cones and thus sum their inputs to represent luminance) and “chromatic” mechanisms (e.g., in cells that instead receive excitatory inputs from one cone type and inhibition from another, so that the response depends on the difference or ratio of the cones’ signals). Thus redundancy reduction provides a rationale for how the cone signals are reorganized in cells in the retina [4]. Finally, the overlap in the cone sensitivities also means that the differences between the cones (which convey color) are many times smaller than the absolute responses (which convey luminance). If our vision depended on these raw cone signals, then the world would vary much more in brightness than in color. However, a further prediction from coding efficiency is that each channel uses its fully dynamic range to code the range of available signals. As a consequence we are much more sensitive to chromatic differences than luminance differences (when equated for cone contrasts), and the range of brightness and color in the world appears perceptually balanced.

The preceding examples illustrate how studies of the color environment have provided insights into general properties of the mechanisms of color vision. For many species the environment also

holds special color signals that are important to detect, because they signal food or mates, and these signals have given rise to more specialized adaptations. For example, most mammals are dichromats, yet humans and other old-world primates evolved a third class of photoreceptor through an evolutionarily recent duplication of the gene coding their longer-wave photopigment. The extra dimension of color afforded by their trichromacy has been shown to be well tuned for detecting ripe fruit among foliage and also for judging the subtle variations in skin tones that signal the health and emotional state of conspecifics [5].

Color Appearance and the Color Environment

Conventional models of color appearance hold that the signals from the three classes of cones are recombined to form three perceptual channels that signal red vs. green, blue vs. yellow, and bright vs. dark. A central issue in color science is how these perceptual dimensions are encoded by neurons of the visual pathway and why particular hue directions are special. Here again the color characteristics of the environment are thought to play a central role.

According to the standard color-opponent model, any stimulus is represented by how it differs from gray (e.g., lighter vs. darker or redder vs. greener). But what establishes gray? The spectral sensitivity of the observer varies in capricious ways depending on factors such as the level of screening pigments or the relative numbers of the different types of cone receptors and thus must be adjusted to match “gray” in the environment. This gray likely corresponds to the average spectral stimulus the observer has been exposed to. Normalizing color responses for this average probably happens at many levels of the visual pathway and at many timescales. For example, the photoreceptors rapidly adapt to changes in the average color or brightness of the scene (e.g., becoming more or less sensitive when the light levels get lower or higher, respectively). These sensitivity changes rescale their relative responses

so that the current average color tends toward gray, an adjustment known as von Kries adaptation, and this is thought to be an important initial component in color constancy (since it tends to remove or discount changes in the average color of the scene introduced by changes in the illuminant). Similarly, opponent channels must be calibrated so that the opposing cone inputs are weighted so that their null points (e.g., where the signal is neither red nor green) correspond to a neutral gray. Sensory adaptation thus plays a critical role in calibrating color vision so that it is correctly adjusted to match the observer’s environment [6]. (Importantly, these adaptive responses are not unique to color, but probably influence all aspects of perception.)

Supporting this, many lines of evidence show that the stimulus that appears gray depends very little on the observer’s actual sensitivity to wavelength. For example, individuals vary widely in the density of the lens pigment, which also steadily builds up with age. This pigment selectively absorbs at shorter wavelengths, and thus the light reaching the retina of an older observer will have a very different spectrum from a typical young observer. Despite this, the physical stimulus that appears white remains constant across the lifespan, because color mechanisms are continuously recalibrated so that they are adjusted to the average spectrum [7]. Receptors in the fovea at the center of vision are also selectively screened by a second inert pigment known as macular pigment, which further filters light at short wavelengths. Thus the fovea and surrounding periphery receive very different stimulation on average. Yet the physical stimulus that looks white again remains largely constant across the visual field, suggesting that each part of the eye can locally adapt to set the achromatic point. Patients undergoing cataract surgery provide a natural experiment for observing these adaptation processes. Cataracts represent an extreme form of lens brunescence, blocking much of the short-wavelength light. Yet the stimulus that appears white is within the normal range for many patients. When the cataract is replaced with a clear lens, the sudden increase in short wavelengths causes the world to appear too blue. However, achromatic settings almost immediately begin to

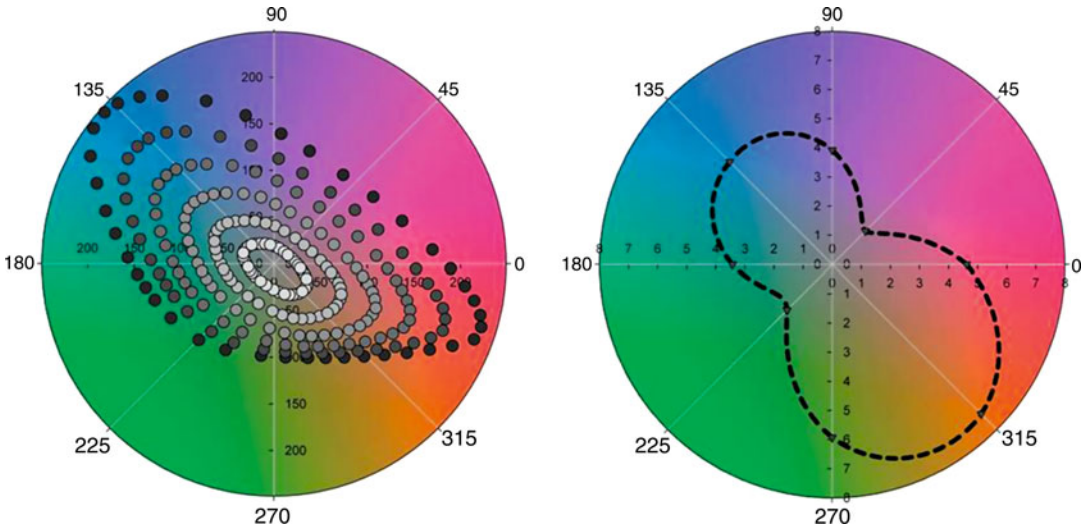
renormalize and continue to slowly drift back to their prior levels over a period that may last for months [8].

Analogous arguments have been advanced to account for the special nature of red-green and blue-yellow as the primary dimensions of color appearance. As noted above, red may be special because it signals ripe fruit or blood. For blue and yellow, a number of studies have pointed out that these hues fall close to the daylight locus (e.g., variations between the sky and direct sunlight), and thus the tuning of the blue-yellow axis is matched to a dominant source of variation in the color environment [9]. In fact, this idea has been taken further to suggest that the special perceptual nature of some hues lies entirely in the environment. In opponent theory, blue and yellow (and red and green) are pure or unique hues because they arise from the pure isolated responses within the blue-yellow or red-green channels. However, a problem for this theory is that neurons with the chromatic tuning predicted by these sensations (e.g., that respond only to pure blue-yellow signals) have yet to be discovered. This has suggested the possibility that observers may learn to perceive the unique hues as special because they are special in the environment and not because they correspond to a unique pattern of activity within the neurons coding color. The number of basic color terms (color words that have high agreement and do not refer to specific objects) and the locations in color space that they denote have also been predicted from analyses of how colors are clustered in images. By this account, red spectra tend to form a distinct and distant cluster of signals, and this salient cluster may be why a prevalent color term for red arose in many of the world's languages [10].

Regardless of their neural basis, it is clear that like gray, the unique hues are not strongly tied to the specific spectral sensitivity of the observer. The wavelength that appears pure yellow shows little dependence on the relative numbers of long- and medium-wavelength-sensitive cones, even though these can vary over an enormous range of ratios. Similarly, the wavelengths chosen as unique hues cannot be predicted from individual differences in lens or macular pigment density and

like gray show little effect of the observer's age. These results could again suggest that the hues are set more by the environment. However, it is important to note that the stimuli perceived as unique hues also vary widely across observers and thus cannot be tightly bound to pronounced and stable properties of the world. Indeed, it remains debated to what extent and in what ways color appearance and color naming reflect properties of the observer, the physical environment, or the cultural environment [11].

A further potential tie of the appearance of blue and yellow to the environment is that, in a number of tasks, observers appear less sensitive to this color axis. For example, when individuals adjust a stimulus to appear white, they are much more tolerant of blue-yellow variations than red-green variations, and differences in the achromatic settings between observers also exhibit greater variance in blue and yellow. Similarly, color discrimination thresholds are (in some but not all cases) higher for blue and yellow. At suprathreshold, blue-yellow variations can have lower effective contrast (e.g., in a color search task), are rated as lower in visual discomfort (which correlates with perceived contrast), and have recently been found to induce weaker cortical responses as measured by functional MRI. One potential explanation for this bias against blue-yellow is that the world itself has a bias toward more blue and yellow. In particular, in many natural outdoor scenes, the variance in chromaticity is higher along bluish-yellowish axes [12]. The visual system adapts not only to the average color in the stimulus but also to the range or contrast. Adaptation to the higher blue-yellow contrasts in typical outdoor scenes may thus leave us less sensitive to these hues. Notably, this bias is also built into the structure of perceptually uniform color spaces (Fig. 1). Color order systems like the Munsell system or spaces like the CIE uniform chromaticity diagram are also stretched along bluish-yellowish axes (though these tend to be closer to an orange-cyan axis). Not surprisingly then, uniform spaces may mirror the structure of the color distributions we are adapted to and in turn might hold clues to the particular color environment we are calibrated for [13].



Environmental Influences on Color Vision, Fig. 1 Uniform color systems and natural image statistics. *Left:* Munsell colors plotted within a color space defined by the two color-opponent axes coded by the retinal: long-versus medium-wavelength cone signals (x axis) and short versus medium and long cone signals (t axis). Sets of points with the same shading represent contours of constant chroma. These are elongated in the bluish-yellowish direction, consistent with weaker sensitivity to these hue

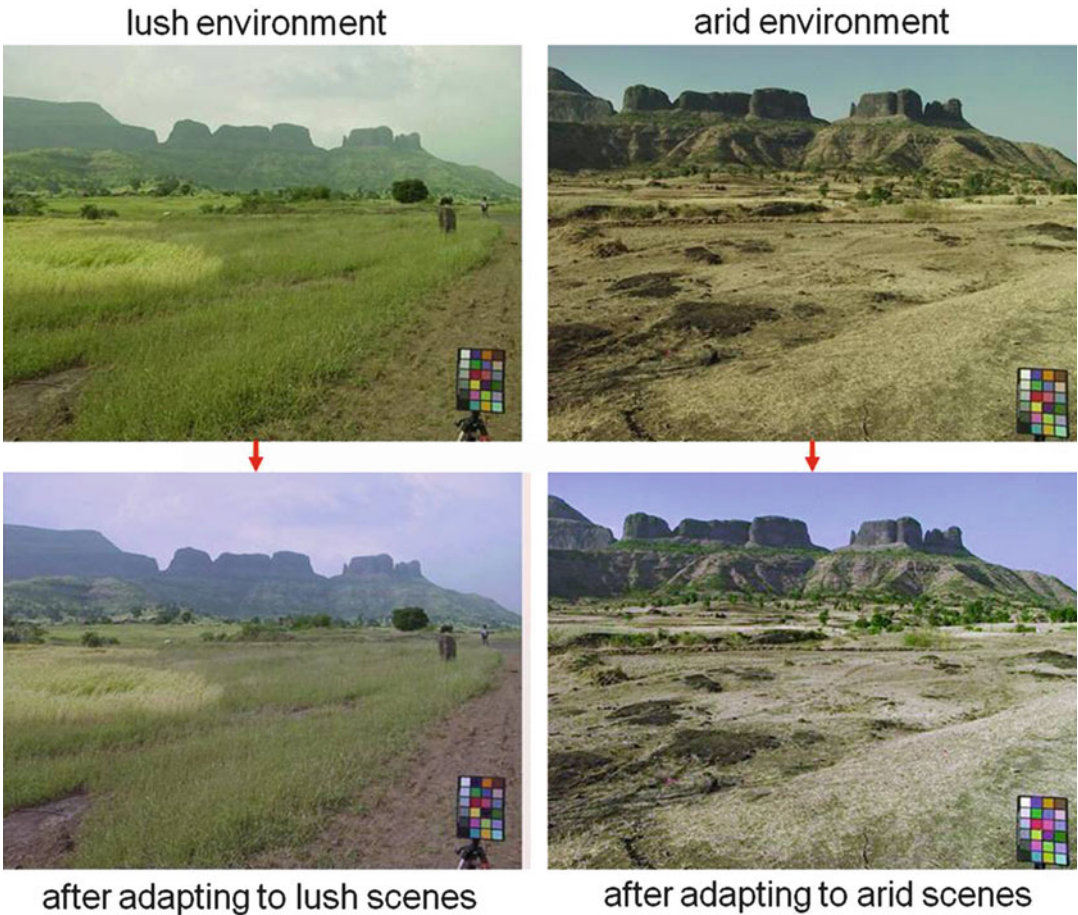
directions. *Right:* color distribution characteristic of natural outdoor scenes plotted in the same space. The scenes have a blue-yellow bias corresponding to the sky and earth. Note that the scaling is chosen to equate sensitivity to each axis, and this roughly matches the range of color along either axis. The weaker sensitivity to blue and yellow may be an adaptation to the greater blue-yellow variance in environmental color

Variations in the Color Environment

Clearly, the world is not stationary in its color characteristics and can instead vary widely from one location to the next or in the same place over time as the lighting or seasons change [12]. If observers are adapted to the colors they are immersed in, then we should expect variations in the properties of their color vision as we go from forest to desert or from summer to winter (Fig. 2). The effects of such environmental changes have not been widely studied, but there are telling signs that they occur. One example is variations in visual sensitivity that could arise from different levels of exposure to the phototoxic effects of sunlight. Exposure to ultraviolet light accelerates the normal yellowing of the lens with age and thus reduces sensitivity to blue. Brown and Lindsey noted an intriguing correlation between areas of the world with high UV-B (at lower latitudes) and the presence of languages which lacked a distinct basic color term for blue [14]. They suggested that populations in high UV-B areas would have more

deficient sensitivity at short wavelengths and would thus be less likely to experience “blue.” On the other hand, as noted above, measures of color appearance do not vary with normal aging of the lens, and thus adaptation may similarly compensate color appearance for these sensitivity losses.

A second example is when observers are placed in artificially colored environments. Neitz et al. exposed subjects for several hours at a time to “red” or “green” worlds by altering their room lighting or by having the subjects wear tinted contacts [15]. This led to persistent changes in the wavelength that appeared unique yellow, consistent with a long-term renormalization in color vision. Subtle differences in color appearance have also been found in studies comparing unique hues and focal colors (the best example of a given color term) across different populations [11]. However, a clear causal link between how much the world varies in color, and how much people in different parts of the world differ in their color perception, has yet to be established.



Environmental Influences on Color Vision, Fig. 2 Simulations of how the world might appear to observers adapted to different environments. *Top*: images of the same setting in wet (*left*) or dry (*right*) seasons. *Bottom*: the same images with the colors adjusted to simulate the effects of visual adaptation to the average color and to the range of colors in the different settings. Note this

shifts the average color toward gray and tends to highlight the more novel colors in the environment (Reproduced with permission from Juricevic, I. and Webster, M.A. (2009). Variations in normal color vision. V. Simulations of adaptation to natural color environments. *Visual Neuroscience* 26, 133–145)

References

1. Simoncelli, E.P., Olshausen, B.A.: Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–216 (2001)
2. Maloney, L.T.: Physics-based approaches to modeling surface color perception. In: Gegenfurtner, K., Sharpe, L. (eds.) *Color Vision: From Genes to Perception*, pp. 387–416. Cambridge University Press, Cambridge (1999)
3. MacLeod, D.I.A.: Colour discrimination, colour constancy, and natural scene statistics (The Verriest Lecture). In: Mollon, J., Pokorny, J., Knoblauch, K. (eds.) *Normal and Defective Colour Vision*. Oxford University Press, London (2003)
4. Buchsbaum, G., Gottschalk, A.: Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* **220**(1218), 89–113 (1983)
5. Mollon, J.D.: Tho' she kneel'd in that place where they grew. . . The uses and origins of primate colour vision. *J. Exp. Biol.* **146**, 1–38 (1989)
6. Webster, M.A.: Adaptation and visual coding. *J. Vis.* **11**(5), 3, 1–23 (2011)
7. Werner, J.S., Scheffrin, B.E.: Loci of achromatic points throughout the life span. *J. Opt. Soc. Am. A* **10**, 1509–1516 (1993)
8. Delahunt, P.B., Webster, M.A., Ma, L., Werner, J.S.: Long-term renormalization of chromatic mechanisms following cataract surgery. *Vis. Neurosci.* **21**, 301–7 (2004)

9. Shepard, R.N.: The perceptual organization of colors: an adaptation to regularities of the terrestrial world? In: Jerome, H., Cosmides, L., Tooby, J. (eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pp. 495–532. Oxford University Press, New York (1992)
10. Yendrikhovskij, S.N.: Computing color categories from the statistics of natural images. *J. Imag. Sci. Technol.* **45**, 409–417 (2001)
11. Webster, M.A., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R., Paramei, G., Dedrick, D. (eds.) *Anthropology of Color*, pp. 29–53. John Benjamins, Amsterdam (2007)
12. Webster, M.A., Mollon, J.D.: Adaptation and the color statistics of natural images. *Vision Res.* **37**, 3283–3298 (1997)
13. McDermott, K.C., Webster, M.A.: Uniform color spaces and natural image statistics. *J. Optic. Soc. Am. A.* **29**, A182–A187 (2012)
14. Lindsey, D.T., Brown, A.M.: Color naming and the phototoxic effects of sunlight on the eye. *Psychol. Sci.* **13**, 506–512 (2002)
15. Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M., Williams, D.R.: Color perception is mediated by a plastic neural mechanism that is adjustable in adults. *Neuron* **35**, 783–792 (2002)

ERP

- [Visual Evoked Potentials](#)

Event Related Potential

- [Visual Evoked Potentials](#)

Evolution

- [Environmental Influences on Color Vision](#)

Expressive Rendering

- [Non-Photorealistic Rendering](#)

Eye Sensitivity Functions

- [Spectral Luminous Efficiency](#)

F

Favorite Colors

► [Color Preference](#)

Fechner, Gustav Theodor

Mark D. Fairchild
College of Science, Rochester Institute of
Technology, Rochester, NY, USA



Gustav Theodor Fechner was a German experimental psychologist and philosopher. He is also

considered by many to be the father of modern psychophysics. Initially, Fechner took a degree in medicine and worked in that area for a while. During that time he began publishing a series of humorous and satirical articles and poems lampooning the medical profession. These were published under the pseudonym Dr. Mises, and one of his most famous publications of this genre was on the *Comparative Anatomy of Angels* (1825). The writings of Dr. Mises certainly provide insight into the breadth and depth of Fechner's intelligence and abilities.

Fechner then moved on to physics by learning about contemporary advances in electricity and magnetism through translation of great French works as well as handbooks of chemistry and physics into German. With this new knowledge in hand, Fechner taught at the University of Leipzig, eventually obtaining a professorship and making distinguished contributions to the field. Interestingly, Dr. Mises also continued to publish on occasion.

All along, Fechner was interested in the mind-body problem and desired to determine an empirical relationship between the physics of the world (and body) and the perceptions of the mind. Some of his studies in this vein included the discovery of perceived colors for slowly flickering black-and-white patterns (known as Fechner's Colors or Fechner-Benham Colors after Benham made the work accessible in English) and the detailed study of color afterimages. It was the study of afterimages that set Fechner off in his next direction. He

was studying afterimages by staring at the sun through colored filters. This led him to give up his chair in physics in 1840 due to induced photophobia from eye injuries that made him an invalid, and overly sensitive to light, for about a decade.

During this period, on October 22, 1850, while lying in bed he finally figured out the basis of linking physical measurements in the environment with human perceptions in the mind that is currently referred to as Fechner's law. Based on his knowledge of Weber's law (coined by Fechner) that, for many perceptions, the ratio of a just-noticeable change in a stimulus to the initial magnitude of a stimulus is a constant, Fechner figured out that the differential equation implied by Fechner could be integrated and assumed that the just-noticeable differences could be summed to predict perceptual magnitude. The resulting relationship, now termed Fechner's law, mathematically suggested that the magnitude of a perception would be proportional to the logarithm of the stimulus intensity.

More modern knowledge tells us that the specific relationship depends on the perceptual quantity and that it is not valid to sum JNDs to predict magnitudes. Nonetheless, Fechner's contribution was very significant in founding the field of psychophysics. Fechner published further details of the theory and practice of psychophysics in his seminal work, *Elements of Psychophysics* (1860), [1] which remains a useful guide for practitioners in the field. Interestingly, Fechner is also credited with being the first to introduce the concept of the median to formal data analysis.

From Boring's introduction in Adler's translation of *Elements of Psychophysics*, we find that Fechner was for 7 years a physiologist (1817–1824), for 15 years a physicist (1824–1839), for 12 years an invalid (1839–1851), for 14 years a psychophysicist (1851–1865), for 11 years an experimental estheticist (1865–1876), and for periods throughout a philosopher.

References

1. G. Fechner: *Elements of Psychophysics* (trans.: H. E. Adler). Holt, Rinehart and Winston, New York (1966)

Fechner's Colors and Behnam's Top

John H. Krantz

Natural Sciences, Department of Psychology,
Hanover College, Hanover, IN, USA

Synonyms

[Behnam's disk](#); [Fechner-Behnam subjective color](#); [Illusory colors](#); [Pattern-induced flicker colors](#); [Subjective colors](#)

Definitions

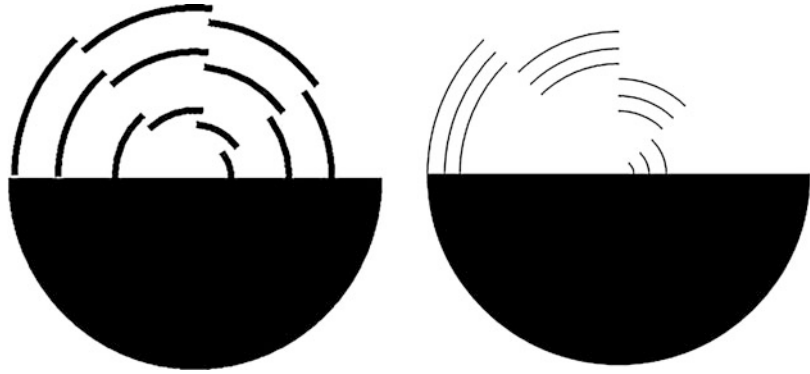
Fechner colors refer to the illusory colors that result from the repeated flashing of black and white patterns. These colors are often referred to as Fechner colors after their discoverer. *Behnam's top* is a pattern of spinning patterns developed for a spinning disk to create these illusory colors that have become a standard.

Basic Phenomenon and History

Fechner colors refer to the creation of illusory or subjective colors through an alternating pattern of black and white stimulus. Fechner reported the discovery of this illusion in an 1838 paper, and his name has been given to these illusory colors. The discovery was accidental as Fechner at the time was involved in the creation of disks to create different shades of gray [1]. From the beginning, a spinning disk was the means for creating these illusory colors. There were several studies throughout the nineteenth century of this phenomenon with spinning disks including studies by such a luminary as Helmholtz. Helmholtz even proposed one of the early explanations for these colors. Each of the early studies had used individually designed disks to create the illusory colors. In 1894, Behnam developed what has become the near-standard disk pattern of producing these illusory colors that he called the "artificial spectrum top" and has come to bear his name [2]. This disk

Fechner's Colors and Behnam's Top,

Fig. 1 Common versions of Behnam's top or disk



is called usually either Behnam's top or Behnam's disk. Figure 1 shows some examples of a Behnam's top. Variations in the disk include having thicker single lines or thinner three parallel lines.

To generate these illusory colors, it is important that the flicker rate of the stimulus is not too slow or too fast. The flicker is noticeable and the stimulus is well short of the critical fusion frequency. The rate of alternation of the black and white areas is not terribly fast with effective stimuli frequencies going from 3 Hz on up with optimal rotational frequencies for the spinning disk of about 4–6 Hz [3, 4]. The direction of the spinning seems to affect the colors experienced [10].

Creating and Measuring Fechner Colors

To create Fechner colors, the disk is usually spun in the frequencies reported above. Colors are experienced in bands located where the black lines will be flickering [5]. While Behnam's top represents a traditional means of creating the colors, researchers have developed many alternatives in trying to understand how these illusory colors occur [4, 7, 9, 10]. It is also not necessary that a spinning disk or top be used. A stationary flickering stimulus can be used [7] though there is some evidence that the colors are not as easily seen [9]. To see an online example that uses a flickering stimulus, the reader is referred to Krantz [8].

Measuring Behnam's top has been approached several ways over the years. The most common

methods use some form of color matching to determine the color experienced by the participant. Schramme [10] used a color matching paradigm where a mirror was used to overlay the color patch onto the spinning disk. Then participants tried to match this patch to the illusory colors they were seeing. Rosenblum, Anderson, and Purple [9] used Munsell color chips to make the color matches which served them well as they were interested if the illusory colors show some of the features of color matching errors seen in dichromats. Jarvis [7] used a binocular matching method where the illusory colors were presented in one eye and the matching color was presented in another eye. In all of these methods, the attempt was to try to determine the color perceived by the observer.

Results from various studies over the years have revealed several interesting features of Fechner colors. First, the colors experienced are always desaturated, that is, somewhat washed out [7]. These colors can vary in their vibrancy in ways that depend upon the method of generation, but natural colors can much more easily be made to be highly saturated. Many reports indicate a wide range of individual differences in the colors experienced [6]. While most visual phenomena have some individual variation, there is a remarkable degree of agreement across observers. If there were not this agreement, then technologies such as color reproduction would not be feasible. This agreement is not seen in Fechner colors. Other researchers have observed that both illumination level and adaptation state alter the colors

experienced, making the adaptation state of observers important to control in any experiment on Fechner colors [3, 4]. Perhaps one of the most interesting findings is that Fechner colors can be obtained with monochromatic, single-wavelength illumination as well as in full-spectrum neutral illumination [3].

Explanations and Future Directions

There have been many different types of explanations offered for the existence of Fechner colors in the almost 200 years since its discovery [1]. The challenge for any theoretical explanation is twofold: how does a monochromatic or neutral color stimulation lead to the perception of colors and what is the role of the flickering rate? Another important factor to consider, often ignored in most theoretical attempts, is the individual differences in the colors reported by observers. Older explanations have involved fatigue in receptors, the role of complementary colors, and even factors related to Hering's original theory of color perception. Most of these explanations have been ruled out by directly examining fatigue, contrast, and limitations of the complementarity of the colors experienced [1]. As more information about cone functioning developed, explanations developed that used ideas about differences in the speed with which different wavelengths of light are processed [3] and information about the relative speed with which the blue or short-wavelength cone responding to flickering stimuli compared to the speed of the other cones responding to the same flickering stimuli have been developed. The hypothesis is that it is the relative slowness of the short-wavelength cone in responding to flickering stimuli that is responsible for Fechner colors. The idea is that with slowly flickering stimuli, the short wavelength cone responds too slowly to pick up the flicker adequately, while the middle- and long-wavelength cones, red and green cones, respectively, still can respond. Thus, these cones do not adequately

cancel each other out in the blue-yellow color opponent channel leading to the perception of colors. Shramme [10] found, using color matches, Fechner colors that fell along the blue-yellow axis of the 1931 CIE diagram that matched these expectations. Despite the sophistication and intuitive plausibility of this explanation, it is still safe to say that there is no generally accepted explanation for these illusory colors. Campenhausen et al. [4] propose a two-step model for Fechner colors that involves both summing of cone inputs in a noncolor opponent pathway and lateral interactions at the level of the horizontal cells. The support for this hypothesis comes from many lines of evidence but includes that Fechner color in areas that do not receive modulated light and that achromatic effects similar to Fechner colors can be produced in rod vision. Shramme [10], whose results seemed to support the possibility that Fechner colors were due to the slowness of the short-wavelength cone, proposed that Fechner colors actually arise in the blue-yellow ganglion cells. The argument here depended upon the ability to manipulate the perceived Fechner color along the blue-yellow axis of the CIE diagram in a manner which would have been hard to predict from the action of the short-wavelength cone alone. Using these two explanations as examples, it can be seen that even the level of the visual system involved is not agreed upon, though all theories do seem to argue that Fechner colors arise in the retina.

While these explanations are intriguing, future developments will be needed to combine, modify, or seek a new direction of explanation for Fechner colors. One feature of the findings that is rarely explained and needs to be is individual variation in the colors experienced. The actual mechanism of the temporal response needs to be further uncovered as this finding increases the evidence for multiple ways that the eye responds to temporally moving or flickering stimuli. It is possible that it is in the variations in temporal response that the individual differences of Fechner colors lie.

Cross-References

- [Color Phenomenology](#)
- [Color Vision, Opponent Theory](#)
- [Cone Fundamentals](#)
- [Motion and Color Cognition](#)
- [Munsell, Albert Henry](#)
- [Psychological Color Space and Color Terms](#)

References

1. Bagley, F.W.: An investigation of Fechner's colors. *Am. J. Psychol.* **13**(4), 488–525 (1902)
2. Behnam, C.E.: The artificial spectrum top. *Nature* **51**, 200 (1894)
3. Brown, J.L.: Flicker and intermittent stimulation. In: Graham, C.H. (ed.) *Vision and Visual Perception*, pp. 251–320. Wiley, New York (1965)
4. Campenhausen, C.V., Hostetter, K., Schramme, J., Tritsch, M.F.: Color induction vs non-opponent lateral interactions in the human retina. *Vis. Res.* **32**(5), 913–923 (1992)
5. Falk, D., Brill, D., Stork, D.: *Seeing the Light*. Harper & Row, New York (1986)
6. Gerjuoy, H., Clarke, F.R.: Fechner colors on television. *Am. J. Psychol.* **71**(3), 606–607 (1958)
7. Jarvis, J.R.: On Fechner-Benham subjective color. *Vis. Res.* **17**, 445–471 (1977)
8. Krantz, J. H.: Behnam's top and Fechner colors. <http://psych.hanover.edu/krantz/BenhamTop/> (1999). Accessed 6 Jun 2012
9. Rosenblum, K., Anderson, M.L., Purple, R.L.: Normal and color defective perceptions of Fechner-Benham colors: implications for color vision theory. *Vis. Res.* **21**, 1483–1490 (1981)
10. Schramme, J.: Changes in pattern induced flicker colors are mediated by the blue-yellow opponent process. *Vis. Res.* **32**(11), 2129–2134 (1992)

Fechner-Benham Subjective Color

- [Fechner's Colors and Behnam's Top](#)

Filling-In

- [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)

Finish, Textile

Parikshit Goswami

School of Design, University of Leeds, Leeds,
West Yorkshire, UK

Synonyms

[Dry or mechanical finishing](#); [Textile processing](#);
[Textile wet processing](#); [Wet or chemical finishes](#)

Definition

Textile finishing can be defined as all processes (chemical and/or mechanical), employed subsequent to textile coloration which impart additional functionality/superior aesthetics to the textile material. Mostly, textile finishing is applied to fabrics (woven, knitted, nonwovens); however, textile finishes can also be applied to fibers and yarns.

Introduction

The primary objectives of textile finishing are to improve the aesthetic and functional properties of textiles. In a broader sense, “finishing” relates to all processes that fabrics might be subjected to subsequent to weaving, knitting, or nonwoven manufacturing processes. The term “finishing” could include fabric preparation (e.g., singeing, desizing, scouring, bleaching, optical brightening process, mercerization process, etc.) and coloration (dyeing and printing); indeed, the combination of these processes is sometimes referred to as “wet-finishing” processes [1]. Another school of thought refers to “finishing” as the final stage of fabric preparation with the objective to prepare the fabric for the consumer and that the term “finishing” concerns only fabric treatment other than fabric preparation and coloration. In this chapter, finishing subsequent to coloration will

be discussed, in which context, it is important to check the compatibility of the finishing process with the treated substrate and the coloration process that has been previously implemented [2].

Classification

The conventional textile finishing methods can broadly be divided into two categories:

- (i) “Dry” or “mechanical” finishes
- (ii) “Wet” or “chemical” finishes

There are, however, finishing operations which combine mechanical and chemical finishes e.g., mercerization (the NaOH treatment of fabric on machines with or without tension).

“Dry” or “Mechanical” Finishes

The mechanical finishing of textiles may range from a simple drying operation to a complicated series of calendaring operations (Schreiner finisher). A few examples of the mechanical finishing of textile are discussed below.

Drying

It is well known that wet fabric can hold a large amount of water and that some of this held water can be easily removed by simple mechanical action. The efficiency of the process to mechanically remove the water from the fabric is vitally important from both cost and environmental points of view. High energy is required to evaporate the remaining water due to its high latent heat of evaporation. Machines such as mangles, centrifuges, and suction-slot machines are used in textile processing to remove water from fabric by mechanical means [3]. Drying machines are employed in textile processing with the purpose of removing the water that has not been removed by mechanical action. The majority of the dryers used in textile processing offer continuous throughput and a few examples of such machines are “hot flue”, “drum dryers”, “perforated drum dryers”, “tumble dryers”, etc. [1].

Calendering

In calendering, the fabric is passed between two heavy rollers. The rollers may vary in hardness, surface speed, and temperature. Fabric properties such as smoothness and luster can be improved using calendering [4]. Calender finishing can be of six types, namely, simple finishing or swizzing calendering, chasing calendering, friction calendering, Schreiner calendering, embossing calendering, and felt calendering. A few examples of calendering are discussed below.

Friction Calendering

The differential speed of the two rollers is carefully selected to produce effects such as “chintz.” The smooth metal roller normally rotates at a higher rate than the soft roller [1, 3]. Durable finishes (Everglaze) can be achieved using this method by adding a cross-linking agent [5].

Embossed Effects

In this technique, a fabric pattern is produced by calendering the fabric using an engraved heated metal roller and a soft roller [4]. To avoid slippage between the rollers, it is important to control the process accurately.

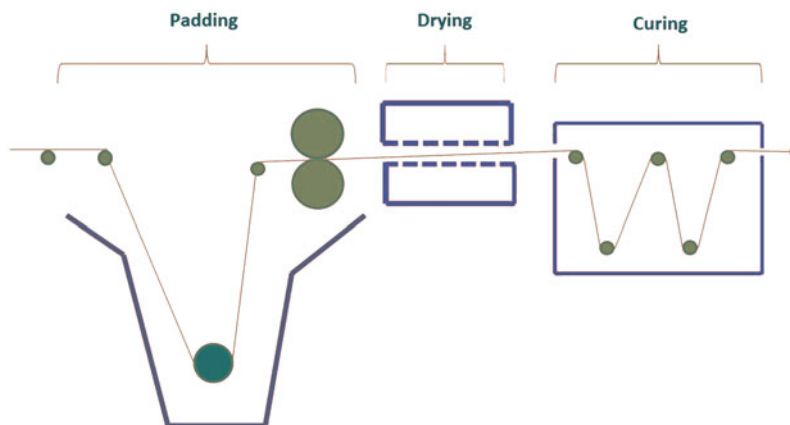
Schreiner Finishing

In “Schreiner,” fine lines engraved on a metal roller are transferred to the fabric. With appropriate fabric construction and by carefully engineering the line direction of the engraving, a soft lustrous handle can be achieved [3]. This treatment is used to give cotton fabric the appearance of silk [1]. Although the wash fastness of this effect is poor [1, 3], it increases the point-of-sale appeal [3].

Napping

Napping is a very effective way for imparting a soft handle to fabrics [6]. Fabrics of low-twist, staple fiber yarn can be used for napping [7]. The napping machine contains metal cards and the napping tool comprises rollers with curved metal wires [1] which pull the fiber ends to the surface of the fabric. One or both sides of the fabric can be napped [7].

Finish, Textile,
Fig. 1 Schematic diagram
 of pad-dry cure operation



F

Sanforizing

Fabrics contain internal tensions as a result of the various manufacturing processes employed, and a pretreatment is required to nullify this effect, since, otherwise, there is a possibility of significant shrinkage of garments during washing. Sanforizing is a controlled mechanical shrinkage process without the use of any chemicals. This treatment should be the last treatment that is applied to the fabric [1].

“Wet” or “Chemical” Finishes

Although chemical finishing has always been an important process for textiles, the demand for high-performance textiles has grown exponentially in recent years and, as a result, so has the demand for chemical finishing [8]. The majority of chemical finishes for textiles are additive, where the finishing treatment results in an increase in substrate mass through sorption or deposition of chemical compounds. However, some chemical finishing can be *subtractive*, in which the treatment results in the fiber substrate losing mass as a result of chemical degradation. While chemical finishing is mainly applied to fabric, it can also be applied to loose fibers, yarns, garments, etc. Based on the life-span of the finish on the textile substrate, textile finishing can also be classified as *transient* and *durable* finishing. The most important aspect in this regard is the resistance of the finish to domestic washing [9].

Application of Chemical Finishes

While the application and formulation of chemical finishes depends on several factors, the most important factors are the composition of the material being treated and the chemistry of the functional chemical. Interaction of the finishing effects, compatibility of different chemicals, and the environmental/green credentials of the process are other factors to consider.

If the functional chemical finishes displays high substantivity towards the substrate, then exhaust (immersion) application methods can be used and any of the textile dyeing machines employed for batchwise dyeing can be used for textile finishing. However, if the applied functional chemical has limited or low substantivity, then a continuous application method can be used. One such method is the pad-dry-cure method wherein the fabric is first immersed in a solution containing the functional chemical, followed by drying and finally “curing” to fix the finish within the fabric (Fig. 1).

Chemical finishes can also be applied to textiles using modern application techniques such as exposure of the textile substrate to plasma employing either low pressure, low-temperature plasma, or atmospheric plasma. Plasma, which is considered as the fourth aggregation state of matter, can be described as a mixture of partially ionized gases in which the constituents are achieved by external energy addition. The main advantage of such plasma treatments are:

- It consumes minimal chemicals compared to conventional finishing.
- No costly drying process is required.
- It is a surface treatment, so the bulk properties of the material are not affected.
- High environmental compatibility.

Important Chemical Finishes

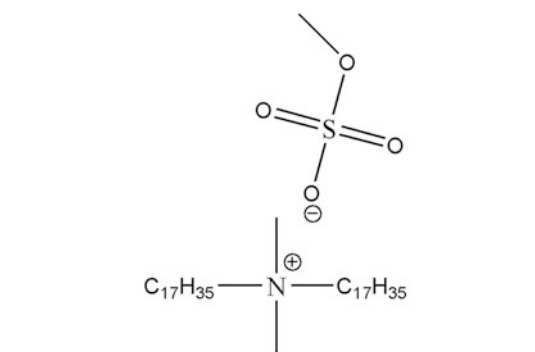
A few commercially important chemical finishes are discussed below. Other types of chemical finishes include, but are not limited to, antistatic finishes, soil release finishes, insect resist finishes, flame-retardant finishes, stain resist finishes, non-slip finishes, UV-protection finishes, finishing to improve color fastness, etc.

Softening Agents

Softeners are one of the most widely used chemicals in textile processing, and as the name suggests, the application of softeners dramatically improves the handle and perceived quality of a fabric. In textile processing, one of the most important functions of softeners is to counteract the harshness imparted by other finishes (e.g., easy care). Softeners are, however, also used extensively in domestic washing formulations.

Although the main effects of the softeners are at the surface of fibers, the small molecules can impart internal plasticization of the fibers by reducing the glass transition temperature (T_g) of the fibrous polymer [8]. Early softeners were based on waxes and oils although modern types can be divided into five categories, namely, cationic softeners, nonionic softeners, anionic softeners, reactive softeners, and amphoteric softeners [3].

Cationic softeners are the most important type of softener for both industrial and domestic applications. The positively charged ends of the cationic softeners orient themselves towards the negatively charged fibers (zeta potential), creating a new surface of hydrophobic carbon chains that are responsible for the excellent softening effect of cationic softeners [8]. Because of the positive charge, these softeners are substantive towards cellulosic fibers, and exhaust application of these finishes is widespread using conventional textile dyeing machines (e.g., winch, jet, beam)

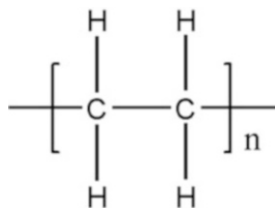


Finish, Textile, Fig. 2 Dimethyldistearylammonium methosulfate

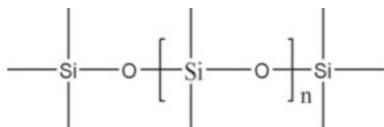
[3]. Cationic softeners are known to exacerbate the soiling propensity of fibers and to inhibit soil removal.

Dimethyldistearylammonium methosulfate is a good example of a popular cationic softener (Fig. 2). The use of silicone-based finishes in textile processing has grown steadily over the last 50 or so years [3]. Silicon-based cationic softeners provide a high degree of softness and a unique hand to fabrics. However, silicon softeners may contain variable amounts of siloxane oligomers, depending on the method of synthesis, which may be a cause of air pollution [8]. A popular range of silicon softeners are based on the weakly cationic amino functional siloxanes. Although these silicon compounds could potentially be applied from mild alkaline conditions, durability is improved by application at pH of 4.

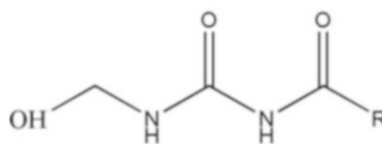
Nonionic softeners are well known to perform multiple functions such as softeners, emulsifiers, stabilizers, extenders, and lubricants [3]. Paraffin waxes and similar materials are the most basic nonionic softeners. Molten polyethylene (Fig. 3) at high pressure can be air oxidized to secure hydrophilic characteristics (mainly carboxyl groups). High-quality, more stable products can be produced by emulsification in the presence of alkali. These products are stable to extreme pH conditions and can withstand the temperatures used for normal textile processing [8]. In recent years, polysiloxanes have gained importance as nonionic softeners (Fig. 4). Due to their limited substantivity, nonionic softeners, with very few



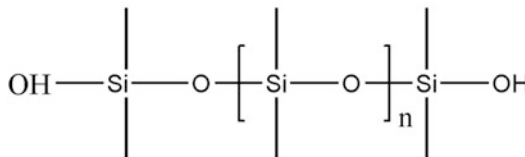
Finish, Textile, Fig. 3 Polyethylene



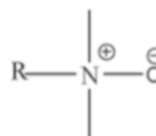
Finish, Textile, Fig. 4 Polydimethylsiloxane



Finish, Textile, Fig. 5 *N*-methylol derivative (R: C₁₇H₃₅)



Finish, Textile, Fig. 6 Compound containing silanol group



Finish, Textile, Fig. 7 Alkyldimethylamine oxide softener (R: long alkyl chain)

exceptions, are normally applied by non-exhaust methods (e.g., padding) [3].

Anionic softeners, due to the presence of the negative charge, are repelled from negatively charged fibers (e.g., cellulosic fibers) which leads to higher hydrophilicity [8]. These were among the first soft finishes to be used commercially and include long-chain alkyl sulfates, sulfonates, and succinates. Anionic softeners are used in specialized areas of application where the physiological activity is low, e.g., medical textiles. As the majority of fluorescent brightening agents are anionic, they are widely used in conjunction with anionic softeners for the resin finishing of white cellulosic fibers [3].

Reactive softeners contain functional groups capable of reacting with particular functional groups in some fibers [8] (e.g., the -OH group of cellulosic fibers). This type of finishing is permanent to washing due to the formation of covalent bonds between the softener and substrate. *N*-methylol derivatives of stearic acid amides and urea-substituted compounds are very successful as reactive softeners (Fig. 5) [3]. One example of a silicon-based reactive softener is given in Fig. 6, which is a modified siloxane and contains functional silanol groups. The silanol group could potentially be replaced by other functional groups such as amines or alcohols.

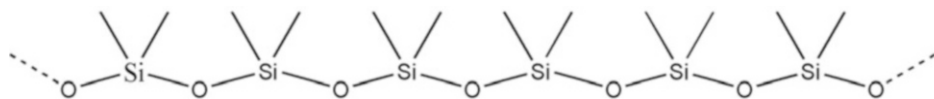
Amphoteric softeners have limited textile applications but are very popular in personal

care products (e.g., shampoo formulations) due to their low toxicity [3]. One example of an amphoteric softener is shown in Fig. 7.

Water and Oil-Repellent Finishes

There is a distinction between “water-repellent” and “waterproof” fabrics. Waterproofing is achieved, for example, by coating the fabric with rubber. Such a treated fabric will not only be impermeable to water but also most notably against air and perspiration; as such, “waterproof” fabrics are uncomfortable to wear. A water-repellent finish, on the other hand, remains permeable to air and is achieved by the application of hydrophobic chemicals to the fabric [1]. These types of “water-repellent” finishes are normally used for clothing and will be discussed in this chapter. As very few textiles are inherently water repellent but none are oil repellent, so an additional process must be added to display these properties.

If the internal cohesive interactions within a liquid are lower than the adhesive interactions



Finish, Textile, Fig. 8 Polydimethylsiloxane

operating between a solid surface and the liquid, then a drop of the liquid will spread on the solid surface. Based on this theory, repellent finishes (both oil and water) work on the principle of reducing the free energy of the fabric surface [8]. In this context, it should be mentioned that alongside the chemical composition of the material, the geometry of the textile surface also plays a significant role in the wettability of the textile [1]. It has long been recognized that superhydrophobic surfaces require a unique combination of two fundamental properties, namely, surface roughness and low surface energy [2]. The functional chemicals used to achieve water-repellent finishing significantly vary in their chemistry, their role being to reduce the surface energy of the fabric by adding lower surface energy chemical groups to the surface [3].

The chemicals traditionally used to achieve water repellence may be divided into a few groups: metal salts, soap/metal salts, wax, pyridinium-based finishes, organometallic complexes, *N*-methylol derivatives, silicone finishes, and fluorochemical finishes [3]. Among these, only fluorocarbon finishes are known to repel both oil and water [3, 12], whereas the other finishes can repel only water. Moreover, perfluorinated derivatives are effective at very low concentrations [1]. Polymeric fluorochemical finishes are typically acrylic or methacrylic polymers with perfluoro side chains. Silicon-based water repellents are also used (Fig. 8).

Easy-Care and Durable Press Finishes

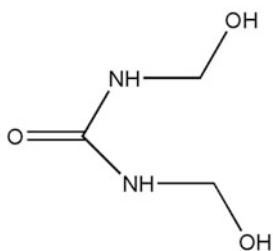
As the name suggests, “easy care” and “durable press” are applied to textiles to impart minimum care properties, mostly to cellulosic fibers and their blends with other fibers. Various terms have been used to describe this area, including easy care, easy-to-iron, no iron, crease resistant, wrinkle resistant, wrinkle free etc. However the

technically correct term is “cellulosic anti-swelling” or “cellulosic cross-linking” finishes [8].

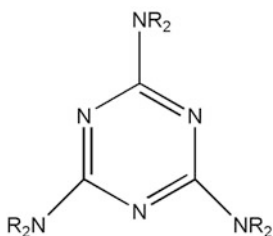
Cellulose (natural or regenerated) is a linear polymer formed by the condensation of β -D-glucopyranose that contains 1 \rightarrow 4-glycosidic linkages. The presence of hydroxyl groups along each chain creates extensive H-bonding both between (intermolecular) and within (intramolecular) the chains. Formation of such intramolecular H-bonds imparts “stiffness” to the cellulose molecules by restricting movement of the 1,4- β -D-glucopyranose units. The -OH groups are also responsible for the hydrophilicity of cellulosic fibers; owing to the high density of the -OH groups, cellulosic fabrics shrink in water and crease upon drying. The main function of easy-care/durable press finishing is to overcome shrinkage and wrinkling by cross-linking the -OH groups.

Although finishing agents are unable to penetrate the crystalline regions of the fiber, the easy-care/durable press finish needs to be of small M_r and preferably high reactivity to enable it to penetrate the amorphous region of the fibers [3].

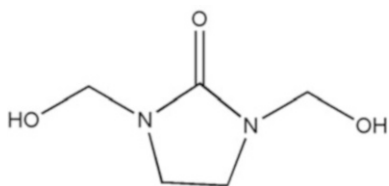
It was demonstrated in the late 1920s that thermosetting resins (e.g., phenol-formaldehyde, urea-formaldehyde) could impart crease resistance to cellulosic fabrics. Since then, numerous publications and patents have described novel cross-linking reagents based on urea-HCHO chemistry [4]. This original type of easy-care products (urea-HCHO resins) used to contain high free formaldehyde, e.g., dimethylolurea (DMU) which is prepared from an excess of HCHO (Fig. 9). These nonreactant types of finishes are only durable to washing temperatures up to 60 °C. Wash fastness can be improved by using reactive resins such as methoxymethylmelamines (Fig. 10).



Finish, Textile, Fig. 9 Dimethylolurea

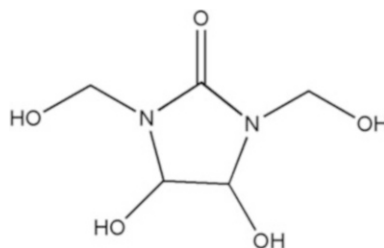


Finish, Textile, Fig. 10 Methoxymethylmelamines
($R=CH_2OCH_3, CH_2OH$)



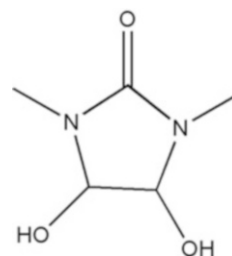
Finish, Textile, Fig. 11 Dihydroxymethyl ethylene urea

The first cyclic urea-reactant finishing agent was dimethylol ethylene urea (DMEU) which displays poor fastness to chlorine bleaching and adversely affects the light fastness of the finished fabric (Fig. 11). Cyclic urea-based finishing agents, *N,N'*-dimethylol-4,5-dihydroxyethylene urea (DMDHEU), comprise ~90 % of the easy-care/durable press finish products that are available in the market [8] (Fig. 12). Due to the presence of two hydroxyl groups at the C-4 and C-5 positions, the reactivity of DMDHEU is low, and therefore, reaction requires a catalyst. Although DMDHEU-containing products contain >0.3 % free formaldehyde, the search for formaldehyde-free finishing agents has spanned several decades [3] due to the toxicological effects associated with formaldehyde. Textiles containing a high level of



Finish, Textile, Fig. 12 *N,N'*-dimethylol-4,5-dihydroxyethylene

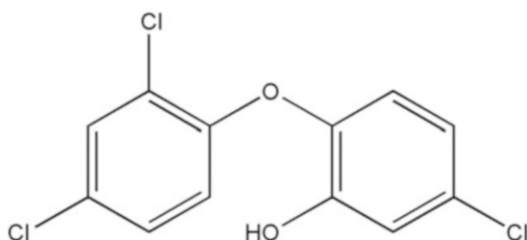
Finish, Textile, Fig. 13 *N,N'*-dimethyl-4,5-dihydroxyethylene urea



formaldehyde can give rise to eczema and allergic reactions; furthermore, HCHO is a suspected human carcinogen [8]. Expensive formaldehyde-free finishing agents such as *N,N'*-dimethyl-4,5-dihydroxyethylene urea (DMeDHEU) (Fig. 13) is used for textile processing. DMeDHEU is less reactive than DMDHEU and therefore requires a stronger catalyst and harsher reaction conditions for successful cross-linking with cellulosic fibers.

Antimicrobial Finishes

Antimicrobial finishes are used to inhibit the growth of or to destroy microscopic organisms [3]. This could be for hygiene purposes, i.e. to protect the user from pathogenic or odor-causing microorganisms or to protect the textile material itself from damage caused by mold, mildew, or rot-producing microorganisms. Formaldehyde is a widely used biocide and preservation product. As discussed in the previous section, the majority of easy-care/durable press finishes contain HCHO, and therefore, these finishes display a small antimicrobial effect. However, for effective antimicrobial action, specific chemical finishes are used. Traditional antimicrobial agents such as copper naphthenate, copper-8-quinolate, and numerous organomercury compounds are strictly



Finish, Textile, Fig. 14 2-(2,4-dichlorophenoxy)-5-chlorophenol

regulated because of their toxicity and potential for environmental damage [8].

A common antimicrobial agent extensively used for cellulosic material is polyhexamethylene biguanide (PHMB) which has long been used in applications such as cosmetics and swimming pools. Along with conventional textile applications, PHMB can also be used in the production of medical textiles. PHMB is cationic and therefore displays excellent durability on cellulosic material [3]. Triclosan (Fig. 14) is another antimicrobial agent that is used extensively in mouthwashes, toothpastes, deodorants, and on textiles. This is nontoxic to humans and used as durable antimicrobial finish on polyester and polyamide fibers and their blends with cotton and wool. Heavy metals such as silver, copper, and mercury can provide antimicrobial effect in the form of the metal or metallic salt. Antimicrobial high-performance textile fibers (e.g., polyester, nylon, etc.) can be produced that include nanoparticles of heavy metals (e.g., silver). Chitosan is a natural biodegradable polymer which is nontoxic and shows microbial resistant.

Conclusions

As a result of increasing demands for superior quality and functionality in textiles, the sophistication of textile finishes and finishing operations has increased. New types of finishing are being developed to provide novel effects such as fragrance finishes, well-being finishes, bionic finishes, as well as finishing for smart textiles, medical textiles etc. Parallel to the search for novel finishes, existing finishes (and application techniques) are

constantly being scrutinized to ensure that processing is as efficient as possible from technical, economical, and ecological perspectives. Textile finishing can be expected to remain a highly valuable tool which can significantly enhance the value of the finished textile product.

Cross-References

- [Colorant, Textile](#)
- [Coloration, Textile](#)
- [Dye](#)
- [Dye, Functional](#)

References

1. Fiscus, G., Grunenwald, D.: Textile Finishing: A Complete Guide. Editions High Tex, Sausheim (1995)
2. Marsh, J.T.: An Introduction to Textile Finishing, 2nd edn. Chapman & Hall, London (1966)
3. Haywood, D. (ed.): Textile Finishing. Society of Dyers & Colourists, Bradford (2003)
4. Vigo, T.: Textile Processing and Properties: Preparation, Dyeing, Finishing, and Performance. Elsevier, New York (1994)
5. Jackman, D., Dixon, M., Condra, J.: The Guide to Textiles for Interiors. Portage & Main Press, Manitoba (2003)
6. Carty, P., Byrne, M.S.: The Chemical and Mechanical Finishing of Textile Material, 2nd edn. Newcastle upon Tyne polytechnic products Ltd, Newcastle upon Tyne (1987)
7. Hargrave, H.: From Fibre to Fabric: The Essential Guide to Quilting Textiles. C & T Publishing, Lafayette (2008)
8. Schindler, W.D., Hauser, P.J.: Chemical Finishing of Textiles. Woodhead Publishing Limited, Cambridge (2004)
9. Tesoro, G.C.: Textile finishes. J Am Oil Chem Soc **45**(5), 351–353 (1968)
10. Daoud, W.A., Xin, J.H., Tao, X.: Superhydrophobic silica nanocomposite coating by a low-temperature process. J Am Ceram Soc **87**(9), 1782 (2004)
11. Balu, B., Breedveld, V., Hess, D.W.: Fabrication of “roll-off” and “sticky” superhydrophobic cellulose surfaces via plasma processing. Langmuir **24**, 4785 (2008)

Fitting (Anglo-Saxon English)

- [Luminares](#)

Fixture (American English)

- [Luminaires](#)

Flame Light Lamps

- [Combustion Lamp](#)

Focal Colors

- [Unique Hues](#)

Forecast

- [Color Trends](#)

Four-Dimensional Color Vision

- [Tetrachromatic Vision](#)

Functionality of Color

Antal Nemcsics¹, Malvina Arrarte-Grau² and Galen Minah³

¹Department of Architecture, Budapest University of Technology and Economics, Budapest, Hungary

²Architecture, Landscape and Color Design, Arquitectura Paisajismo Color, Lima, Peru

³College of Built Environments, Department of Architecture, University of Washington, Seattle, WA, USA

Synonyms

[Color effects to humans and the environment](#);
[Color roles](#); [Color uses](#)

Definition

The use, informative, and aesthetic impact of color on humans and the environment.

Color and Function

Human environment may be considered as a system, since man and environment elements are mutually determinant in any sense – an interrelation concerned with demand of man for his environment. The function of the built environment is based on social demand, necessity raised to a social level.

Within the system of man and elements of his environment, structural relations are defined and sustained by complex functions. This complex is composed of three types of functions: utility, aesthetic, and informative ones. Utility function is understood as the designation and purpose of environment elements. Aesthetic functions bring about properties of environment elements which enable one to experience utility functions. Informative functions involve properties of environment elements by which their functions, uses, and operation become understandable for man. These preliminaries underlie the present day approach to environment color design, considering the totality of utility, aesthetic, and informative functions as the system of demands.

► [Environment color design](#) has to serve the expression of the complex function of environment elements. Various functions of color-bearing environment elements are strictly interrelated and are prone to change into each other, a characteristic to be considered in the methodology of environment color design as a design process. It is necessary to develop design methods suitable for simultaneous and differentiated consideration and coordination of these components [1–3].

Color function components of environment color design as a design process can be deduced from the man to color relations. The man to color relation cannot be regarded in an abstract way. Colors are always associated to some object, phenomenon, or process, implicating into this relation their complex functions. In this respect, the theory

of environment color design has been concerned with the possibility to express function by the color of the environment element. It can be stated, for instance, that more saturated colors of the color space, with longer dominant wavelengths, act dynamically, hence suit to express functions involving dynamism. The less the saturation of these colors, the lower their dynamism. The greater the ► [hue](#), saturation, or ► [lightness](#) differences between members of color complexes to express a function, the more dynamic is the function expressed by them. The intensity of the expression of function is more affected by the variation of lightness differences than by saturation differences. Again, it is found that a function is expressed not so much by a single color but by a complex of several colors. Accordingly, the expression of function by the environment elements contributes to ► [harmony](#) relations of colors bore by the element, and Coloroid color harmony relations may be combined with function expression indices [3, 4].

Disclosure of the rules of function expression cannot rely solely on visual information, since it is physiologically possible that stimuli perceived by one of the sensory organs should create perceptions normally transmitted by another sensory organ.

Utility Function of Colors

Environment is the space for human activities aimed at satisfying human demands. Since the beginnings of society, architecture has been expected to meet human demands. Analysis of and reckoning with functions necessarily took place in every period, even if functional demands have significantly changed and developed through the ages. The scientific analysis of functions and a conscious, integrated attention to all the requirements could not, however, develop earlier than in contemporary age, at a higher level of social development. This has led from early functionalism, the narrow interpretation of function, to an extended, wide-range functional approach, to the complex consideration of the integer system of relations inclusive of demands, requirements, and the color environment.

Of course, this functional approach itself is undergoing development, although in its germs it has been present in every stage of development throughout the history of architecture and, recently, in ► [environment color design](#). Nowadays, however, it has become more conscious, unambiguous, and scientific, and it has risen to be an important factor of the up-to-date approach. The relationship between structure and function in architecture has been recognized, and this knowledge unavoidably affected environment color design. The ideas of environment color designers have to be directed by modern scientific thinking.

Activities in various areas of the built environment have become extremely differentiated. Functions have become so manifold that categorization is not only difficult but may lead to misunderstandings. Nevertheless, functions need to be categorized; otherwise the relations between functions and colors cannot be delineated, and the role of colors in the utility function of the environment cannot be studied.

According to the activities accommodated, the environment can be divided into working spaces, community spaces, living spaces, and traffic spaces.

To demonstrate the usable function of colors in different institutions, hospitals will be dealt with in the next paragraphs. Functions of hospital buildings are very special. For some people, a hospital provides accommodation for a shorter or longer but nevertheless limited time, while for others it is a working place. Therefore, the demands of both groups have to be respected not only concerning construction and equipment but also coloration. The more so since these demands are sometimes contradictory [2, 5, 6].

In most cases, the state of mind of a patient is a priori somewhat abnormal. The lonely, worried patient, so to say forcedly dragged from his habitual environment, is in need of being addressed and protectively accommodated by his new environment. On the other side, the work of the medical staff requires a great deal of concentration, and excessive color effects may distract their attention or even disturb his work. These anomalies have led to the absurd dispute among color dynamists

as to whom should priority be given in the hospital ward.

In a hospital, honoring of the utility functions can be done according to different aspects. Coloration should help orientation. Applying a certain color for doors of rooms with identical or similar functions may eliminate much uncertainty. Doors not for patients may be colored so as to convey this information. Proper color design may express connections between rooms. Any obtrusive color sensation has to be avoided in examination rooms and surgeries. In bathrooms, natural color effects of the sun, sand, and water have to be mimicked. The patient finds his picture in the mirror pale and sickly if it is being colored by near strong reflections from a green surface. A similar pale-colored impression arises when the patient observes his face neighbored by fiery orange or fiery red colors. But in rooms frequently contaminated by blood, green should be applied as complementary to red. Ward walls possibly need cheerful, warming, approaching, but not very saturated colors. Eventual chromotherapeutic effects have to be exploited. A febrile condition in itself yearns for a cool environment. In a condition of depression, a multicolored environment has to be preferred. In a labor ward or an intensive care unit, a positively reassuring coloration has to be applied.

Aesthetic Function of Colors

Just as any work of art, every element of the environment, and the environment itself, is an inseparable unity of content and form. Just as the human environment is a space for human functions, environment elements serve somehow the realization, accomplishment, and completion of human functions, and by all that they fulfill their own function. Thus, the human environment is up to its aesthetic function if it expresses its utility function in conformity with unity of content and form – where content is the utility function and form is the shape and color of the elements [1, 3, 5].

The essential condition of an aesthetic effect is unity between content and form, as so defined. Furthermore, the objects forming the content of the environment, or indeed the entire environment, are essentially functional. It follows that

the integral expression of content can be realized by the harmonious functioning of all the components of the entire space. The practical-utility and spiritual-mental components of function are interdependent. The mental arises, and is inseparable, from practical. Now it is not realistic any longer to believe that the aesthetic form of an object or space should be possible without knowing its function. Neither is it believed that there exist universal aesthetic formulae applicable in every case nor that it is possible to force upon the environment any fine ► [color harmony](#) without knowledge of its function. In environment color design, it is also necessary to consider how much importance is attributed to its practical functions from the point of view of human life as a whole. That is to say, any work and activity has its emotional, mental, and spiritual attachments. Therefore each object, tool, or space claims some of these attachments, depending on its role, importance, and function in human life.

Form becomes a necessity in the course of a “function experience,” imprinting on consciousness a harmony sensation as inseparable unity between content and form. Of course, the sight of a color complex may give aesthetic pleasure, but when separated from the content and function of the space, this pleasure is rather superficial, as it does not offer a full space sensation.

To be able to create a form expressing the message of the built environment, the designer has to be acquainted with relationships between environment structures, the so-called composition relationships, such as that of material-function-form, conditions of space effects, or the role of space-time-motion in an environment.

Informative Function of Colors

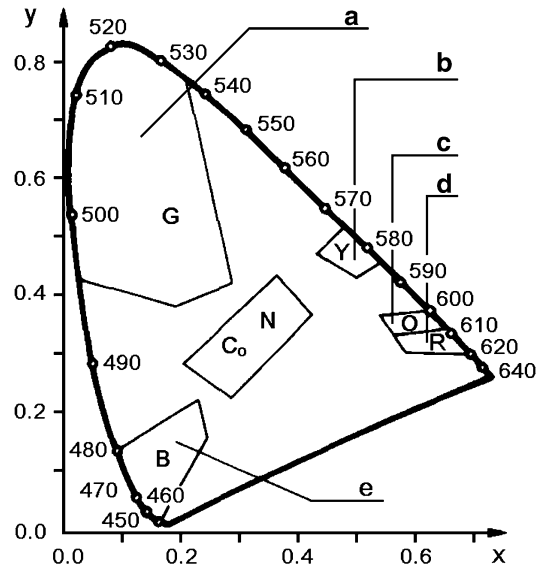
Information functions of environment serve to interpret for man the functions, the ways of utilizing, and the operation of the environment and its elements. An important part of the informative functions of the environment is expressed by color information. Depending on the message, color information may be interpreted as logic information or as artistic or aesthetic information.

Logic and aesthetic information are carried by the same elements, but a different structure

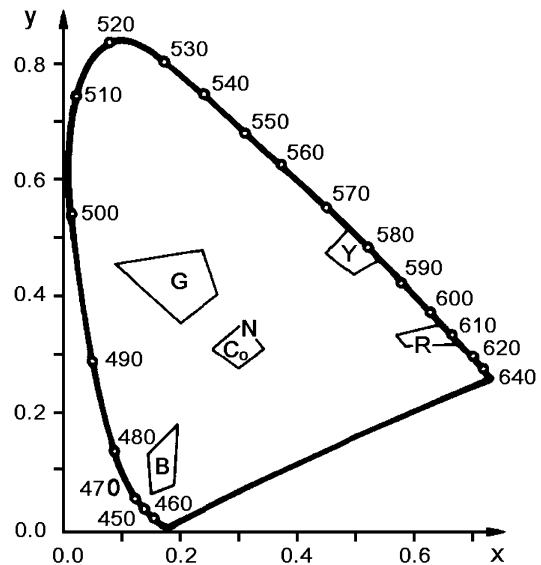
belongs to every form of message. They can be characterized partly by their different visual systems, by the differentiation of their complexity and structure, and partly by the psychical differences of their messages. A message is transmitted by highlighting, contracting, and grouping some visual codes in the information-bearing surface or space, while disregarding others. A group of colors attracts attention when it excels by regularity and its structure is well recognizable. Recently, the analysis of these relations has been tackled with the methods of semiotics. Although these studies mostly concerned other than visual structures, the results can also be applied to color dynamics.

Logic information is conveyed by standard codes; they are practical and strictly mental and transmit knowledge. They prepare decisions of receivers and control their behavior and attitudes. Logic information is transmitted, e.g., by internationally agreed safety color signals. Locations of these colors in the CIE diagram are seen in Fig. 1. For instance, green means information, orange warning, red prohibition, and blue instruction. A special field of the built environment is traffic. Traffic signal colors are plotted in Fig. 2. Recently, colors indicating various technological processes have been standardized, as seen in Fig. 3. A special field of technology is the color signals of pipelines. Color signals for the fluid carried have been standardized, including different uses of the same fluid. For instance, there are different signal colors for drinking water, for utility hot water, condensing water, hot water for heating, soft water, neutralizing sewage, and other applications. In Fig. 4, pipeline colors are plotted in the given Coloroid sections [3, 4].

Color information of aesthetic content is mostly emotional. It expresses inward conditions and the wish to exert mental and emotional effects based on a common semantic knowledge. Because of their operative and recording functions, visual codes are not only carriers of the message, and social ideas of the color-designed space as a work of art, but may also display the attitude typical of the artistic subject and culture. It follows from all these that color dynamics creates complex color conditions and puts them into

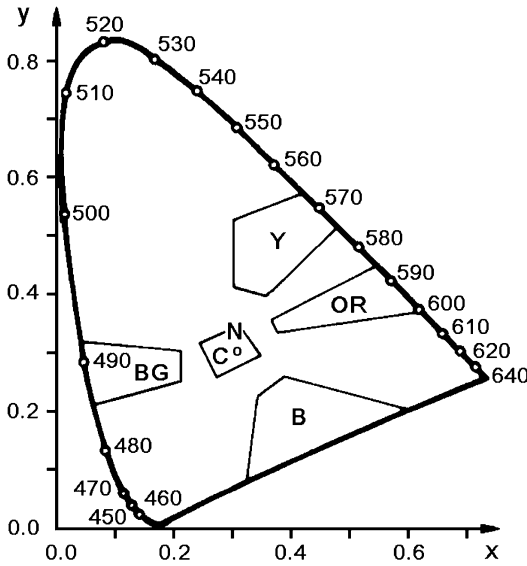


Functionality of Color, Fig. 1 Safety signal colors in the CIE diagram: (a) information, (b) warning, (c) prohibition, (d) fire, (e) instruction, N neutral complementary domain, Y yellow, O orange, R red, B blue, G green

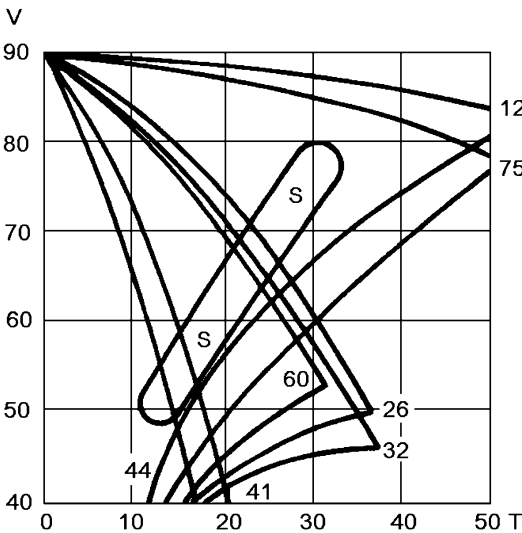


Functionality of Color, Fig. 2 Traffic light colors in the CIE diagram: Y yellow, R red, B blue, G green, N neutral complementary domain

a colored world; thereby it has a manifold message, that is, it both expresses and molds human consciousness and emotions. In other words, **environment color design** is an artistic activity



Functionality of Color, Fig. 3 Technology signal colors in the CIE diagram: *Y* yellows, *OR* orange reds, *B* purples and violets, *BG* bluish greens, *N* neutral complementary



Functionality of Color, Fig. 4 Signal colors (*S*) to distinguish pipelines in Coloroid sections. The horizontal axis shows the Coloroid saturation scale, the vertical axis, and the Coloroid lightness scale. Numbers represent Coloroid hues

on a large scale, a piece of art existing in the built space, and as such, it has not only to cope with its specific iconic task but also to meet the special conditions of color dynamics.

Functionality of Color in Architecture

Color, as a component of a building's image may result directly from the constructive process as is the case of vernacular buildings from the past and present applied as in the local tradition, or otherwise it may be decided at the end of the construction when a protective coating is required. Nowadays architects prepare complete designs, with finishes included, though this is not always the case with color specifications, often left for the final phase. In one way or another, the many functions of color may be used in favor of architectural façades and exposed structures.

Color can turn buildings into landmarks, spoil the panorama or contribute to the identity of a place. Architectural color design may be considered a matter for experts, but because building exteriors are part of the public realm, color is inevitably subject to acceptance or rejection from users: owners, pedestrians and drivers. Color is a primary perception and, as such, it affects human beings at physical and psychological levels. Architectural color works simultaneously from the behavioral the practical and for the specific interest of architecture.

Color Functionality from a Practical Point of View

The physical properties of color may be widely utilized in architecture. Deciding building colors goes hand in hand with technical aspects of the material, as each one has particular ► [appearance](#) characteristics and comes with a ► [palette](#) of color specific to the product. The outermost layer of the building should be designed to resist weather and environmental conditions, such as low and high temperatures, rain, wind, dust, and pollution.

Durability and maintenance are considerations to have in mind when thinking of exterior colors finishes, as some conceal dust better than others. As regards paints finishes in certain cases the best solution is to apply a coat of economic paint in order to have a clean façade considering that dark and saturated tones discolor easily. Natural cladding materials in appropriate colors may also be practical for reducing maintenance of high-rise buildings.

Color as part of visual ergonomics is given great importance in interior design, considering the time spent indoors. For the areas implicated, the colors of building exteriors are also relevant from the point of view of *visual comfort*. Color is used to compensate the ► **glare** effect, produced when high levels of luminosity from white or light-colored surfaces reflect back in the eye, causing discomfort and stress.

The climatic and natural lighting conditions of each location should be analyzed before taking color decisions, especially when critical situations occur, as in high latitudes during winter, when bright colors are preferred for *visual aid*, as opposed to dark nuances.

Design for energy conservation is a present trend in architecture. In order to ensure *thermal comfort*, sophisticated finishes are as important as design principles of insulation, orientation, ventilation, and color. In hot climates light colors and reflective materials serve to avoid overheating of roofs, while black surfaces, preferably opaque, are used in building exteriors for absorbing the maximum heat.

Color Functionality from a Behavioral Point of View

Color applied to building exteriors has the potential to benefit the user, improving and facilitating interaction with the surroundings. Color serves as an *informative* tool. It acts as a reference for content, identity, and location, but in order to accomplish its functions, it should be applied carefully, for color randomness and excess work in the opposite direction.

Color codification for *safety*, as used in industrial building, may be applied to architectural design in a conventional manner, for marking emergency exits and secure zones, or applied in a versatile way to building parts and elements for *orientation*. On border scale, Familiarization with large structures as fairs, residential blocks and commercial centers, may be successfully achieved by color design.

The semiotic function of color applied to architecture contributes to *clarity*. This color function is particularly helpful to drivers, as distinctive tones and combinations can be perceived from a

distance. When approaching a market or a day-care center, an ordinary person could anticipate its role by the external colors, when these are in accordance with the use of the building.

The functions of color extend to the symbolic aspect as ► **hues** convey *meaning*. For example, a red door on a black wall would something about the owner's personality, maybe an artist. Semiotics in a more strict sense applies to nationality and affiliation, as in Olympic colors. Similarly, corporate *identity* requires color in precise nuances and combinations. The colors of an oil company or a bank may be transferred to the façades of a building exterior. However, the preservation of the architectural character is in the hands of the designer.

Building exteriors acts as a backdrop to human activities, especially in cities. By maintaining a *balance* between excitement and relaxation in urban color schemes, monotony may be avoided, and outdoor spaces, enhanced. The *aesthetic* aspect of architecture is relevant to develop positive attitudes and a sense of belonging in the user.

Color Functionality from an Architectural Perspective

Architectural color is based on the formal characteristics of the building, such as composition style, material and proportions. Historical data and usage are also part of the information to be considered in the design. The setting, natural or man-made, is of great influence in the perception of the external appearance of buildings. Ideally, color design should meet the requirements of each project in the formal, semantic and contextual aspects.

Color serves for the *optimizes building design* by reinforcing style and concept. For functional and aesthetic purposes new constructions deserve a mindful color proposal.

Revaluation of existing buildings may be achieved through color design. An ordinary construction may be transformed by color, emphasizing its character, neutralizing its obtrusive effect or making unsightly parts less conspicuous. Buildings could recover their authentic character by *color restoration*. Proposals may be based on traditional ► **color palettes**, specific for the

location and architectural period. As an alternative to date nuances that approximate to the original ones could be applied according to the architectural style.

Facade *decoration* may be used to complement building fronts. This mode of coloring plays a key role in balancing architectural compositions to make them more pleasant on behalf of the human scale.

Art and design in the form of pictures, supergraphics, and *trompe l'oeil* paintings may be used in the renovation of unsightly elevations. These serve to visually integrate or decompose façades, giving value to blind walls and revitalizing of outdoor spaces.

On a greater scale color has the potential to *regulate the impact* of obtrusive structures within a context, in urban and natural settings.

A Major Consideration in Architectural Color: Scale of Perception

Building finishes and colors have various functions according to the impact produced at different scales or distances of perception. This refers to the position from user to object, in this case a building. The relevant scales for architectural color perception are the architectural scale, which considers the building as a three-dimensional object; the detail scale, which focuses on its parts and language elements; and the context scale, which considers the setting, natural or man-made.

On the architectural scale color affects the perception of a building as it would in the case of a sculpture, having a direct influence on the *definition* of volume, plane, and proportions and, consequently, in the *articulation of three-dimensional form*. Ideally, at this scale, the object and its coloration produce an integral result.

On the *detail scale* color is important for achieving *order and balance* within a façade or architectural composition. The different parts of the building present an opportunity for sophisticated color design. Depending on the characteristics of the architecture, it may go from structure and volume to doors and moldings, and the respective materials, textures and shadows.

On the *context scale* the color of a building determines the degree of visual *attachment or*

detachment to the surroundings. Building-context integration depends on multiple conditions apart from color: orientation, position, size composition, and material quality, amongst others. With ingenuity, color may work to make a building inconspicuous or even camouflage. On the other hand, it is easier to make a building stand out by the use of bright or unusual colors. By following a color master plan, a building acts as part of a broader context, from street and block to town and region.

Building color is a changing component; its equilibrium is fragile, and a subtle variation may easily modify the whole. Color is a design tool, but it may act as double-edged sword, as its potential may be reversed to cause discomfort, disorientation and ► [COLOR POLLUTION](#). In the best interest of the user, architectural color should be handled in a knowledgeable way.

F

Color Tectonics in Architecture

Tectonics in architecture and urban design is the discipline dealing with the principles of design, structure, construction, and ornamentation. Color tectonics refers to the function that color plays in accentuating these principles.

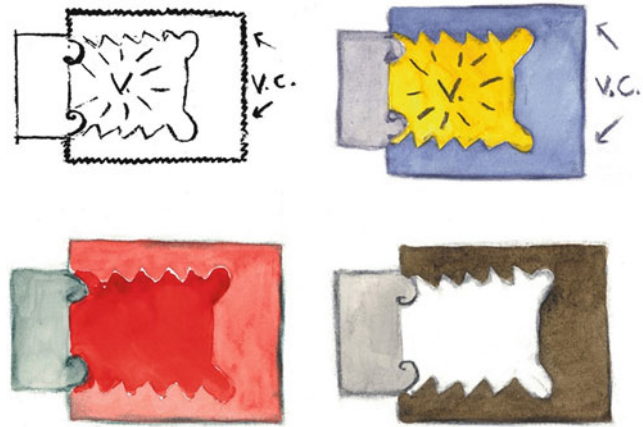
Color and form are codependent in built form. In architecture color plays many roles throughout the design process. Color is used dynamically in diagrams in the conceptual phase. In the design development phase, color can enhance the perception of physical form and define interior space. In detailing color can express the parts of the building that contribute to the whole. These color and form strategies will be present as material color in the finished building. In the architecture, color can be functional as an enhancement to form or abstract as expressive media. Color can be associative by carrying cultural values and meaning, as well as eliciting emotional response. In all cases color combined with form becomes another layer of interpretation and clarification of the intentions of the designer [7].

Conceptual Phase

A building, like a city, has an order based upon a clear hierarchical relationship of parts to the

Functionality of Color,

Fig. 5 Color adding nuance to a concept diagram by Louis Kahn



whole. Architecture begins with a concept expressed both verbally and in diagram. In the conceptual phase of the design process, line drawings are used to represent an abstract relationship of the essential parts of the building. These parts can be described metaphorically or formally, and the relationship of these parts to one another creates the generative idea that is the point of departure for the design. These relationships are often expressed as a dialogue between oppositions such as public/private or active/passive. They can also represent events in the experience of architecture such as hierarchy, separation, connection, transition, and assimilation. The drawings are usually monochromatic, but if one assigns colors to these diagrams representing the character of the part (i.e., red represents active functions and blue passive, or saturated hues might be dominant and muted hues subordinate), then color juxtaposition and ► [contrasts](#) can set up a more visually dynamic relationship of these parts in the diagram. These color choices in the conceptual phase are chosen for their dynamic relationships in juxtaposition rather than by functional criteria, but these colors will often influence choices in the design development phase such as the delineation of three-dimensional form, building materials, and lighting.

A repertory theater is a place where actors must project their voice to the audience without acoustical support. A design concept for a repertory theater by Louis Kahn was first expressed metaphorically as “a violin within a violin case.” The

diagram was a box representing the outer shell of the building or the “violin case” with an irregular form inside representing the “violin,” an acoustically designed theater. In black and white, this is just a simple diagram. With color, however, the violin and violin case form an expressive dialogue between these parts, making them more dynamic and meaningful (Fig. 5).

Design Development

In the design development phase of a design process, color is used to accentuate or diminish three-dimensional form. The exploration of perceptual effects created by color, particularly the spatial effect of color, color and atmosphere, and the familiarity with the principles of camouflage [8] are helpful tools in this pursuit (Fig. 6). The color decisions here are based on how the eye perceives color and form. The accentuating and disguising of form with color are the techniques supporting color tectonics. Using color in models in this phase establishes color juxtapositions and ► [contrasts](#) that will be present in the final building. Model studies similar to the experiments by Lois Swinoff [9] are valuable exercises.

The purpose of color tectonics is to use color to clarify and enhance the perception of building form. This form is made up of parts arranged hierarchically that unite to shape the whole. This color/form carries the aesthetic and expressive intentions of the architect, and clarity of expression is usually the goal. In some instances, however, there is an intention to disguise and actually

Functionality of Color,

Fig. 6 Transforming two solid cubes and a void with a color illusion

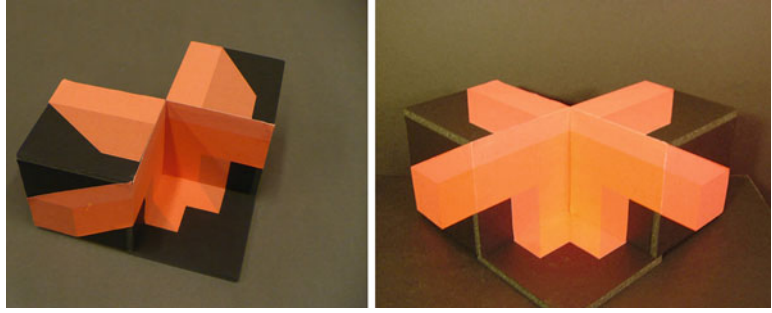
**Functionality of Color,**

Fig. 7 Color used for “deconstruction.” Portland Building by Michael Graves, Portland, Oregon



deconstruct the true form, and color becomes a powerful tool in this endeavor. This was the case in the Portland Building during the period known as deconstruction in architecture (Fig. 7).

Detail

The details of buildings carry expressive content. Structural details communicate what elements in the building are supporting and load bearing and what are non-load bearing and serve to enclose the building and define space within

it. Many functional parts of a building are repetitive, and through this repetition patterns and rhythms are established which support the unity of the building as a whole. ► **Color contrasts** provide a powerful tool for making these parts visible in a meaningful way. In the main entry facade to the Fort Wayne Repertory Theater, light gray concrete is contrasted with darker brick to both characterize the structural forces within a large wall and to accentuate the main entry (Fig. 8).



Functionality of Color, Fig. 8 Color contrast emphasizing structural detail. The Arts United Center by Louis Kahn, Fort Wayne, Indiana

Color Imagery

In the final or building phase of the design process, color choices are more specific. To this point color has been an integral tool in clarifying the conceptual and design intentions of the architect and not as a secondary consideration in the design process. The final color/form imagery will both convey these intentions and also carry meaning through cultural associations, symbolism, and emotional response. The use of colors from the natural and built environmental contexts can give the architecture a sense of place as well [10]. In this process color is a principal consideration through all phases. The imagery experienced in the architecture will carry these design ideas in form and structure through perceptual color effects, but there will be an additional layer of meaning with the associative uses of color and the aesthetic imagery from experiencing the architecture as a whole in its environmental context.

Cross-References

- ▶ [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- ▶ [Color Appearance](#)
- ▶ [Color Dynamics](#)
- ▶ [Color Harmony](#)
- ▶ [Color Palette](#)
- ▶ [Color Pollution](#)
- ▶ [Environmental Color Design](#)
- ▶ [Unique Hues](#)

References

1. Pickford, R.W.: Psychology and visual aesthetics. Hutchinson, London (1972)
2. Birren, F.: Science and art, objective and subjective. *Color Res. Appl.* **10**, 180–186 (1985)
3. Nemcsics, A.: Colour dynamics. Environmental colour design. Ellis Horwood, New York (1993)

4. Nemcsics, A.: The coloroid colour system. *Color Res. Appl.* **5**(2), 113–120 (1980)
5. Frieling, H.: *Farbe in Kultur und Leben*. Battenberg Verlag, Munich (1963)
6. Nemcsics, A.: Das Farbenpräferenz-Indexzahlensystems im Dienste der farblichen Raumgestaltung. *E'KME Tud. Kozl.* **13**(1/2), 21–261 (1967)
7. Minah, G.: Color as idea. *Colour Des. Creat.* **2**(3), 1–9 (2008)
8. Rao, J.: *Introduction to camouflage and deception*. Metcalf House, Delhi (1999)
9. Swinoff, L.: *Dimensional color*. Birkhauser, Boston (1989)
10. Lenclos, J., Lenclos, D.: *Colors of the world, the geography of color*. W.W. Norton, New York (2004)

Fundamental Colors

► [Primary Colors](#)

G

Galaxy Color Magnitude Diagram

Michael H. Brill
Datacolor, Lawrenceville, NJ, USA

Definition

A scatter graph of galaxies showing the relationship between each galaxy's absolute magnitude and its estimated temperature, or between optical and perceptual proxies for these quantities. The construction of such a diagram is similar to that of the Hertzsprung-Russell diagram, but the interpretation in terms of physical properties is not as precise.

Absolute Magnitude and Temperature Scales

To compute a galaxy's absolute magnitude, it is treated as a point-like object, whereupon its radiation is corrected (by the inverse-square law) to a distance of 10 pc, and the absolute magnitude is computed as the number of $10^{-0.4}$ attenuations of similarly compensated power from a reference to achieve a match.

The temperature is also the same "effective temperature" as is used in the H-R diagram. However, galaxies are visible from much farther away than individual stars, so a galaxy's recessional redshift exerts an appreciable influence on its

effective temperature. Because light from farther galaxies (and greater redshift) requires more time for light to travel to the Earth, the galaxy color-magnitude diagram, if it were not redshift corrected, would depict an earlier universe at the red end of the temperature scale. However, in practice the temperature is redshift corrected to be in the "rest frame" of the galaxy. In this case, the galaxies divide into two clusters: bright and reddish as opposed to dim and bluish. The existence of such clusters places strong constraints on theories of galactic formation.

History

Galaxy color-magnitude diagrams came into wide use starting in 1990, when the Hubble Space Telescope was launched into orbit. The Space Telescope was able to render high-contrast images outside the Earth's atmosphere without encountering the absorption, scattering, and haze that beset Earth-bound telescopes. In digital images conveyed to earth from the Space Telescope, it was possible to see the colors, not only of stars, but of whole galaxies that are so distant as to be too dim to measure on the Earth.

Coordinate Options

Color-magnitude diagrams are not always defined in terms of fundamental physical properties such

as those described above. In one embodiment, the galaxy color-magnitude diagram coordinates are defined by the log outputs of two sensors, typically called *g* and *r*, the sensors having different spectral sensitivities. The abscissa records *r* (which stands in lieu of the absolute magnitude), and the ordinate records *g*–*r* (which relates to the temperature). At the top of the diagram, one finds the red sequence of galaxies (elliptical galaxies), and at the bottom one finds the blue cloud (spiral galaxies).

Cross-References

- [Color-Magnitude Diagrams](#)
- [Hertzsprung-Russell Diagram](#)

Gamut Volume

Philipp Urban

Fraunhofer Institute for Computer Graphics
Research IGD, Technische Universität Darmstadt,
Darmstadt, Germany

Synonyms

[CIECAM02](#) (Standards: CIE); [CIELAB](#) (Standards: CIE)

Definition

The gamut volume describes how much of a color space a gamut occupies. Hence, it depends on the color space in which the gamut is represented. In order to be used as a meaningful quantity to compare gamuts, a representation within a highly perceptually uniform color space is recommended. Color spaces typically used to calculate gamut volumes are CIELAB or CIECAM02 [1]. Please note that these spaces are only nearly perceptually uniform. Hence, spaces with improved uniformity [2, 3] are sometimes used as well mathematically;

the gamut volume V is defined by the volume integral

$$V = \iiint_G dc_1 \, dc_2 \, dc_3,$$

where G is the gamut, and c_i are the color space coordinates, e.g., $(c_1, c_2, c_3) = (L^*, a^*, b^*)$. The gamut volume is expressed in *cubic color space units* and mostly rounded to thousands for device and media gamuts [4].

In a perceptually uniform color space, the gamut volume is nearly monotonically related to the number of distinguishable colors within the gamut, i.e., the larger the gamut volume, the larger is the number of distinguishable colors in the gamut. Mathematically, the number of distinguishable colors n within the gamut may be expressed as the smallest number of just noticeable distance (JND) spheres covering the gamut, i.e.,

$$n = \min \left\{ |\mathbb{S}| \mid G \subset \bigcup_{S \in \mathbb{S}} S \right\},$$

where G is the gamut, \mathbb{S} is a set of JND spheres, and $|\mathbb{S}|$ are the number of elements in \mathbb{S} . It is worthy to mention that the JND depends on the viewing condition, i.e., the luminance level, viewing field, adaptation state, etc. Hence, the gamut volume and the number of distinguishable colors within the gamut are not two sides of the same coin. However, assuming similar viewing conditions, gamut volumes can be reasonably compared to determine the gamut that includes more distinguishable colors.

The gamut volume is only one quantity to compare gamuts and does not give any information about colors covered by the gamut. A gamut that includes highly relevant colors for a distinct application should be preferred over a gamut that does not cover these colors regardless of the gamut volume.

Calculation

A major problem of accurately calculating the gamut volume is the non-convex shape of the

gamut within a perceptually uniform color space. For output devices, such as printers, displays, or projectors, a good approximation of the gamut is given implicitly by an accurate device model

$$\mathcal{M} : P \rightarrow C,$$

i.e., a predicting function that maps the control value space P (e.g., $P = \text{RGB}$, CMY) into the color space C used for gamut representation (e.g., $C = \text{CIELAB}$, CIECAM02 , etc.). The gamut G can be parametrized by the device's control values, i.e., $G = \mathcal{M}(P)$. Using this parametrization and assuming that \mathcal{M} is an injective differentiable function with continuous partial derivatives, the gamut volume can be computed using the substitution rule for integration

$$\begin{aligned} V &= \iiint_{G=\mathcal{M}(P)} dc_1 \, dc_2 \, dc_3 \\ &= \iiint_P |\det(J\mathcal{M}(p_1, p_2, p_3))| dp_1 \, dp_2 \, dp_3, \end{aligned}$$

where $J\mathcal{M}$ is the Jacobian matrix of \mathcal{M} . Since P has typically cubic shape, the latter integral can easily be calculated numerically. Furthermore, no representation of the rather complex gamut boundaries within C is required.

While mathematically elegant this method is often inapplicable, e.g., for image gamuts or if the control value space of a device has more than three dimensions, such as for multi-ink printers (CMYK , CMYKOG , etc.). In this case, a gamut boundary descriptor (GBD) [5] describing the gamut boundary by a tessellation within C can be utilized. This tessellation is then used as the boundary of a tetrahedralization that fills the entire interior of the gamut. In most well-behaved cases, tetrahedralization can be simply generated by forming tetrahedra with each of the surface triangles and one inner point, e.g., the centroid of the GBD vertices [4]. To approximate the gamut volume, the volumes of all tetrahedrons within the tetrahedralization are added up. The volume of a tetrahedron with vertices a , b , c and d is given by $1/6|\det(a-b, b-c, c-d)|$. For devices using

additive color mixture, such as displays, the gamut is convex in the CIEXYZ color space. A tetrahedralization of the gamut in CIEXYZ space allows a convenient partitioning into subgamuts for further consideration. Since a transformation,

$$\mathcal{F} : \text{CIEXYZ} \rightarrow C,$$

from CIEXYZ into a perceptually uniform color space C can be assumed to be a diffeomorphism, the following equation applies:

$$G = \bigcup_{T \in \mathbb{T}} \mathcal{F}(T),$$

where G is the gamut in C , and \mathbb{T} is the set of all tetrahedrons within the tetrahedralization. Hence, the gamut volume can be computed as follows [6]:

$$\begin{aligned} V &= \iiint_G dc_1 \, dc_2 \, dc_3 \\ &= \sum_{T \in \mathbb{T}} \iiint_{\mathcal{F}(T)} dc_1 \, dc_2 \, dc_3 \\ &= \sum_{T \in \mathbb{T}} \iiint_T |\det(J\mathcal{F}(X, Y, Z))| dX \, dY \, dZ, \end{aligned}$$

where $J\mathcal{F}$ denotes the Jacobian matrix of \mathcal{F} . It should be noted that the meaningfulness of the computed gamut volume – with respect to its relation to the number of distinguishable colors – is usually more effected by the uniformity of the color space C rather than by the computational method.

Applications

Most often gamut volumes are used to compare imaging systems with each other or with standard gamuts (e.g., sRGB or Adobe RGB). Also the volume of gamut intersections is used and related to standard gamuts, e.g., to state that a display's gamut covers more than 95 % of the Adobe RGB gamut.

The gamut volume is sometimes used as an objective function to optimize imaging systems (e.g., displays or printers). Particularly, the spectral or colorimetric characteristics of printing inks or display primaries are optimized to maximize the gamut volume. Most of such optimizations are subjected to other beneficial properties of imaging system or feasibility constraints.

References

1. CIE Publication No. 159.: A Colour Appearance Model for Colour Management Systems: CIECAM02. CIE Central Bureau, Vienna (2004)
2. Thomsen, K.: A euclidean color space in high agreement with the cie94 color difference formula. *Color. Res. Appl.* **25**, 64–65 (2000)
3. Urban, P., Schleicher, D., Rosen, M.R., Berns, R.S.: Embedding non-euclidean color spaces into euclidean color spaces with minimal isometric disagreement. *J. Opt. Soc. Am. A* **24**(6), 1516–1528 (2007)
4. Morovic, J.: *Color Gamut Mapping*. Wiley, Hoboken (2008)
5. Morovic, J., Luo, M.R.: Calculating medium and image gamut boundaries for gamut mapping. *Color. Res. Appl.* **25**, 394–401 (2000)
6. Rodriguez-Pardo, C. E., Sharma, G., Speigle, J., Feng, X.-F., Sezan, I.: Efficient computation of display gamut volumes in perceptual spaces. In: *IS&T/SID, 19th Color and Imaging Conference*, pp. 132–138. San Jose (2011)

Ganglion cells

Arne Valberg

Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway

Trichromatic theories of color mixture have had a tendency to overemphasize the role that L, M, and S-cone receptors have for color *perception*. The realization, however, that these physiological units can only explain the laws of additive color mixture and color matching (metamerism) has led to an end of this bias. Modern cone-opponent color vision models combine inputs of one excitatory cone type in the center and another inhibitory type (or types) in the surround of the cell's

receptive field (e.g., L-center–M-surround, giving an “L–M” ON-center cell, or an M-center–L-surround, “M–L” ON-center cell (also called increment cell). The opposite configuration is found in “–L + M” and “–M + L” OFF-center cells (decrement cells). Ganglion cells with S-cone inputs can be characterized as “M–(S + M)” and “S–(L + M)” cells [1–5]. This organization in spatially antagonistic receptive fields and parallel pathways has moved the scientific interest from the ancient idea of cones as color units, toward neural processing in retinal ganglion cells.

Retinal ganglion cells are situated near the inner surface of the retina. They receive cone signals modified by horizontal, bipolar, and amacrine cells and have long axons extending into the lateral geniculate nucleus (LGN) and from there to the various visual centers of the brain. Ganglion cells are spiking neurones of different sizes, connections, and spectral selectivity. Nerve impulses (spikes) are “all or none” electrical signals of about 1 ms duration each. The number of spikes per second is used as code for transmitting information from one cell to another. Excitation of a cell results in an increased firing frequency, whereas inhibition results in a reduced rate of firing. There are about 1.5 million different retinal ganglion cells in the human retina, each type forming a mosaic of receptive fields with a close to ideal covering factor (neighboring dendritic fields do not overlap [6]). In the central fovea a single ganglion cell is connected to a single cone in the center of its receptive field. Midget retinal ganglion cells are connected in an opponent way to L- and M-cones and have rather small dendritic fields. Cells making contact with S-cones (the so-called bistratified cells [7]) do this in two different synaptic layers and seem to have a coinciding (coextensive) center/surround structure of their receptive fields.

In addition to color coding in cone-opponent parvocellular (PC)- and koniocellular (KC)-ganglion cells (about 80 % of all retinal ganglion cells) with a sustained firing pattern, one also finds transiently responding, nonspectrally selective, magnocellular (MC) cell types (about 10 % of all retinal ganglion cells). These very contrast-

sensitive cells are called parasol- or diffuse ganglion cells and have spectral sensitivities matching the primate psychophysical spectral sensitivity function $V(\lambda)$ to luminance contrast. They combine L- and M-cones in an additive way (L + M) in both center and surround of their receptive field. MC-cells also form parallel pathways as they are either of the ON-center, OFF-surround, or OFF-center, ON-surround types. Since these cells have close to $V(\lambda)$ spectral sensitivity and large receptive fields, they are excellent candidates as neural substrates for Minimally Distinct Borders (MDB) and flicker-photometric luminance [7].

A small number of retinal ganglion cells are photosensitive with their own photopigment, melanopsin, and contribute to the circadian rhythm and papillary light reflex but little or nothing to vision [8].

In the ganglion cells slow, standing potentials of the cones and bipolar cells are transformed into a series of nerve impulses. The number of spikes of a particular cell contains information about the intensity or strength of a visual stimulus. Information about a stimulus' compound features such as contrast, form, color, movement, direction of motion, distance, etc. is conveyed by activation of more or less specialized cell types. However, a particular attribute, say orange color of medium lightness, may result in activation in more than one cell type. The orange "percept," for instance, is best regarded as a result of the combined response of several ganglion cell types (not necessarily monitored by a separate cell). The perception and scaling of lightness, for example, may be a result of interaction between nerve processes that are elicited by light increments and light decrements in parallel Increment- and Decrement channels [9]. Equal responses of a low-level cell such as a certain polarization of a cone receptor's potential or the number of nerve impulses/s of a retinal ganglion cell may result from exposure to several different stimuli, such as changes in color, intensity, size, orientation, or combinations of these. The combination of visual attributes that makes up a complex visual object can be regarded as a result of integration, correlation, and comparisons of these lower units' activities.

Retinal ganglion cells project to similar cell types in the parvo- (PC) and koniocellular (KC) layers of the lateral geniculate nucleus (LGN), and from there information is directed to specific visual centers of the brain, depending on the "trigger features" they encode.

References

1. Wiesel, T.N., Hubel, D.H.: Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J. Neurophysiol.* **29**, 1115–1156 (1966)
2. DeValois, R.L.: Analysis and coding of color in the primate visual system. *Cold Spring Harb. Symp. Quant. Biol.* **30**, 567–579 (1965)
3. Lee, B.B., Martin, P.R., Valberg, A.: A physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *J. Physiol.* **404**, 323–347 (1988)
4. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 242–265 (1984)
5. Valberg, A., Seim, T., Lee, B., Tryti, J.: Reconstruction of equidistant color space from responses of visual neurones of macaques. *J. Opt. Soc. Am.* **A3**, 1726–1734 (1986)
6. Dacey, D.M.: The mosaic of midget ganglion cells in the human fovea. *J. Neurosci.* **13**, 5334–5355 (1993)
7. Dacey, D.M., Lee, B.B.: The "blue-on" opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature* **367**, 732–735 (1994)
8. Hattar, S., Liao, H.W., Takao, M., Berson, D.M., Yau, K.W.: Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002)
9. Valberg, A., Seim, T.: Neural mechanisms of chromatic and achromatic vision. *Color Res. Appl.* **33**, 433–443 (2008)

Ganzfeld

Sunčica Zdravković

Department of Psychology, University of Novi Sad, Novi Sad, Serbia

Laboratory for Experimental Psychology, University of Belgrade, Belgrade, Serbia

Definition

A ganzfeld ("whole field" in German) is an absolutely homogeneous region of space that covers an

observer's entire visual field. It can be of any single uniform wavelength and intensity.

History

The ganzfeld was first introduced in 1930 [1], by German Gestalt psychologist Wolfgang Metzger (1899–1979). Metzger was interested in elementary visual phenomena produced by impoverished visual conditions. The ganzfeld contains no luminance border, luminance ramp, or texture. The light reaching the eye is absolutely equal from all possible directions. Such conditions, called whiteout, can occur naturally during a snowstorm or in an airplane flying through clouds.

Perceptual Experience in the Ganzfeld

The typical visual phenomenology, for a person of normal sight, includes the emergence of "Eigengrau" [2], a uniform gray fog of indeterminate depth. This percept occurs regardless of the physical intensity and wavelength of the ganzfeld. Any color in a ganzfeld will eventually fade to filmy gray. The experience of a bright ganzfeld differs from that of a completely dark ganzfeld. Total darkness implies a lack of visual objects and is somewhat less disorienting. However, disorientation in a bright ganzfeld is fast and compelling.

From the beginning, the observer is unable to fixate any part of the field and becomes conscious of the vitreous opacities and blood vessels within the eye. Perceived luminance gradually diminishes, the original color slowly bleaches, and the whole visual field turns into a gray fog. This fog sometimes has a barely noticeable washed-out tone complementary to its actual color. Prolonged exposure to a ganzfeld results in more complex percepts. Sometimes even patterns (dots, lines) are reported. Because hallucinations have been reported in extreme cases, the method has been used outside of vision science. Sometimes prolonged exposure (10–20 min) leads to intermittent losses of vision (so called "blank-out"). Participants sometimes cannot report if their eyes are open or closed.

In traditional studies, phenomenological reports were obtained from observers trained in providing such introspective reports "post hoc" from memory. The limits of this introspection, especially in the case of complex visual experience, have been questioned. This constitutes a problem for a phenomenon that might involve disorientation and consciousness-altering states. Alternative methods include "on-demand" reports either elicited by the experimenter or volunteered by the observer. This allows a random sample of momentarily experienced states with no involvement of memory. Unfortunately some of the same criticisms are applicable here, too. This method also requires training in introspection.

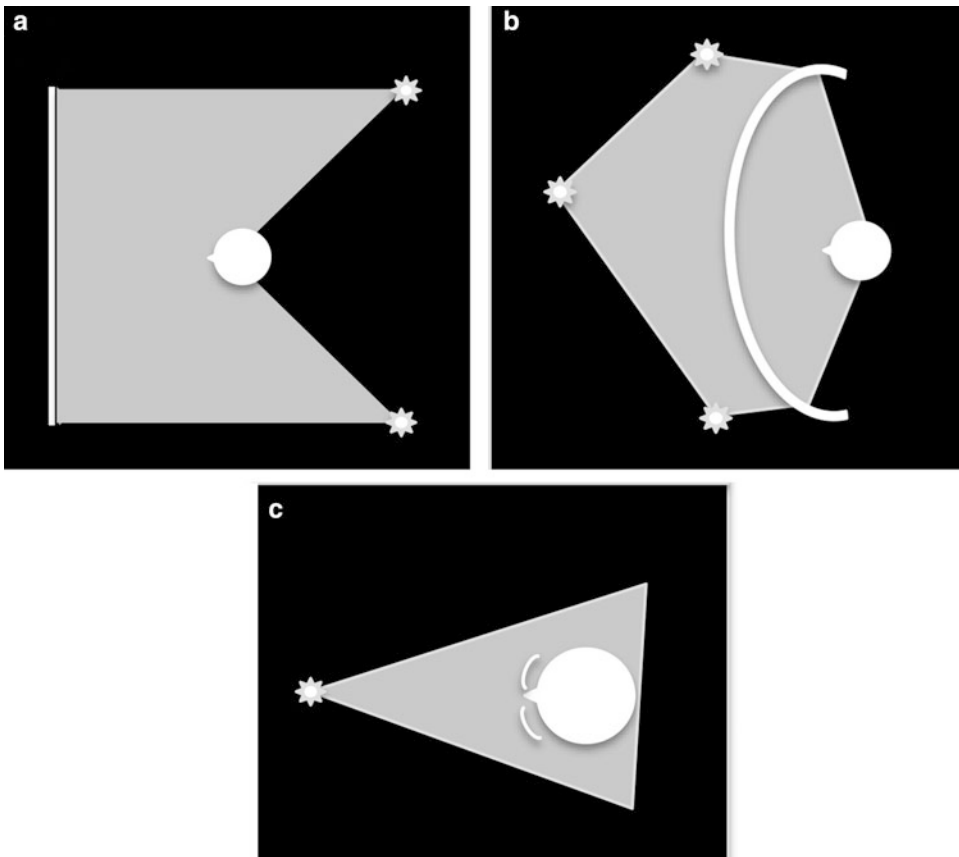
Producing an Experimental Ganzfeld

Although a primitive ganzfeld can be achieved by placing a person in complete darkness, creating a proper ganzfeld of a desired brightness is not a trivial challenge.

In the 1930s, observers were placed in front of either a smooth homogeneous wall (Fig. 1a) or a large sheet of paper [1]. A simple way to create a ganzfeld is to cover each eye with a section of a table tennis ball (Fig. 1c). These need to be trimmed to fit the shape of the face surrounding the eye, leaving no extra gaps for illumination to leak in, and they can be attached to the face using cosmetic cement. Industrial production does not guarantee a completely uniform thickness of the plastic, but this is not a problem because the surface is too close to the eye for focus. Another option is placing the head inside a large hemisphere, which is uniformly illuminated from outside (Fig. 1b). A somewhat better solution is the usage of translucent contact lenses. This method provides a high degree of homogeneity.

More modern studies tend to use computer-controlled xenon lamp-based D-ILA projector creating (usually red) ganzfeld.

Finally any change in luminance or color, including blinking, restores perception. Given that the ganzfeld effect can be achieved fast, this



Ganzfeld, Fig. 1 Traditional spatial arrangements of Ganzfeld. (a) Observer in Metzger's experiments was placed in front of a uniform wall lit from behind. (b)

Domes lit from outside. (c) Lenses or table tennis balls placed in front of the eyes functioning as light diffusers

is not necessarily a problem. If prolonged viewing is required, the eyelids can be taped open to prevent blinking, but the eyeballs need to be kept moist. This is easily achieved if the eye is completely covered by the table tennis ball section, as described.

Theoretical Importance

The ganzfeld clearly demonstrates that our senses in general (vision in particular) function to detect change. Invariant stimulation does not produce a sensory response. Framed in Gestalt terminology, an unstructured physical field cannot produce perception due to the lack of perceptual organization.

Observable Characteristics

Brightness. The full ganzfeld phenomenon is achieved on average after 5–7 min [3]. During that time, with no change in stimulation, perceived brightness slowly decreases until it reaches the final value.

If the luminance in a ganzfeld is subjected to subliminal changes over time and space, observers can still detect the correct direction of change [4]. For a completely homogenous ganzfeld (9.1 footlambert), changing 1,000-fold per hour at a constant rate, observers detected the direction after 9.7 min.

Wavelength [5]. Although the ganzfeld produces an achromatic, dark gray percept, irrespective of original conditions, the initial

wavelength still dictates some characteristics. Longer wavelengths fade faster than short wavelengths. Shorter wavelengths create additional sensation of field darkening.

Monocular versus binocular ganzfeld. It has been shown that a ganzfeld presented binocularly (i.e., viewed with two eyes) does not show a loss of visual perception (so called “blank-out”) typical for monocular viewing [6]. However, recent findings show that a gradual loss of brightness and saturation (so called “fade-out”) is measurable both for monocular and binocular ganzfeld. Fade-out is however determined by the original field intensity and wavelength [7].

Sex differences [8]. Female observers are more responsive to longer wavelengths. They also preserve visual sensation shorter and report a fewer “blank-out” effects than male observers.

Explanation

The patterns that appear during the short ganzfeld exposures, such as zigzag lines or dot fields, are accounted for by early retinal processes. Receptive and other retinal cells exhibit spontaneous activity, oversaturation, and inhibitory reactions. Prolonged exposure elicits complex percepts, which presumably involve the central nervous system. The loss of vision is probably the consequence of brain discarding uninformative visual input about invariant stimulation.

There is a correlation between “blank-outs” and alpha activity in the resting EEG [9]. Alpha activity (8–12 Hz) is otherwise typical for relaxed states with no stimulation and closed eyes.

Related Phenomena

Aftereffects. If the ganzfeld is induced by yellow table tennis balls, after their removal the world will tend to look bluish, due to chromatic adaptation. Also during the initial bleaching of the original ganzfeld color, sometimes the percept does not remain neutral but adopts a pale complementary color.

Artificial scotomata. Eye movements can be counterfeited by fixating a large portion of visual field to stimulate an identical part of retina. This leads to “fade-out” and loss of percept within seconds. To maintain normal perception, a constant change of stimulation is necessary and this is achieved by eye movements. The absence of eye movements in large areas of the visual field, known as artificial scotomata, leads to perceptual filling in, creating large uniform surfaces.

Sensory deprivation. Although the ganzfeld represents extremely impoverished stimulation, the phenomenon represents perceptual but not sensory deprivation. In the former case, we have simply unstructured stimulation, while in the latter, physical intensity of stimuli is minimized.

Dark field vision. When observers are placed in total darkness in a relaxed state, with the eyes closed, they often report the appearance of phosphenes, believed to be the consequence of spontaneous firing of the cells. Phosphenes are similar to the percepts in the early stages of ganzfeld exposures.

Cross-References

- [Adaptation](#)
- [Afterimage](#)

References

1. Metzger, W.: Optische Untersuchungen am Ganzfeld. *Psychol. Forsch.* **13**, 6–29 (1930)
2. Knau, H., Spillmann, L.: Failure of brightness and color constancy under prolonged Ganzfeld stimulation. *SPIE Proc.* **2657**(04), 19–29 (1996)
3. Knau, H.: Thresholds for detecting slowly changing *Ganzfeld* luminances. *J. Opt. Soc. Am. A* **17**, 1382–1387 (2000)
4. Schubert, J., Gilchrist, A.L.: Relative luminance is not derived from absolute luminance. *Invest. Ophthalmol. Vis. Sci.* **33**(Suppl.), 1258 (1992)
5. Gur, M.: Color and brightness fade-out in the ganzfeld is wavelength dependent. *Vision Res.* **29**(10), 1335–1341 (1989)
6. Bolanowski, S.J., Doty, R.W.: Perceptual blankout of monocular homogeneous fields (Ganzfelder) is prevented with binocular viewing. *Vision Res.* **27**, 967–982 (1987)

7. Gur, M.: Perceptual fade-out occurs in the binocularly viewed Ganzfeld. *Perception* **20**(5), 645–654 (1991)
8. McGuinness, D., Lewis, I.: Sex differences in visual persistence: experiments on the Ganzfeld and afterimages. *Perception* **5**(3), 295–301 (1976)
9. Cohen, W., Cadwallader, T.C.: Cessation of visual experience under prolonged uniform visual stimulation. *Am. Psychol.* **13**, 410 (1958)

Geniculate Pathways

► [Magno-, Parvo-, Koniocellular Pathways](#)

Gestalt Grouping Principles (or Laws)

► [Perceptual Grouping and Color](#)

Ghost Image

► [Afterimage](#)

Glare

Wout van Bommel
Nuenen, The Netherlands

Definition

Condition of vision of persons within the illuminated area of a lighting installation in which there is discomfort and/or a reduction in the ability to see details or objects, caused by an unsuitable distribution of luminance and/or an unsuitable range of luminance values [1].

Note that conditions of discomfort or reduction of vision for persons not within the illuminated area are not covered by this definition. That condition is caused by obtrusive light into areas where the light is not meant to be. The effect is called light pollution and is dealt with in a separate entry.

Forms of Glare

Glare can take the form of disability glare that is the form which reduces visual performance. It can also take the form of discomfort glare which causes discomfort. The two forms are different phenomena. They may occur in combination but this needs not always to be the case. Disability glare is especially dependent on the amount of light falling on the eye and hardly on the luminance of the glare sources. Also the spectrum of the glare source has no influence. Discomfort glare is very much dependent on the luminance of the glare sources and also on their size. The spectrum of the glare source plays a role as well.

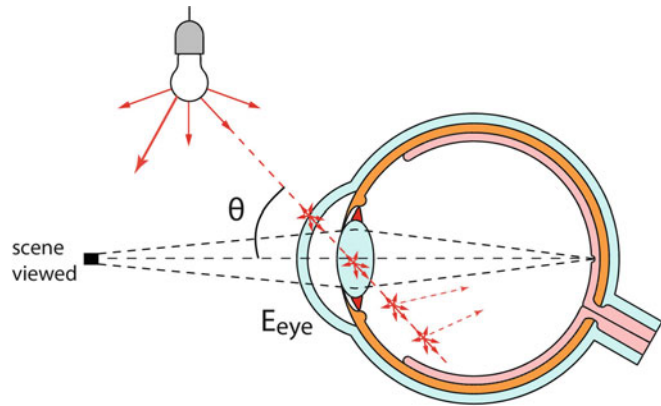
Glare has been studied mainly through research into how persons react to glare. Still much has to be learned about the mechanism behind the glare effects.

Disability Glare

Disability glare can best be understood by considering the light scatter taking place in the eye. A sharp image of the scene in the direct field of view is focused on the retina of the eye (Fig. 1). At the same time, light coming from a glare source that is outside the direct line of sight is partly scattered in the eye lens and eyeball due to their imperfect transparency. Part of the scattered light is redirected toward the retina overlaying the retinal image. It reduces the contrast of the retinal image of the scene. The scattered light produces a visual sensation that can be likened to the drawing of a bright veil across the field of vision. The effect can be compared to some extent to viewing through a net curtain or through a bride's veil. The veil can be considered as having a luminance – the veiling luminance – proportional to the degree of scatter in the direction of the fovea. The final disability effect is dependent on the combination of veiling luminance and scene luminance.

Holladay [2] found empirically the veiling luminance (L_v) to be dependent upon the illuminance on the eye (E_{eye}) in a plane perpendicular to the line of sight and the angle (θ) between the viewing direction and the direction of light incidence from the glare source (see Fig. 1 again):

Glare, Fig. 1 Light scatter in the eye due to a glare source



$$L_v = k \frac{E_{eye}}{\theta^2}$$

The factor k (with the dimensions degree²/steradian) is an age factor, its value being standardized at 10 for observers aged between 20 and 30 years. This value increases with age at a rate of roughly 0.2 per year.

This equation has been used by the lighting community for over 70 years, although there are limitations on its validity, among others to not too small angles [3].

For multiple glare sources, the total veiling luminance is obtained by adding the veiling luminances of the individual sources, thus

$$L_v = \sum_{i=1}^n L_{vi}$$

Discomfort Glare

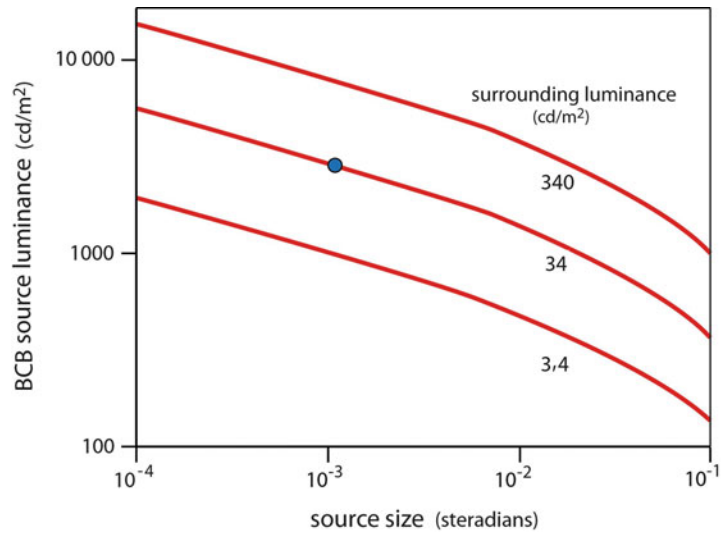
Discomfort glare has been and is being studied by the subjective response to it. The basis has been laid by Luckiesh and Guth [4] already in the 1940s. They used as the criterion for discomfort glare the sensation occurring at the border between comfort and discomfort, which they called the BCD sensation. In a situation with a single glare source, it depends on the luminance of the glare source in the direction of the observer, the luminance of the surrounding field, the solid angle of the glare source from the observer's position, and, finally, the displacement of the glare source from the observers line of sight (the off-line angle is usually used to characterize this).

Figure 2 illustrates the dependency of the BCD sensation on glare source luminance, solid angle of the glare source, and luminance of the surrounding field. Figure 3 shows the influence of the displacement of the glare source. The figures are redrawn from the original 1949 publication of Luckiesh and Guth. As can be seen from Fig. 3, a vertical displacement lowers discomfort glare more rapidly than a horizontal displacement. For example, the source luminance at a displacement angle of 40° in the vertical plane can be around 4,000 cd/m², while the same displacement in the horizontal plane only allows for a source luminance of some 1,000 cd/m², all for a same BCD sensation.

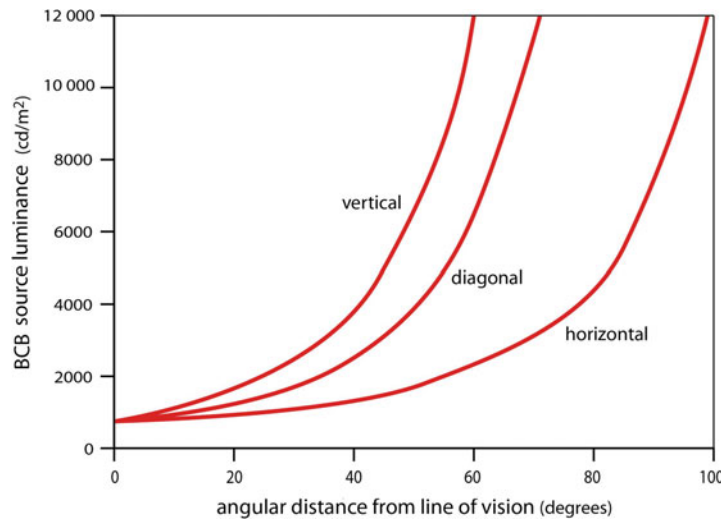
Discomfort glare research for interior lighting has hardly taken into account a possible effect of the spectrum of the glare source. Most of the studies have been carried out with incandescent lamps (the early studies) and with fluorescent lamps with color temperatures around 4,000 K.

Early research of discomfort glare for road lighting has shown that sodium gas discharge lamps (especially low-pressure sodium lamps) have a slight advantage over whiter light sources [5, 6]. With the introduction of LED light sources that can have many different white-light spectra, research on the effect of the spectrum has been intensified and indicates that high color temperature glare sources are more critical than lower color temperature white-light glare sources [7–9]. Results are often given on the basis of a

Glare, Fig. 2 The relationship between BCD glare source luminance and size of glare source (on the line of vision) for three different luminances of the (uniform) surrounding field [4]. The blue dot gives the starting situation used in Fig. 3



Glare, Fig. 3 The relationship between BCD glare source luminance and vertical, diagonal, and horizontal displacement of the glare source from the line of sight [4]. Starting point is the situation indicated with the blue dot in Fig. 2

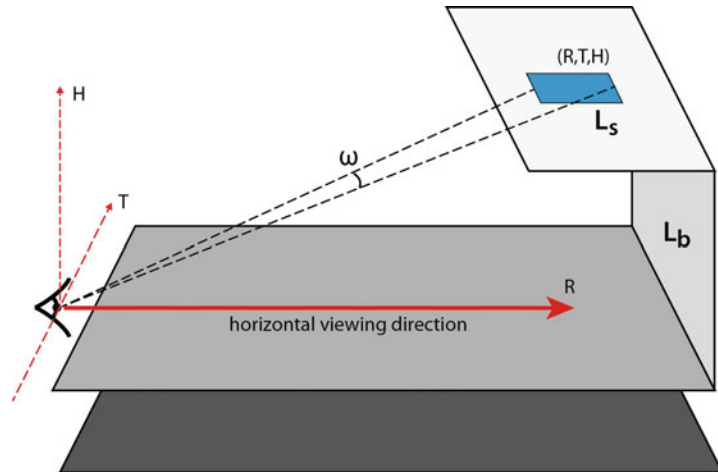


nine-point assessment scale with one standing for “unbearable glare” and nine for “unnoticeable glare.” This scale is often referred to as the “de Boer scale” [6]. Interim results show that the difference on this scale amounts roughly to 0.5 to 0.9 when comparing high pressure sodium light (2,000 K) with light of 3,000 K, in the advantage of high pressure sodium light. Comparing LEDs of 3,000 K with 8,000 K LEDs, the difference amounts roughly 0.5 to 0.7 in the advantage of the lower color temperature LEDs.

Glare Evaluation Systems

Systems for the practical evaluation of glare from lighting installations have been more or less independently developed for the different lighting applications fields. So the glare evaluation systems for interior lighting, for road lighting for motorized traffic and for pedestrians and slow moving traffic, for sports floodlighting, and for car head lighting are all different and will be dealt with separately.

Glare, Fig. 4 Geometric situation as the basis for UGR calculation



Interior Lighting

It is generally accepted that when in interior lighting discomfort glare is controlled sufficiently, disability glare will be limited sufficiently as well. Therefore glare evaluation systems for interior lighting normally are discomfort glare evaluation systems. Until the mid-1990s, many different discomfort glare evaluation systems for interior lighting were used in different parts of the world. All these systems take the four parameters described in the section above under “Discomfort Glare” into account. In 1995 the International Lighting Commission CIE introduced a consensus system called the “Unified Glare Rating (UGR)” system [10]. This system has since been widely adapted. The glare rating scale UGR has been developed so that a glare rating unit of 1 corresponds to the smallest detectable difference in discomfort glare. A high value indicates a high sensation of discomfort glare and a low value little discomfort glare. The practical range for most interior lighting installations is from 10 to 30. Most specifications for office lighting installations require a UGR value of maximum 19. For many industrial lighting installations, the limit is higher: 22 or 25.

For the geometric situation sketched in Fig. 4, UGR is given by the formula:

$$\text{UGR} = 8 \log \left\{ \frac{0, 25}{L_b} \sum \frac{\omega L_s^2}{p^2} \right\}$$

with:

L_b is the background luminance (cd/m^2).

L_s is the glare source luminance (cd/m^2).

ω is the solid angle of the bright parts of each luminaire at the observer's eye (steradian).

p is the Guth position index for each glare source.

The Guth position index values are given by CIE [10] in dependence of the displacement of the glare source from the line of sight, i.e., in dependence of R , T , and H in Fig. 4.

The UGR formula fulfills the requirement of additivity of multiple glare sources. Also when a glare source is split up in two glare sources with half dimensions but the same luminance, the UGR value remains the same.

The UGR formula can also be written as

$$\text{UGR} = 8 \log \left\{ \frac{0, 25}{L_b} \sum \frac{I_s^2}{A d^2 p^2} \right\}$$

with:

I_s is the intensity of the glare source in the direction of the observer.

A is the projected area (perpendicular on the sight of line) of the bright parts of the luminaire.

d is the distance between the glare source and the observer.

This makes it instantly clear that discomfort glare increases with increase in luminous intensity of the glare source, with decrease in size of the glare source (more compact source) and with decrease in background luminance. The latter means that UGR is also dependent on the reflectances of the room surfaces of the interior: the darker the walls and ceiling, the higher the risk for discomfort glare.

The UGR formula permits for detailed glare evaluations for different observer positions in the room and for different horizontal viewing directions. If one wants to judge the overall glare situation of an interior lighting installation with just one single UGR value, a reference observer and reference luminaire layout should be used as defined by ISO-CIE [11]. The observer position is standardized as being in the middle of each wall with the line of sight horizontal and straight ahead. The worst UGR value (highest value) of the resulting four values for these positions is taken as the value that has to fulfill the specification. The reference luminaire layout is defined as a uniform array of luminaires with 1:1 spacing to height ratio in the longitudinal and transversal directions (the height is taken as the luminaire height above eye level). Specifications of UGR values for these reference conditions are called limiting UGR values and are specified as UGR_L .

Road Lighting for Motorized Traffic

In the distant past, discomfort glare was part of the glare evaluation system for road lighting: de Boer's glare control mark [5, 6]. Today almost all glare evaluation systems for road lighting for motorized traffic are solely based on disability glare while they suggest that lighting installations designed to limit disability glare will be generally acceptable as regards discomfort glare [12].

In the section "Disability Glare," it has been concluded that the final disability effect of glare is dependent on the combination of the veiling luminance (L_v , according to the Holladay formula) and the scene luminance. Under road lighting conditions, the average scene luminance is defined by the average road surface luminance (L_{av}). The measure for disability glare for road lighting,

called threshold increment (TI), is dependent on L_v and L_{av} . It is defined as the percentage of extra contrast that would be required to compensate for the loss of contrast vision due to the actual glare situation (based on an object seen under an angle of 8 min of arc). For the typical road lighting conditions defined by the range of road lighting luminances commonly used for road lighting, an average screening angle of car windows of 20° and a viewing direction to a point on the road 90 m in front of the observer, TI is given by the relationship [11]:

$$TI = 65 \frac{L_v}{L_{av}^{0.8}}$$

TI is expressed in percent. In specifications for road lighting for motorized traffic, TI limits range from 10 % to 20 % dependent on traffic characteristics and type of road geometry.

Road Lighting for Slow Moving Traffic

The problem of glare is less critical for pedestrians than it is for motorists. This is primarily because of the much lower speeds involved where pedestrians are concerned. The pedestrian has far more time in which to adapt to changes in brightness in his visual field and is therefore less likely to be so blinded as to come into collision with an unseen obstacle in his path. A pedestrian is more likely to be troubled by discomfort glare than by disability glare. In practice, the severest glare sensation will be caused by individual bright luminaires appearing near to the direct line of sight. This is because the pedestrian continuously scans the whole environment to orientate himself and indeed regularly looks more or less straight into a luminaire. It is sensible, therefore, to limit the luminance of the individual luminaires for critical angles of emission. Of course the bright area of the luminaire plays here a role as well. As has been shown earlier, the more compact the light source, the higher the risk for discomfortable glare. CIE [13] defines on the basis of luminance of the luminaire under an elevation angle of 85° (L_{85}) and of the size of the bright part of the luminaire (A_{85}) the so-called luminaire glare index:

$$\text{Luminaire glare index} = L_{85} A_{85}^{0,5}$$

Limiting values for this luminaire glare index vary from 4,000 for mounting heights of up to 4,5 m to 7,000 for mounting heights higher than 6 m.

Car Head Lighting

Disability glare caused by car headlamps is evaluated by considering the loss of contrast vision because of the luminous veil created by the headlamps of oncoming cars. The Holladay formula for veiling luminance is less suitable for evaluations of glare from oncoming cars because the angle (θ) between the viewing direction and the direction of light from the oncoming car can be very small. The Holladay formula, where θ^2 is in the denominator, has validity only for angles larger than 1° . The Fry veiling luminance formula [14] has been shown to be suitable for the typical car headlamp conditions:

$$L_v = \frac{9, 2E_{eye}}{\theta(\theta + 1, 5)}$$

The most commonly used model for discomfort glare evaluations of car headlamps is that developed by Schmidt-Clausen and Bindels [15]. It expresses discomfort glare (W) in the nine-point scale of de Boer as described earlier (one standing for “unbearable glare” and nine for “unnoticeable glare”) with the expression

$$W = 5 - 2 \log \left\{ \frac{E_{eye}}{0, 02 \left(1 + \sqrt{\frac{L_{adapt}}{0, 04}} \right) \theta^{0,46}} \right\}$$

Outdoor Sports and Area Floodlighting

On sports grounds or outdoor working areas illuminated with floodlights, disturbing glare will occur for viewing directions directly toward the floodlights. Although this problem cannot be avoided, the practical consequences can be minimized by paying careful attention to the siting of the floodlights relative to the main viewing

directions for the situation and type of sport considered. Disturbing glare can, however, also occur to both players and spectators for viewing directions not directly toward the floodlights but toward the playing field. The degree of disturbance for these latter viewing directions is dependent upon factors such as the types of floodlights, their arrangement, mounting heights, and aiming directions. An extensive study in the 1980s looked for a relation between these parameters of the lighting installation and glare appraisals made by both players and laymen on many different floodlit playing grounds and stadiums [16]. Based on the results of this study, CIE [17] defined the glare rating GR:

$$GR = 27 + 24 \log \left(\frac{L_{vl}}{L_{ve}^{0,9}} \right)$$

with:

L_{vl} is the veiling luminance (according to Holladay formula) produced by the luminaires (floodlights) for an observer at the actual position on the field for the actual direction of view toward the field.

L_{ve} is the veiling luminance (according to the Holladay formula) produced by the lit environment.

L_{ve} can be approximated for most playgrounds and stadium layouts with the formula:

$$L_{ve} = 0, 035 L_{av}$$

with:

L_{av} is the average luminance of the horizontal area being observed from the actual position and for the actual viewing direction toward the field.

Limiting values for GR vary from 55 (lighting for sports training or lighting for work with very rough tasks) to 50 (lighting for sports competition or lighting for work with rough to medium tasks) and 45 (lighting for work with fine tasks).

Validity of Evaluation Systems for LEDs

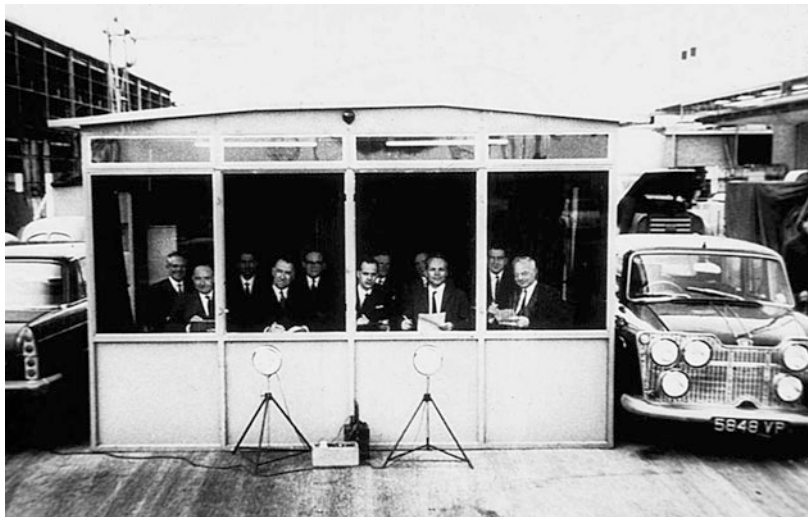
Figures 5, 6, and 7 show photographs of actual glare tests carried out for, respectively, interior lighting, road lighting, and sports floodlighting. It is important to realize that the research on which today's glare evaluation systems is based date mainly from the 1940s, 1950s, and 1960s (for sports floodlighting from the 1980s). Of course the glare effects on human beings have not been changed since then. However, the lighting

products in use have been changed tremendously. For example, the glare evaluation systems for interior lighting have only extensively been studied for opal and prismatic luminaires whereas already since long mirror optical systems are in use. The most important change has been the introduction of LED light sources. All the present-day glare evaluation systems have not originally been studied for the conditions prevailing with LED luminaires.

Glare, Fig. 5 Interior lighting glare study carried out in the 1950s [18]



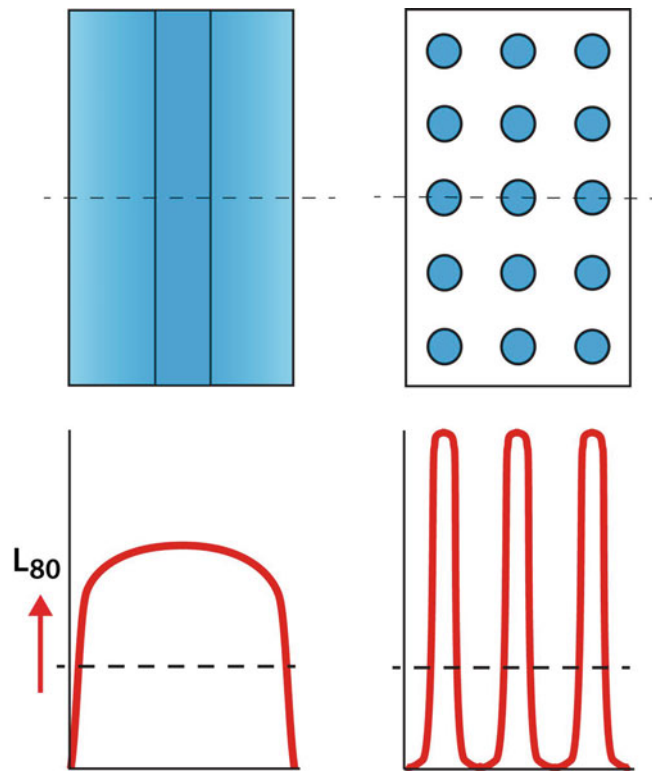
Glare, Fig. 6 Road lighting glare assessments being carried out in the 1950s [19]



Glare, Fig. 7 Sports floodlighting glare assessments being carried out in the 1980s [16]



Glare, Fig. 8 Uniform and nonuniform luminance distribution of the luminous part of a fluorescent tube and LED luminaire, respectively



LEDs can be produced in many more white-light spectrum versions than could be done with the earlier light sources. It has already been mentioned that there is an influence of glare source spectrum on discomfort glare. Some preliminary guidance has been given under the heading

“Discomfort Glare” at the beginning of this entry. More detailed research is however required.

LED luminaires usually consist of multi-LED panels and consequently the luminous surface of most LED luminaires is not uniform. Compare in Fig. 8 a typical fluorescent tube luminaire with

a LED luminaire with the same average luminance. It has been shown that it is not correct to use in the glare formulas of the various evaluation systems the straightforward average luminance of the LED luminaire [20]. Also in those formulas where the size of the bright parts of the luminaire plays a role, it is not self-evident what forms the size of the bright parts of a LED luminaire.

Individual LEDs are very small. With small optics around each individual LED of a multi-LED luminaire, it is possible to produce more pronounced light distributions than was possible with more conventional light source luminaires. The effects of these pronounced light distributions on glare have not been studied in the original glare evaluation studies but have an effect [21].

In many research institutions, research is being carried out on all typical LED aspects mentioned above. It will lead either to adaptations of the present glare evaluation systems or to complete new systems.

Cross-References

- [Automotive Lighting](#)
- [Interior Lighting](#)
- [Light Pollution](#)
- [Luminaires](#)
- [Road Lighting](#)
- [Sports Lighting](#)

References

1. CIE Publication: S 017/E: 2011, International lighting vocabulary. International Commission on Illumination, CIE, Vienna 17–492 (2011)
2. Holladay, L.L.: Action of a light source in the field of view in lowering visibility. *J. Opt. Soc. Am.* **14** (1927)
3. CIE Publication: 146:2002, CIE equations for disability glare. International Commission on Illumination, CIE, Vienna (2002)
4. Luckiesh, M., Guth, S.K.: Brightnesses in visual field at borderline between comfort and discomfort (BCD). *Illum. Eng.* 650–670 (1949)
5. CIE Publication: 31:1976, Glare and uniformity in road lighting installations. International Commission on Illumination, CIE, Vienna (1976)
6. De Boer, J.B.: Observations on discomfort glare in street lighting; influence of the colours of light, 13th CIE congress Zürich, CIE. International Commission on Illumination, CIE, Vienna (1955)
7. Bullough, J.D., Fu, Z., Van Derlofske, J.: Discomfort and disability glare from halogen and HID headlamp systems. SAE Technical paper series 2002-01-0010, Michigan (2002)
8. Akashi, Y., Asano, S., Kakuta, Y.: Visual mechanisms of discomfort glare sensation caused by LEDs, CIE Publication, x038:2013, Proceedings of CIE Centenary Conference “Towards a New Century of Light” Paris, 327–330, Vienna (2013)
9. Niedling, M., Kierdorf, D., Völker, S.: Influence of a glare sources spectrum on discomfort and disability glare under mesopic conditions, CIE Publication, x038:2013, Proceedings of CIE Centenary Conference “Towards a New Century of Light” Paris, 340–347, Vienna (2013)
10. CIE Publication: 117:1995, Discomfort glare in interior lighting. International Commission on Illumination, CIE, Vienna (1995)
11. ISO-CIE Publication, CIE S008/E-2001: International Standard, Lighting of indoor work places. International Commission on Illumination, CIE, Vienna (2002)
12. CIE Publication: 115:2010, Lighting of roads for motor and pedestrian traffic. International Commission on Illumination, CIE, Vienna (2010)
13. CIE Publication: 136:2000, Guide to the lighting of urban areas. International Commission on Illumination, CIE, Vienna (2000)
14. Fry, G.A.: Evaluating disability effects of approaching automobile headlights. *Highw. Res. Bull.* **89**, 38–42 (1954)
15. Schmidt-Clausen, H.J., Bindels, J.T.H.: Assessment of discomfort glare in motor vehicle lighting. *Light. Res. Technol.* **6**(2), 79–88 (1974)
16. Van Bommel, W.J.M., Tekelenburg, J., Fischer, D.: A glare evaluation system for outdoor sports lighting and its consequences for the design practice, CIE Proceedings: CIE 20st Session. Amsterdam, 1983, International Commission on Illumination, CIE, Vienna (1983)
17. CIE Publication: 112:1994, Glare evaluation system for use within outdoor sports and area lighting. International Commission on Illumination, CIE, Vienna (1994)
18. Balder, J.J.: Erwünschte Leuchtdichten in Büroräumen. *Lichttechnik* **9**, 455 (1957)
19. De Boer, J.B.: Fundamental experiments of visibility and admissible glare in road lighting, CIE 12th Session, Stockholm, CIE. International Commission on Illumination, CIE, Vienna (1951)
20. Higashi, H., Koga, S., Kotani, T.: The development of evaluation for discomfort glare in LED lighting of indoor work place: the effect of the luminance distribution of luminous parts on subjective evaluation, CIE Publication, x038:2013, Proceedings of CIE Centenary Conference “Towards a New Century of Light”, Paris, 648–662, International Commission on Illumination, CIE, Vienna (2013)
21. Zhu, X., Demirdes, H., Gong, X., Lai, J., Heynderickx, I.: The luminaire beam-shape influence on discomfort glare from Led road lighting luminaires, CIE Publication, x038:2013, Proceedings of CIE Centenary Conference “Towards a New Century of Light”, Paris, 648–662, International Commission on Illumination, CIE, Vienna (2013)

Glints

► [Texture Measurement, Modeling, and Computer Graphics](#)

Global Illumination

Tobias Ritschel
Department of Computer Graphics, MPI
Informatik, Saarbrücken, Germany

Synonyms

[Color bleeding](#); [Indirect illumination](#); [Interreflection](#); [Light bounces](#)

Definition

Global illumination (GI) is illumination reflected more than once inside a scene.

Introduction

After light was emitted, it has to be reflected before it can arrive at the eye of an observer or a sensor. When this reflection occurs once, it is called local or direct illumination. Illumination which has been reflected more than once is referred to as global or indirect illumination (GI). Reflection of light is sometimes called a bounce. If light is reflected two times from a surface, it is called the one-bounce indirect illumination; if it is reflected three times, it is called two-bounce; etc. As physical reflectance is below one, only a finite number of bounces contribute to the scene illumination.

Example

A simple example of GI is best visible when putting a rectangular achromatic cardboard next

to a perpendicular chromatic (e.g., red) piece of cardboard. When directional light is cast perpendicular onto the red cardboard (and consequently no light is cast on the achromatic cardboard), the white cardboard is still illuminated due to the indirect and colored (reddish) light from the chromatic (red) cardboard.

Phenomena

Depending on the reflective properties of the surfaces, GI can create different visual phenomena. For Lambertian surfaces (such as the aforementioned cardboard), light is reflected in all directions equally following a cosine distribution. For specular (mirrorlike) surfaces, light is distributed more in directions similar to the mirror direction (a Dirac distribution). The resulting concentration of light on the receiver side is called a caustic.

GI is subject to similar principles that govern local illumination: if a surface blocks the line of sight between a surface point A and another surface point B, no light is exchanged between A and B, and B is said to lie in A's indirect shadow. In the same way, a mirror can reflect indirect light as well as it can be refracted.

Strictly speaking, multiple reflections or refractions are also GI. However, the term GI is mostly used if the reflection is a true convolution that accounts for multiple directions and not only a single refracted or reflected one. At the same time, lighting from extended (area) light sources, such as the sky, is sometimes referred to as GI as well.

GI tends to be more colorful than local illumination, as light is filtered at every reflection, which is never perfectly white. This contributes to the color mood of a scene but can also be a difficulty, e.g., in white balancing where the illuminant is complicated.

Global illumination is present also in scenes without hard surfaces but participating media. Here, local illumination corresponds to single scattering, and global illumination corresponds to multiple scattering.

Formalization

Conceptually, linking all points where light is reflected results in a light path; the process of illumination along such a path is called light transport. The vertices of the light path can be labeled according to the type of reflection occurring. The light vertex is denoted as L , a diffuse reflection as D , a specular reflection as S , and the eye as E . As an example of this notation, “LDE” describes light that was emitted from the light and bounced diffusely once before arriving at the eye; “LDDE” denotes one indirect diffuse bounce, “LDSE” the reflection of a diffuse surface, and “LSDE” a caustic.

The rendering equation [1] is a formal way to describe global illumination. It defines the radiance L_o leaving a differential surface element at location \mathbf{x} with normal $n(\mathbf{x})$ in direction ω as

$$L_o(\mathbf{x}, \omega_o) = L_e(\mathbf{x}, \omega_o) + \int_{\Omega} L_i(\mathbf{x}, \omega_i) f_r(\mathbf{x}, \omega_i, \omega_o) < n(\mathbf{x}) \cdot \omega_i >^+ d\omega_i,$$

where L_e is the radiance emitted at location \mathbf{x} in direction ω , Ω is the space of directions in the upper hemisphere above \mathbf{x} , L_i is the incoming light arriving at \mathbf{x} from direction ω , f_r is the bi-directional reflectance distribution function (BRDF) which is the ratio outgoing radiance and incoming irradiance, and $<\bullet>^+$ is a dot product that is clamped at zero.

The difficulty of solving this equation is due to the fact that the incoming light and outgoing light appear on both sides of the equation. Solving the equation for a given scene means to find L such that the equation holds.

Computation

For computer graphics, GI is more challenging to compute than local illumination. Including GI in image synthesis leads to more realistic and visually appealing results. The color mood in a scene is to a large extent related to GI. Several approaches exist to compute it.

The most general approach is Monte Carlo ray tracing: For every pixel multiple rays that model geometric optics are intersected with the scene [2]. Whenever a ray hits a scene surface, it is continued in a random direction. After a number of hits, the iteration is stopped, and local as well as global illumination is accumulated according to the combined reflection properties at the hit positions and the local illumination at the last hit. This procedure can simulate all light effects but takes long to compute or results in noisy images.

Radiosity takes a different approach and discretizes the scene surface into finite elements and simulates the energy exchange between them by solving a system of equations [3]. Its basic principle is the form factor that describes the light arriving at element A from element B. Radiosity is best at simulating diffuse indirect illumination but can be extended to specular light transport. It does not scale well to large scenes or highly specular materials.

If only very few paths from the light illuminate a location, it will appear noisy using Monte Carlo ray tracing. As a solution, instead of starting light paths at the eye, they can be started at the light source in a process called light tracing.

When paths are started from both the eye and the light and end points of sub-paths are connected, the procedure is called bi-directional path racing [4].

For notoriously difficult light paths, where only a low percentage of all possible paths contribute a lot of energy to an image, the solution to the rendering equation can be found using Metropolis light transport sampling [5]. The approach generates random light paths, which are changed (mutated) such that more relevant paths are explored in greater detail than less relevant paths.

At third alternative is photon mapping [6]. It is performed in two passes. The first one is similar to light tracing but stores the hit locations into a spatial data structure (the photon map). In the second path, rays are traced from the eye, but every time a surface is hit, the illumination is constructed from the local density of photons inside the photon

map. Photon mapping is the best established method to render caustics.

Several approximate approaches to compute GI exist. Instant radiosity [7] approximates the indirect lighting using a high number of virtual point lights. Lightcuts [8] is an approach to compute GI from the high number of virtual point light in sublinear time. GI rendering at interactive frame rates can be achieved using graphics hardware [9].

If GI is due to extended light sources, and not due to multiple bounces, Image-based lighting (IBL) is applicable [10]. Here, the light entering a scene from a direction is stored into every pixel of a (high dynamic range) spherical raster image. To use this image for rendering, it is preprocessed once and can later be used to compute illumination easily.

A simplified version of GI is ambient occlusion (AO) [11]. AO is the special case of illumination by a constantly white, infinitely extended directional area light source. The resulting illumination is similar to the one found on a cloudy day and contains only directionless and soft shadows.

Cross-References

- [Non-Photorealistic Rendering](#)
- [Sampling Problems in Computer Graphics](#)

References

1. Kajiya, K.: The rendering equation. *Comput. Graph.* (Proc. ACM SIGGRAPH) **20**(4) (1986)
2. Pharr, M., Humphreys, G.: *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann, San Francisco (2010)
3. Cohen, M.F., Wallace, J.R.: *Radiosity and Realistic Image Synthesis*. Morgan Kaufmann (1993)
4. Lafortune, E., et al.: Bi-directional path tracing. *Proc. Comput. Graph.* 145–153
5. Veach, E.: *Robust Monte Carlo methods for light transport simulation*. Ph.D. thesis, Stanford University (1997)
6. Jensen, H.W.: *Realistic Image Synthesis Using Photon Mapping*. AK Peters, Natick (2001)
7. Keller, A.: Instant radiosity. *Proc. SIGGRAPH* (1997)
8. Walter, B., et al.: Lightcuts: a scalable approach to illumination. *ACM Trans. Graph.* (Proc. SIGGRAPH). **24**(3) (2005)
9. Ritschel, T., et al.: The state of the art in interactive global illumination. *Comput. Graph. Forum* **31**(1), 160 (2012)
10. Debevec, P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Proc. SIGGRAPH* (1998)
11. Zhukov, S., et al.: An ambient light illumination model. *Proc. Rendering Tech.* 45–55 (1997)

Gloss Meter

Bor-Jiunn Wen

Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

Synonyms

[Glossmeter](#)

Definition

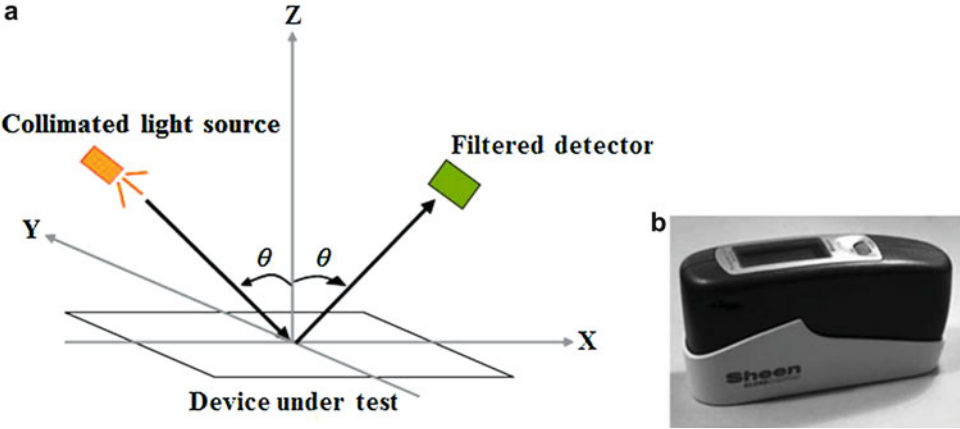
A gloss meter is an instrument that is used to measure the specular reflection gloss of a surface. Gloss is determined by projecting a beam of light of an intensity at a fixed angle onto a surface and measuring the amount of reflected light at an equal but opposite angle.

Introduction

The gloss meter measures gloss intensity under a fixed set of illumination and viewing conditions [1–3]. Figure 1 shows the configuration of both illumination source and a detector across a limited range of observation reception angles. The measurement results of a gloss meter are related to the amount of reflected light from a black glass standard with a defined refractive index. The ratio between the reflected and incident light for the specimen across a range of angles is compared with the ratio for the gloss standard, and is recorded as gloss units (GU).

Applications

There are a number of different geometries available for gloss measurement, according to the types of



Gloss Meter, Fig. 1 (a) Schematic diagram of gloss measurement and (b) Gloss meter

surface to be measured. For non-metals, such as coatings and plastics, the amount of reflected light increases with a greater angle of illumination, as some of the light penetrates into or is reflected from the surface, which is absorbed or scattered in the material depending on its color. Metals have a much higher reflection and are therefore less angularly dependent.

Many international or national standards are available according to the different types of gloss and materials, such as paint, ceramics, paper, metals, plastics, etc. For example, ISO 2813 specifies the measurements of specular gloss in the paint industry. Many industries use specific gloss meters for their quality control work to ensure consistency in their manufacturing processes. The automotive industry is a major user of the gloss meter, with applications extending from the factory floor to the repair shop.

In the case of transparent materials, the gloss readings can be increased up to 2000 GU because of multiple reflections in the bulk of the material. For these applications, it is common to report the measurement results in percentage reflection against the incident light. For each application, the types of gloss meter and their operation procedures must be strictly specified for inter-comparison. Of the above, the angle of illumination is of most importance. Table 1 defines the measuring angle depending upon different surface characteristics: matte, glossy, and semi-glossy. As can be seen in Table 1, the gloss range means the initial measurement based on 60° geometry. The reading

Gloss Meter, Table 1 Gloss type

Gloss type	Gloss range	Measure with
Semi-gloss	10–70	60° geometry
High gloss	>70	20° geometry
Low gloss	<10	85° geometry

is used to determine which type of geometry is to be used. For example, if the initial reading is 40 GU, the result will be 40 GU because it is within the semi-gloss range, using only 60° geometry. If the initial reading is 78 GU, a new reading should be obtained using a 20° geometry gloss meter.

Gloss Meter Industry Standards

The following lists various national or international standards for measuring gloss [4]. Many industries have adopted 20°/60°/85° geometries as specified in ISO2813/ASTM D523.

General Gloss Measurement

Standards	Description
ASTM D523 1999 (USA)	Test method for specular gloss. The principal ASTM specular gloss standard. Very similar to ISO 2813
ASTM D3928 1998 (USA)	Test method for the evaluation of gloss or sheen uniformity

(continued)

Standards	Description
ASTM D4039 1999 (USA)	Test method for the reflection haze of high-gloss surfaces
ASTM D4449 1999 (USA)	Test method for the visual evaluation of gloss differences between surfaces of similar appearance
ASTM D5767 1999 (USA)	Test methods for the instrumental measurement of the distinctness of image gloss of coating surfaces
ASTM E340 1997 (USA)	Test methods for the measurement of the gloss of high-gloss surfaces by goniophotometry
MFT 30–064 (South Africa)	Local version of ASTM D523
JIS Z8741 1997 (Japan)	Method of measurement for specular glossiness

Paint

Standards	Description
ISO 2813 1994 (International)	Paints and varnishes – determination of the specular gloss of non-metallic paint films at 20°, 60°, and 85°. The principal ISO specular gloss standard. Very similar to ASTM D523
BS 3900: part D5 1995 (UK)	Methods of testing for paints – optical tests on paint films – measurement of specular gloss of non-metallic paint films at 20°, 60°, and 85°
DIN 67530 1982 (Germany)	Reflectometer as a means of assessing the specular gloss of smooth painted and plastic surfaces
NFT 30–064 1999 (France)	Paints – measurement of specular gloss at 20°, 60°, and 85°
AS 1580 MTD 602.2 1996 (Australia)	Paints and related materials, methods of testing – introduction and list of methods
JIS Z8741 1997 (Japan)	Specular glossiness – method of measurement
SS 18 41 84 1982 (Sweden)	Paints and varnishes – measurement of the specular gloss of non-metallic paint films at 20°, 60°, and 85°

Plastics

Standards	Description
BS 2782: Pt 5, Method 520A 1992	Methods of testing plastics – optical and color properties, weathering – determination of specular gloss. Similar to ISO 2813
ASTM D2457 1990	Test method for the specular gloss of plastic films and solid plastics specifies the primary standard as a perfect mirror with a defined gloss value of 1000. 20°, 60° and 45°; the 45° method is the same as ASTM C346 for ceramics

Metals

Standards	Description
BS6161: part 12 1987	Methods of testing for anodic oxidation coatings on aluminum and its alloys – measurement of specular reflectance and specular gloss at angles of 20°, 45°, 60° or 85°. Reference standard BS 3900: part D5 (1980); technically equivalent to ISO 7668; replaces BS 1615:1972. At 45°, the dimensions of the source image and receptor aperture are as for 60°. Squares with sides equal to the shorter sides of the rectangles are also recommended. Alternatively, total reflection in a 45° prism is used as a reference; source image and receptor aperture are then circular, both with an angular diameter $3.44^\circ \pm 0.23^\circ$. (1.5 mm \pm 0.1 mm at a focal length of 25.4 mm)
ISO 7668 1986	Anodized aluminum and aluminum alloys – measurement of specular reflectance and specular gloss at angles of 20°, 45°, 60° or 85°
ISO 5190	Anodizing of aluminum and its alloys – evaluation of the uniformity of the appearance of architectural anodic finishes – determination of diffuse reflectance and specular gloss. ECCA T2 (European Coil Coating Association). Specular gloss at 60°

Paper

Standards	Description
DIN 54502 1992	Testing of paper and board; reflectometer as means of assessing the gloss of paper and board
ASTM D1223 1998	Test method for the specular gloss of paper and paperboard at 75°. Has unusual converging beam geometry. Specifies the primary standard as black glass of a refractive index of 1.540, not 1.567, at the sodium D-line, having a defined gloss value of 100
ASTM D1834 1995	Test method for 20° specular gloss of waxed paper. Another unusual converging beam geometry, different from the previous one
TAPPI T480 OM-90 1990 (USA)	Specular gloss of paper and paperboard at 75°. Same text as for ASTM D 1223
TAPPI 653 1990	Specular gloss of waxed paper and paperboard at 20°. Same text as for ASTM D 1834
JIS – Z8142 1993 (Japan)	Testing method for 75° specular gloss

Furniture

Standards	Description
BS 3962: part 1 1980	Methods of testing the finishes of wooden furniture – assessment of low angle glare by measurement of specular gloss at 85°. Similar to ISO 2813: 1978

Floor Polish

Standards	Description
ASTM D1455 1987	Test method for 60° specular gloss of emulsion floor polish. Reference standard ASTM D 523

Ceramics

Standards	Description
ASTM C346 1987	Test method for 45° specular gloss of ceramic materials. Reference standard ASTM D 523
ASTM C584 1981	Test method for 60° specular gloss of glazed ceramic whitewares and related products. Reference standard ASTM D 523 {Sheen}

Fabrics

Standards	Description
BS 3424: method 31: part 28 1993	Testing coated fabrics – determination of specular gloss

Cross-References

- [Reflectance Standards](#)
- [Standard Measurement Geometries](#)

References

- Hunter R.S.: Methods of Determining Gloss. NBS Research Paper RP 958, J. Res. NBS, 18, 77, 281 (1937)
- Jones, L.A.: The gloss characteristics of photographic papers. J. Opt. Soc. Am. 6, 140 (1922)
- CIE 15: Colorimetry, 3rd edn (2004)
- Gloss Meter Industry Standards. Available at: <http://www.gloss-meters.com/GlossIntro5.html>. Accessed 21 Mar 2014

Glossmeter

- [Gloss Meter](#)

Glow Discharge Lamps

- [Neon Lamp](#)

Goethe, Johann Wolfgang von

Rolf G. Kuehni
Charlotte, NC, USA



J. K. Stieler, 1828

Biography

Goethe was born on August 28, 1749, in Frankfurt, Germany, to a lawyer and the daughter of the mayor of Frankfurt. He studied law in Leipzig and Strasbourg (France). Based on his early fame as a poet and novelist, Goethe, at age 26, was invited by the 18-year-old Duke of Sachsen-Weimar to join his court as an advisor. He moved to Weimar where he remained until his death on March 22, 1832. At times he worked as a diplomat and a public servant for the duke. He became known as the preeminent German poet, novelist, and playwright. The philosophical nature of much of his writing and his battle with Kant influenced several later philosophers, such as Hegel, Schopenhauer, Nietzsche, and Wittgenstein. Goethe's

interests were very broad, including pictorial art and certain aspects of science, most importantly in the latter the metamorphosis of plants, where he displayed a pre-Darwinian point of view, as well as the source and nature of color experience. He considered his *Farbenlehre* (Theory of Colours) [1] the most important of his works.

Goethe's views on science were strongly influenced by the seemingly very complex relationship between objectivity and subjectivity. He believed strongly in experimentation, with variability in conditions and multiple replications, which might lead to objective truths. Circa 1793, Goethe wrote an essay *Der Versuch als Vermittler von Objekt und Subjekt* (The Experiment as Mediator Between Subject and Object) which he shared with his friend, the poet Friedrich Schiller, in 1798 in which he described the value of scientific experiments as a bridge between subjectivity and objectivity and the possibility of establishing via multiple experiments of this kind the existence of objective laws of nature [2]. It was only published in print in 1823.

Major Accomplishments/Contributions

His interest in color began early in his life and received input during his extended journey to Italy in 1787 where he began studying Leonardo da Vinci's manuscripts on painting. His first writings on the subject of color date from 1791 *Beiträge zur Chromatik* [Contributions to Chromatics], 1792 *Von den Farbigen Schatten* (Concerning Colored Shadows), and 1794 *Versuch, die Elemente der Farbenlehre zu entdecken* (Attempt at Discovering the Elements of Color Theory). *Zur Farbenlehre*, in three volumes, was published in 1810, the same year P. O. Runge (well acquainted with Goethe) published his *Farben-Kugel* (Color Sphere). The first volume of *Farbenlehre* presents his description and interpretation of physiological colors, physical colors, and chemical colors; a chapter on the possibly singular natural source of color experience; a chapter about the relationship of color science to other intellectual domains such as philosophy, mathematics, general physics, natural history,

etc.; and one on the general esthetics of color. Volume 2 is an extended attack on and critique of the color work of Isaac Newton, its main component being the “unnatural” methodology used by Newton when splitting daylight into its spectral components, replaced by Goethe’s method of investigating the effect of the prism when viewing contrasting boundaries in daylight. The result in the former case is the classical spectrum of colors and in the latter the series of so-called boundary or edge colors (for an informative discussion of boundary colors, see Ref. [3]). Recent complex optical experiments have shown that there is no discrepancy between Newton’s and Goethe’s findings (e.g., Ref. [4]). The third volume consists of an extended, if somewhat prejudiced, review of the history of color science (*Geschichte der Farbenlehre*), beginning with the ancient Greeks and ending with contemporaries of Goethe.

Goethe’s fundamental view of perceived colors was that they are the result of interaction between lightness and darkness, an idea already introduced by Aristotle. Goethe saw it confirmed by his experience with boundary colors when viewing edges of white and black fields through a prism, for example, viewing through it the hexagonal charts of Fig. 1, illuminated by daylight.

Goethe devised a color circle consisting of six hues in three complementary pairs, with a purplish red on top, followed on the left, warm side by orange and yellow, and on the cold right side by violet and blue, with green on the bottom. Colors diametrically opposite are approximately complementary. Such opposing colors were viewed by Goethe as harmonic pairs. He also used their relationships in the circle to demonstrate presumed psychological effects.

The response to Goethe’s *Farbenlehre* was mixed. However, in a commentary of 1892 Hermann von Helmholtz said: “And I for one do not know how anyone, regardless of what his views about colors are, can deny that the theory [Goethe’s] in itself is fully consequent, that its assumptions, once granted, explain the facts treated completely and indeed simply” [5].

The English painter Charles Eastlake published in 1840 a translation of Part I as “Theory of Colours” [6]. It became an important source



Goethe, Johann Wolfgang von, Fig. 1 Partial collection of color-related equipment and materials owned by Goethe in the Goethe-Nationalmuseum in Weimar

of information for painters, such as J. M. W. Turner. In an introduction to a modern version of the translation in 1970 (remaining in print today), D. B. Judd wrote: “This book can lead the reader through a demonstration course not only in subjectively produced colors (after images, light and dark adaptation, irradiation, colored shadows, and pressure phosphenes), but also in physical phenomena detectable qualitatively by observation of color (absorption, scattering, refraction, diffraction, polarization, and interference)” [7].

References

1. von Goethe, J. W.: *Zur Farbenlehre*, 3 vols. Cotta, Tübingen (1810)
2. von Goethe, J.W.: *Gedenkausgabe der Werke, Briefe und Gespräche*, Bd. 16, pp. 844–855. Artemis Verlag, Zürich (1949)
3. Bouma, P.J.: *Physical Aspects of Colours*, 2nd edn. St. Martin’s Press, New York (1971)
4. Rang, M.: Goethe’s *Farbenlehre* und ihre technische “Aufrüstung” – nicht gegen Newton, sondern mit Newtonscher Optik. In: Wolf Schmidt, G. (ed.) *Colours*

- in Culture and Science, pp. 246–273. Tredition, Hamburg (2011)
5. von Helmholtz, H.: *Goethes Vorahnungen kommender naturwissenschaftlicher Ideen*. Paetel, Berlin (1892)
 6. von Goethe, J. W.: *Goethe's Theory of Colours* (trans.: Eastlake, C. L.). Murray, London (1840)
 7. von Goethe, J. W.: *Goethe's Theory of Colours* (trans.: Eastlake, C. L.). MIT Press, Cambridge, MA (1970). Introduction by D. B. Judd

Goniochromatism

► Iridescence (Goniochromism)

Goniometer

► Goniophotometer

Goniophotometer

Cheng-Hsien Chen
Center for Measurement Standards, Industrial
Technology Research Institute, Hsinchu, Taiwan

Synonyms

Goniometer

Definition

A goniophotometer is an instrument for measuring light distribution in terms of the luminous intensity or illuminance distribution of a light source, a luminaire, a self-luminous display, or a surface by changing the position of a photometer or a goniometer at a clearly defined geometry. A typical goniophotometer usually consists of a mechanical device or robotic jib to support the luminaire to be measured and a photometer or a

spectroradiometer, and allows their relative position to be changed to form a spherical locus [1].

By analyzing the optical data of the definite angle of spectral irradiance, the total luminous flux, spectral radiant flux, and color quantities can be obtained. Spectral radiant flux refers to the sum of all the radiant energy emitted per unit of time, by a light source over a specific range of a spectrum. Spectral radiant flux is the basis for deriving important luminaire parameters such as total luminous flux, radiant flux, luminous efficacy, chromaticity, correlated color temperature, color rendering index, etc.

Overview

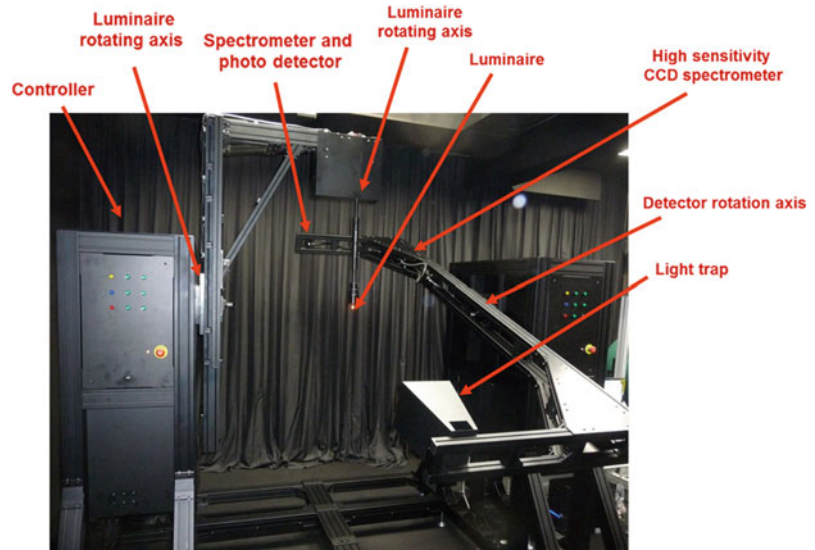
A typical goniophotometer includes a luminaire and a photometer. There are three common types of goniophotometer, according to the different geometrical arrangements of the two components. The first type of goniophotometer rotates the luminaire along the two assigned perpendicular axes with the photometer fixed. The second type of goniophotometer rotates the photometer instead of the luminaire. The third type of goniophotometer rotates the luminaire along one axis, and its photometer moves along the other axis. The relative orientation of the luminaire and the photometer is then adjusted [2–4].

However, it is complicated to design a goniophotometer, compared with an integrating sphere to measure the luminous flux of the luminaire, because goniophotometric measurements require a large, dark room, a long period for measurement, and great precision of movements. Moreover, for some designs, the luminaire must be rotated, which can cause errors because of its sensitivity to the burning position and draft of the radiant flux. Some luminaire-mounting mechanisms of the goniophotometer can also generate shadow, which affects the luminous flux measured [5, 6] (Fig. 1).

The spectral radiant flux measurement goniophotometer uses a spectroradiometer to measure the spectral power distribution of the

Goniophotometer,

Fig. 1 The International Commission on Illumination type C, spectral-based goniophotometer



luminaire at a specific angle. In practice, the spectroradiometer should be calibrated, and the stray light during measurement should be eliminated to achieve measurement accuracy. Stray light is one of the most common sources of measurement inaccuracies for the spectroradiometer. Traditionally, a tunable laser was used to correct the stray light [7]. A novel spectral stray light correction technique for spectroradiometer was then developed. It removes, primarily through an optical frequency comb technique, all stray light that does not fall within a certain fixed wavelength range. This prevents stray light, from indeterminate light sources within the spectroradiometer, from interfering with the measurement process, thus enhancing the accuracy of the measurement [8].

Figure 2 shows a spectral-based goniophotometer, including a detector head module and photodiode array type spectroradiometer set in the jib of the goniophotometer. The luminaire is located in the center of the goniophotometer. The detector head module is arranged according to different predetermined geometries. The key to achieving great measurement accuracy is to avoid rapid movement, which can cause air flow and vibration.

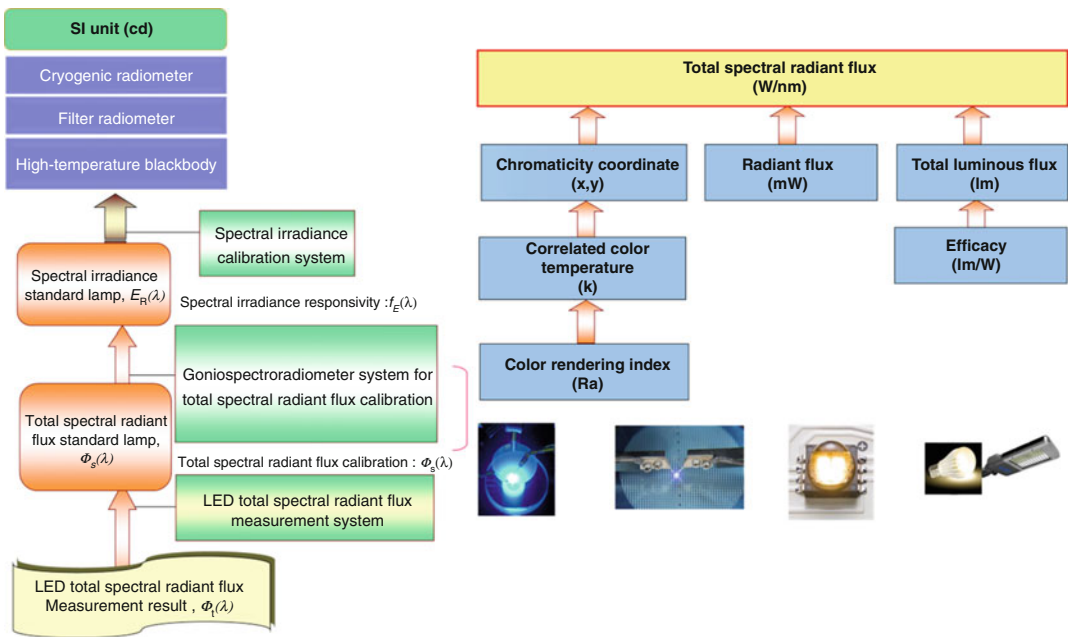
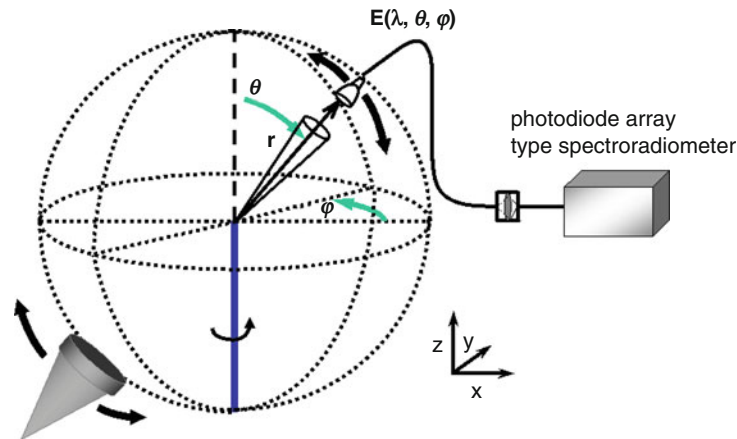
The spectral radiant flux, by integrating the spectral irradiance in the definite angle, may be calculated using the equation below:

$$\Phi(\lambda) = r^2 \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} E(\theta, \phi, \lambda) \sin \theta d\theta d\phi \quad (1)$$

where $\Phi(\lambda)$ is the spectral radiant flux in W/nm unit; $E(\lambda)$ is the spectral irradiance in W/m² nm unit, and r is the radius between the reference plane of the detector and the luminaire.

Light-emitting diode (LED) lighting parameters consist of lighting comfort, product safety, lifespan, and characteristics. Spectral radiant flux is the basis for radiometry and chromaticity parameters and specifications, from which total luminous flux, radiant flux, luminous efficacy, chromaticity, correlated color temperature, color rendering index, etc., can be derived, as shown in Fig. 3. LED total spectral radiant flux is measured using an LED total spectral radiant flux measurement system. The total spectral radiant flux standard lamp carries the total spectral radiant flux and transfers the standard from the goniophotometry measurement system. By using a spectral irradiance standard lamp, the spectral radiant flux standard is transferred to the irradiance standard and traced to SI units. Therefore, the spectral radiant

Goniophotometer,
Fig. 2 Schematic diagram
of a spectral-based
goniophotometer



Goniophotometer, Fig. 3 The metrological traceability chain of the total spectral radiant flux standard lamp calibration system

flux measurement standard can be traced to the SI unit, the candela (cd).

Compared with the sphere-based instruments, a spectral-based goniophotometer

(goniospectroradiometer) provides absolute measurement with greater accuracy, than a calibrated flux standard lamp with a known total spectral radiance. Goniophotometers also report the spatial light

distribution data and measure the luminous flux of the particular light distribution of a luminaire, such as a parabolic aluminized reflector lamp (PAR lamp) or a curved lamp. Also, a photodiode array type spectroradiometer should have spectral response calibration, stray light correction, sensitivity, and linearity certification to ensure accurate measurement.

Cross-References

- [Spectral Power Distribution](#)
- [Spectroradiometer](#)
- [Standard Lamp](#)

References

1. International Commission on Illumination.: ILV: International Lighting Vocabulary. CIE S017/E: 2011, Vienna (2011)
2. International Commission on Illumination.: The Photometry and Goniophotometry of Luminaires. CIE 121:1996, Vienna (1996)
3. International Commission on Illumination.: The Measurement of Absolute Luminous Intensity distribution. CIE 70:1987, Vienna (1987)
4. Illuminating Engineering Society of North America.: Electrical and Photometric Measurements of Solid-State Lighting Products. IES LM-79-08, America (2008)
5. Chen, C.H., Pong, B.J., Jia, Y.D., Lin, H.L.: Lamp orientation dependence of an integrating sphere response for directional light source in luminous flux measurement. *NCSLI Measure J. Meas. Sci.* **8**(2), 42–47 (2013)
6. Lindemann, M., Maass, R.: Photometry and colorimetry of reference LEDs by using a compact goniophotometer. *MAPAN* **24**(3), 143–152 (2009)
7. Zong, Y., Brown, S.W.: Simple spectral stray light correction method for array spectroradiometers. *Appl. Optics* **45**(6), 1111–1119 (2006)
8. Lui, T.A., Wu, G.N., Hsu, S.W., Pong, J.L., Zong, T.Y.: Optical spectrometer calibration technique based on mode-locked laser comb. *Meas. Inf* **154**, 45–48 (2013)

Graininess

- [Texture Measurement, Modeling, and Computer Graphics](#)

Grassmann, Hermann Günther

Rolf G. Kuehni
Charlotte, NC, USA



Grassmann portrait

Biography

Grassmann was born on April 15, 1809, in Stettin in Pomerania, near the Baltic Sea (today Szczecin in Poland), the third of 11 children of a pastor and high school mathematics teacher and his wife. After passing through high school, Grassmann moved to Berlin to study theology, with later addition of mathematics and sciences. In 1844 he published his major work in mathematics, *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik* (linear extension theory, a new branch of mathematics) [1], a general calculus of vectors. *Ausdehnungslehre* was not a success, apparently clearly ahead of its time. In 1846 Grassmann received an award for expanding on a mathematical problem sketched earlier by Leibniz. Grassmann married in 1849, and he and his

wife had 11 children. His father, though teaching at a high school, had been named professor a few years before he passed away in 1852. In that same year Hermann Grassmann assumed the position of mathematics professor his father had held at the Stettin Gymnasium. In his later years, unhappy about the continuing lack of attention to his mathematical efforts, he became interested in the history of languages, learned Sanskrit, and prepared a dictionary and a translation of the sacred collection of Indian Vedic hymns, the *Rigveda*, one of the oldest extant written records in an Indo-European language, dating to the mid-second millennium BCE [2]. Both works immediately gained much admiration and support from linguists. Grassmann died on September 2, 1877, in Stettin.

Major Accomplishments/Contributions

In 1852 Hermann von Helmholtz published an article, based on his results of experimental work in color mixture with a spectroscope of his own design, in which he concluded that Newton's structural design of his color circle, based on his own experiments with mixing spectral lights, must be in error and that there are only two spectral colors, blue and yellow, that when mixed result in colorless appearance [3]. Grassmann applied mathematical logic to the problem and in 1853 published a paper *Zur Theorie der Farbenmischung* (On the Theory of Color Mixture) [4] in the same journal as Helmholtz, claiming that Helmholtz was likely in error. He postulated four "assumptions" about color mixture:

1. Every impression of color may be analyzed into three mathematically determinable elements – hue, intensity of color, and brightness of the intermixed white.
2. If one of two mingling lights is continuously altered (while the other remains unchanged), the impression of the mixed light is also continuously changed.
3. Two colors, both of which have the same hue and the same proportion of intermixed white,

also give identical mixed colors, no matter what homogeneous colors they may be composed of.

4. The total intensity of any mixture is the sum of the intensities of the lights mixed.

A modern interpretation of the content of assumptions 2–4, as provided by Wyszecki and Stiles [5], are the following four laws:

1. Symmetry law: If color stimulus **A** matches stimulus **B**, then stimulus **B** matches stimulus **A**.
2. Transitivity law: If **A** matches **B** and **B** matches **C**, then **A** matches **C**.
3. Proportionality law: If **A** matches **B**, then **aA** matches **aB**, where **a** is a positive factor of the radiant power of the stimulus.
4. Additivity law: If **A** matches **B** and **C** matches **D**, then **(A + D)** matches **(B + C)** (applicable to additive mixtures).

These laws do not explicitly consider variations in conditions, in eye adaptation, or variation in color-matching functions between observers [see e.g., 6].

The laws are considered fundamental components of a trichromatic theory of color vision. In response to Grassmann's paper, after modifying his spectroscope, Helmholtz redid the color-mixing experiments and was able to determine multiple complementary pairs in the spectrum, thus confirming the basic validity of Grassmann's laws [7].

References

1. Grassmann, H. G.: Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik. Leipzig: self-published by Grassmann (1844). A second edition was published in 1862: Die Ausdehnungslehre, Vollständig und in strenger Form bearbeitet. Enslin, Berlin
2. Grassmann, H. G. A.: Wörterbuch zum Rig-Veda. Brockhaus, Leipzig (1873–1875). Rig-Veda (1876–77). Erster Theil: Die Familienbücher des Rig-Veda; Zweiter Theil: Sammelbücher des Rig-Veda. Leipzig: Brockhaus.
3. Helmholtz, H. v.: Über die Theorie der zusammengesetzten Farben. Unger, Berlin (1852)

4. Grassmann, H. G.: Zur Theorie der Farbenmischung. Poggendorffs Ann. Phys. Chem. **89**, 69–84 (1853). English translation (1854): On the theory of compound colors. Philos. Mag. **4**, 254–264. An edited and abbreviated version with explanatory remarks is found in MacAdam, D. L.: Sources of Color Science, pp. 53–60. MIT Press, Cambridge, MA (1970)
5. Wyszecki, G., Stiles, W.S.: Color Science, 2nd edn, p. 118. Wiley, New York (1982)
6. Brill, M. H., Robertson, A. R.: Open problems in the validity of Grassmann's laws. In: Schanda, J. (ed.) Colorimetry, Understanding the CIE System. Wiley, Hoboken (2007)
7. Helmholtz, H. v.: Über die Zusammensetzung von Spectralfarben. Poggendorffs Ann. Phys. Chem. **95**, 1–28 (1855)

Green-Blindness

► Deuteranopia

Guild, John

Robert W. G. Hunt
Department of Colour Science, University of
Leeds, Leeds, UK



John Guild was a British scientist who worked at the National Physical Laboratory (NPL) at Teddington in England. Guild spent the majority of his professional life making outstanding contributions to the development of a wide variety of optical instruments and techniques [1].

In the 1920s he wrote several papers, published in the Transactions of the Optical Society, describing the fundamentals of colorimetry which have formed the basis of the discipline ever since.

He measured, for seven observers, the way in which the colors of the spectrum are matched by beams of red, green, and blue light added together. This work, together with a similar study carried out by David W. Wright, with ten additional observers, formed the basis of the international standard for measuring color, the CIE 1931 Standard Colorimetric Observer. The quality of this experimental work was so high that this standard, although now more than 80 years old, is still in universal use. Guild was largely responsible for making this work the basis of the international system of colorimetry that is still in use today. In collaboration with T. Smith, also of the NPL, he devised the transformation of the experimental results into the XYZ system that is the form in which the standard is used.

In spite of having played a key role in establishing the basis and original standard of colorimetry, after 1931 Guild transferred his interests to other spheres of work and took little further part in the subject. However, he retained his interest in colorimetry and was instrumental in the formation of the Colour Group in the United Kingdom where he was the second chairman of the group (1943–1945) and one of the first honorary members in 1966 [1].

Some of his more important and relevant publications pertaining to this field are listed:

- J. Guild: An equipment for visual spectrophotometry. Trans. Opt. Soc. **26**, 74–94 (1924–1925)
- J. Guild: The transformation of trichromatic mixture data: algebraic methods. Trans. Opt. Soc. **26**, 95–108 (1924–1925)
- J. Guild: The geometrical solution of colour mixture problems. Trans. Opt. Soc. **26**, 139–174 (1924)

- J. Guild: A trichromatic colorimeter suitable for standardisation work. *Trans. Opt. Soc.* **27**, 106–129 (1925–1926)
- J. Guild: A criticism of the monochromatic-plus-white method of colorimetry. *Trans. Opt. Soc.* **27**, 130–138 (1925)
- J. Guild: On a new method of colorimetry. *Trans. Opt. Soc.* **27**, 139–160 (1925–1926)
- T. Smith, J. Guild: The CIE colorimetric standards and their use. *Trans. Opt. Soc.* **33**, 73–134 (1931–1932)
- J. Guild: The colorimetric properties of the spectrum. *Phil. Trans. Roy. Soc. (London) A* **230**, 149–187 (1931)

References

1. The Colour Group (Great Britain): The Advancement of Colour, A History of the First Fifty Years of the Colour Group (Great Britain). The Colour Group (Great Britain), London (1941–1991)

H

Halftone Layer Superposition Equations

► [Demichel Equations](#)

Halogen Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Iodine lamps](#); [Quartz halogen lamps](#)

Definition

Lamps that produce light as a result of an electrical current through a metal wire, contained in a transparent small bulb that heats the metal to incandescence. The bulb also contains halogen that reacts with the vaporized metal so that bulb blackening by vaporized metal is minimized.

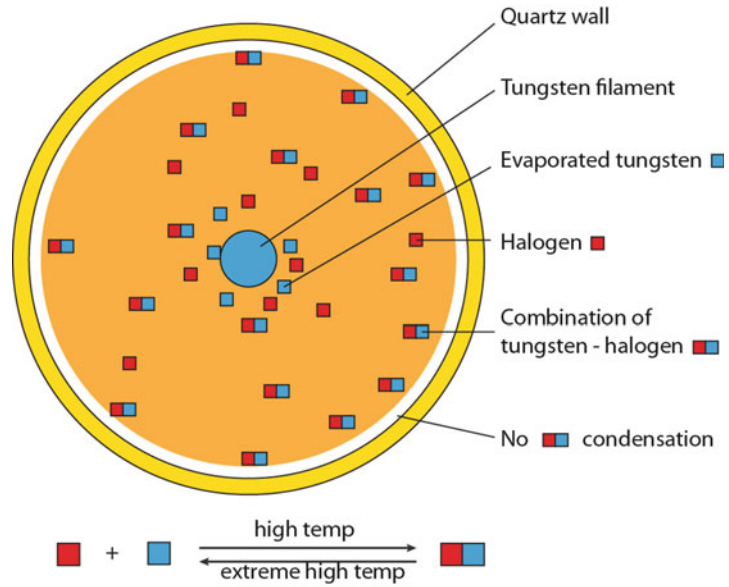
Halogen Lamps

During the operation of an incandescent lamp, tungsten evaporates from the filament and settles on the coldest place inside the lamp (the bulb wall), causing lamp blackening, which leads to a considerable depreciation of the light output during lamp life. In order to keep lamp blackening, and the corresponding light loss, within acceptable limits, bulbs of normal incandescent lamps are relatively large. But in order to be able to operate small-bulb high-light-output incandescent lamps, special measures have to be taken to prevent bulb blackening, which in these small bulbs would quickly lead to unacceptable light losses. The solution to this problem is to introduce halogen into the bulbs. These lamps are called halogen incandescent or, in short, halogen lamps [2, 3]. Thanks to their compactness they are extremely suitable for use in small reflectors to create well-defined light beams. They are widely used for accent lighting and in car headlamps. However, because of the relatively low efficacy and short lifetime of halogen lamps, the more efficient and longer-life gas discharge and solid-state lamp alternatives have become increasingly more important in all these segments.

Halogen Cycle Principle

One way to eliminate bulb blackening is to add a small quantity of halogen (bromide or iodine) to the fill gas. Under the influence of the hot filament, evaporated tungsten particles chemically

Modified and reproduced from [1] by permission of Philips Lighting, Eindhoven, The Netherlands. © 2012 Koninklijke Philips Electronics N.V.

Halogen Lamp,**Fig. 1** The halogen cycle [1]

combine with the halogen particles. If the temperature is sufficiently high (bulb temperature at least 250°C), the resulting mixture will remain floating in a gaseous state, so preventing it from condensing on the coldest part of the lamp bulb to create bulb blackening. The extremely high temperature in the vicinity of the filament causes a chemical reaction that splits the mixture into its original components, tungsten and halogen particles, and the former return to the filament. This phenomenon is called the halogen cycle. It is illustrated in Fig. 1.

A tungsten particle that escapes from one spot of the filament does not return to exactly the same spot. As a result, the lamp will eventually burn out, because there will always be some part of the filament that will become weak over the course of time, but clearly later than in a normal incandescent lamp. This means that the filament can be heated up to a higher temperature (up to 3,000 K instead of 2,750 K as in the case of a normal incandescent lamp) while having a longer life. Halogen lamps offer two to five times the life of normal incandescent lamps, rising from 1,000 to 2,000–5,000 h. The higher filament temperature also increases both the luminous efficacy by 10–50 % relative to normal incandescent lamps and the color temperature up to 3,000 K.

Materials and Construction

Bulb

Since a halogen bulb is so small, it becomes so hot that normal glass, as used with incandescent lamps, cannot be used because it would melt. Halogen bulbs are therefore made out of quartz, which can withstand high temperatures and mechanical stress. In the manufacturing process it can be handled similarly to glass (Fig. 2).

Filament

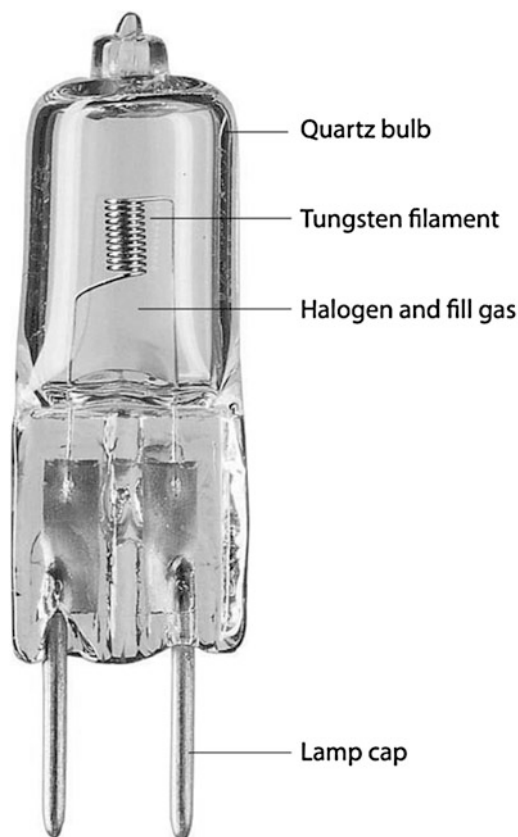
The single- or double-coiled tungsten filament can be placed axially or transversely in the halogen capsule. The placement has consequences for both the efficacy and the light distribution of the lamp.

Fill Gas

As in normal incandescent lamps, a gas filling of krypton or xenon is used to reduce filament evaporation.

Lamp Cap

Halogen lamps are available with a large variety of lamp caps and corresponding bases: two-pin caps and twist caps that ensure that the optical



Halogen Lamp, Fig. 2 Halogen lamp with its components. The bulb can be as small as approximately 10×15 mm

center of the lamp is always in the correct position, ceramic lamp caps for high-voltage, high-lumen-output lamps that become very hot, and “normal” Edison and bayonet caps for high-voltage halogen lamps that can be used just as normal incandescent lamps.

Properties

Energy Balance

Thanks to the higher working temperature of the halogen incandescent lamp, it is more efficient than a normal incandescent lamp. Some 12 % of the input power of a halogen lamp is radiated as visible light. Compare this with the 8 % of a normal incandescent lamp. The remaining power is lost as heat (through conduction, convection, and infrared radiation).

Luminous Efficacy

At 15–25 lm/W, the luminous efficacy of a halogen lamp is a factor 2–2,5 higher than that of an incandescent lamp. With the infrared-reflecting coating technology (see next section “[Product Range](#)”), the efficacy increases to some 35 lm/W. As with normal incandescent lamps, the lower the wattage, the lower the efficacy. There are halogen lamps where priority has been given to a long life at the expense of a somewhat lower efficacy (e.g., 5,000 h and 15 lm/W), and conversely, there are those where the emphasis is on efficacy and not so much on lifetime (e.g., 2,000 h and 25 lm/W).

Although the efficacy of halogen lamps is clearly higher than that of normal incandescent lamps, it is relatively low compared with that of gas discharge and solid-state light lamps.

Lumen-Package Range

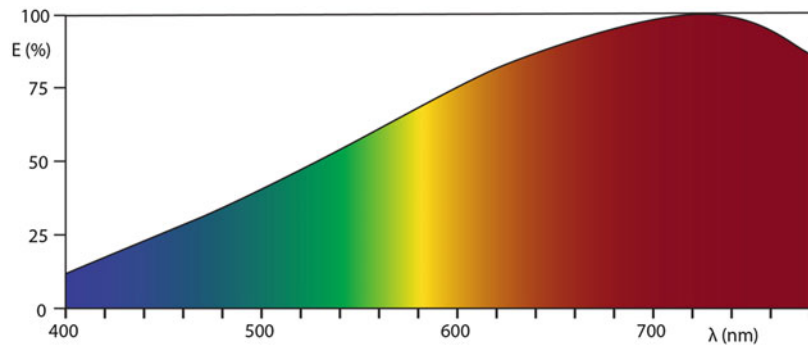
Common types of halogen lamps are available in the range from some 50 to 2,000 lm (corresponding wattage range approximately 5–100 W). Double-ended mains-voltage halogen lamps are available in versions up to some 25,000 lm (1,000 W version).

Color Characteristics

Halogen lamps have a continuous spectrum radiating more energy at long wavelengths than at short wavelengths (Fig. 3). Because of the somewhat higher temperature of the filament in halogen lamps compared to that of the filament in normal incandescent lamps, the spectral energy distribution shifts slightly toward shorter wavelengths. Depending on the version (lower efficacy/longer life or higher efficacy/shorter life), halogen lamps have a color temperature of between 2,800 and 3,000 K, respectively. Their color temperature is always slightly higher than that of normal incandescent lamps, thus giving them a somewhat cooler white light. The color rendering index is 100.

Lamp Life

Normal halogen lamps have a lifetime that is at least twice as long as that of normal incandescent lamps, thus at least 2,000 h. As explained above,

Halogen Lamp,**Fig. 3** Relative spectral energy distribution of a halogen lamp [1]

with some halogen lamps priority is given to efficacy at the expense of lifetime: viz., the 2,000 h versions. Versions where the priority is given to lifetime can have a life of up to 5,000 h.

Lamp-Lumen Depreciation

Thanks to the halogen regenerative cycle, lamp blackening is minimal. Consequently, lumen depreciation with halogen lamps is very small.

Burning Position

With certain exceptions, halogen lamps have a universal burning position. The exceptions are the high-voltage, high-wattage (750 W or more) double-ended types. Here the coiled tungsten filament is so long that with a position that is not near horizontal, the coil would sag so much that the individual coil windings would touch each other, leading to a short circuit.

Run-Up and Reignition

Like normal incandescent lamps, halogen lamps give their full light output immediately after switch on and after reignition.

Switching

The switching behavior of halogen lamps is the same as that of normal incandescent lamps.

Dimming

Halogen lamps can be dimmed in the same way as normal incandescent lamps. Just as with normal incandescent lamps, the phase-cutting system is the more efficient system and is the usually employed. Below a certain dimming point the

lamps cool down so much that the halogen cycle will no longer function. From this point on the filament starts evaporating, just as with a normal, dimmed incandescent lamp. The smaller bulb size of the halogen lamp leads to more blackening in this situation than is the case with normal incandescent lamps. However, since the filament in the dimmed situation is cooler than in the undimmed situation, dimming does not have a real negative effect on the lamp life.

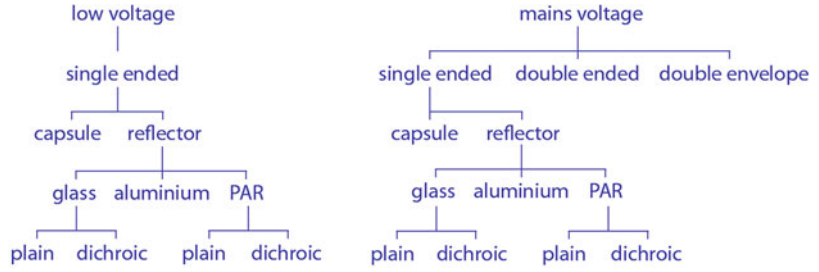
Mains-Voltage Variations

The behavior of halogen lamps as a result of an overvoltage is the same as with normal incandescent lamps. Only a few percent of overvoltage results in a drastically reduced lamp life. For example, a permanent overvoltage of 5 % reduces the lamp life by 50 %.

UV Component

Normal incandescent lamps radiate little UV. This is because the spectrum quickly falls off at the short-wavelength part of the spectrum. What little UV that remains is reduced to near zero because normal glass (the bulb material of incandescent lamps) is a very good absorber of UV radiation. Halogen lamps radiate more UV because of the higher operating temperature of the filament and because of the fact that normal quartz bulbs, unlike glass bulbs, do not absorb UV radiation. In order to limit harmful or damaging UV radiation from halogen lamps, the quartz used today for most halogen lamps is doped with UV-absorbing material (UV-blocking quartz).

Halogen Lamp,
Fig. 4 Range of halogen lamps



Halogen Lamp,
Fig. 5 Single-ended halogen reflectors lamps. From left to right: aluminum reflector with antiglare screen, glass mirror reflector, and cool beam dichroic glass reflector; under pressed-glass, high-voltage, halogen lamp (PAR)



H

Product Range

Grouping of Halogen Lamps

As far as operating voltage is concerned, halogen lamps can be placed into two groups (Fig. 4).

- Lamps operating on low voltage (6 V, 12 V, and 24 V, where the 12 V versions are most common)
- Lamps operating on mains voltage

Low-voltage halogen lamps are usually single ended. Mains-voltage versions can be either single ended or double ended, according to whether the electrical connection is at one end of the lamp (Fig. 5) or on two separated ends of the lamp (double), as illustrated in Fig. 6.

Single-ended halogen lamps are available both as “naked” bulbs (capsules such as in Fig. 7) and as reflector lamps. The reflector versions exist in three different types: mirror-coated glass reflectors, aluminum reflectors, and pressed-glass PAR reflectors (Fig. 5). The mirror-coated glass reflectors are produced with plain mirror-coating versions and with a dichroic coating (so-called cool-beam type; see later in this chapter).

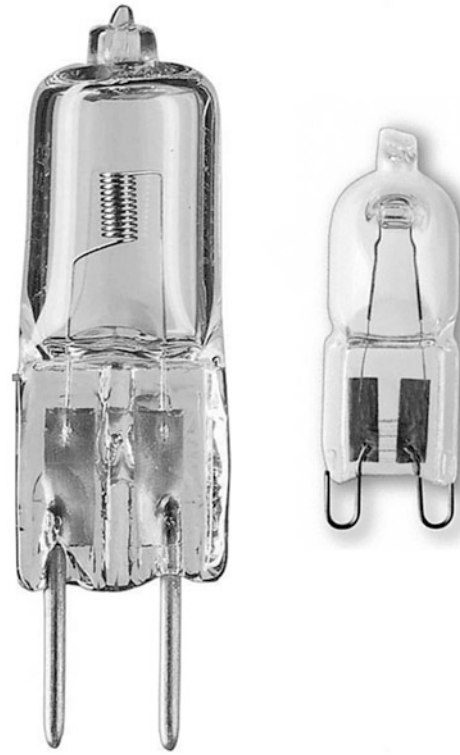
Some mains-voltage halogen bulbs (both single-ended and double-ended) are housed in an extra outer glass bulb or glass tube with the familiar Edison or bayonet lamp cap. These are called double-envelope halogen lamps (Fig. 8). They are the more-energy-efficient replacements for normal incandescent lamps.



Halogen Lamp, Fig. 6 Double-ended mains-voltage halogen lamp

Cool-Beam Halogen Lamps

A special version of halogen lamps, the so-called cool-beam lamps, makes use of the interference principle. They consist of a small halogen bulb (capsule) housed in a reflector with an interference metal coating, also known as a $\frac{1}{4} \lambda$, or dichroic coating (Fig. 9). This coating splits the radiation coming out of the halogen capsule into an infrared part (heat), which is transmitted through the coating and not reflected by it, and a visible part that is reflected but not transmitted. The practical result is that the mirror coating reflects practically all of the light while allowing the infrared to pass through it to the back of the reflector, taking about two-thirds of the heat from the light beam. Thus, only one-third of the generated heat is contained in the light beam. This is why these lamps are referred to as “cool-beam” or “dichroic” halogen lamps. When using this type of lamp in an



Halogen Lamp, Fig. 7 Low-voltage halogen capsules



Halogen Lamp, Fig. 8 Double-envelope mains-voltage halogen lamps

installation, it is important to ensure that the heat radiated from the rear of the lamp can be dissipated backward. The thickness of the coating layer is slightly different for radiation arriving at

Halogen Lamp,

Fig. 9 Dichroic interference coating (shown in yellow) in a cool-beam halogen lamp [1]

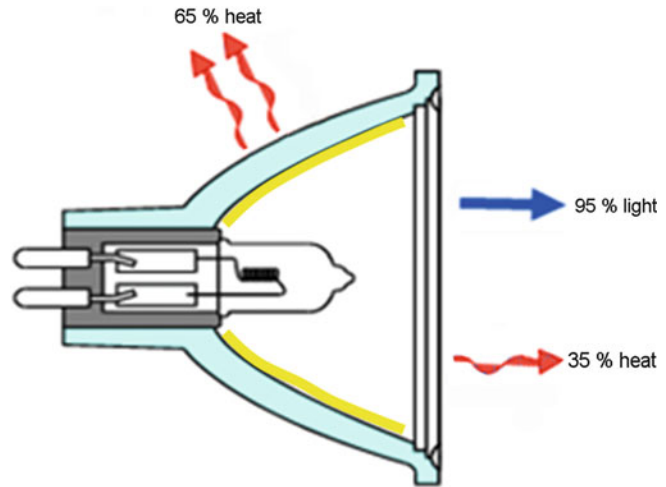
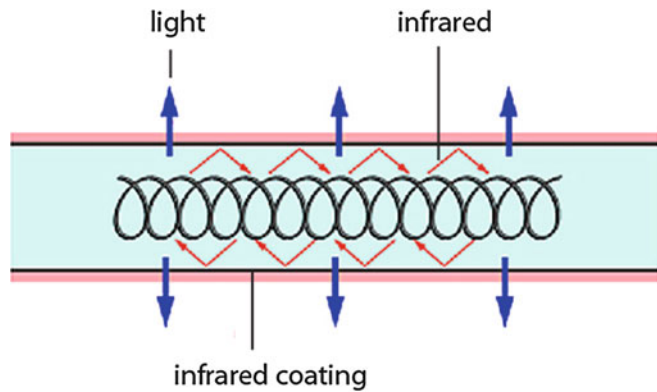
**Halogen Lamp,**

Fig. 10 Infrared reflective coating based on the interference principle used to reflect the infrared radiation of the filament back to the filament while allowing visible light to pass through [1]



the coating from different directions. This also means that the wavelength of the radiation where it is split into reflected and transmitted components is slightly different for different directions.

Halogen Lamps with Infrared-Reflective Coating

Another special version of the halogen lamp also makes use of the interference coating principle but for a completely different reason. Here the reason is thermal recovery: the interference coating applied to the halogen bulb is now so dimensioned that it reflects the infrared radiated from the filament back to the filament while allowing visible light to pass through (Fig. 10). In this way less external energy is needed to keep the filament at the required temperature, thus improving the efficacy of the lamp (by up to 25–35 lm/W).

Moreover, the heat contained in the light beam is reduced by approximately 40 %.

Cross-References

- [Incandescence](#)
- [Incandescent lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Course Book. Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn. Illuminating Engineering Society of North America, IESNA, New York (2011)

Hard-Edge Painting

- [Color Field Painting](#)

HDR

- [High Dynamic Range Imaging](#)

HDRI

- [High Dynamic Range Imaging](#)

Headlamps

- [Automotive Lighting](#)

Helmholtz, Hermann Ludwig von

Rolf G. Kuehni
Charlotte, NC, USA



Image by Ludwig Knaus, 1881

Biography

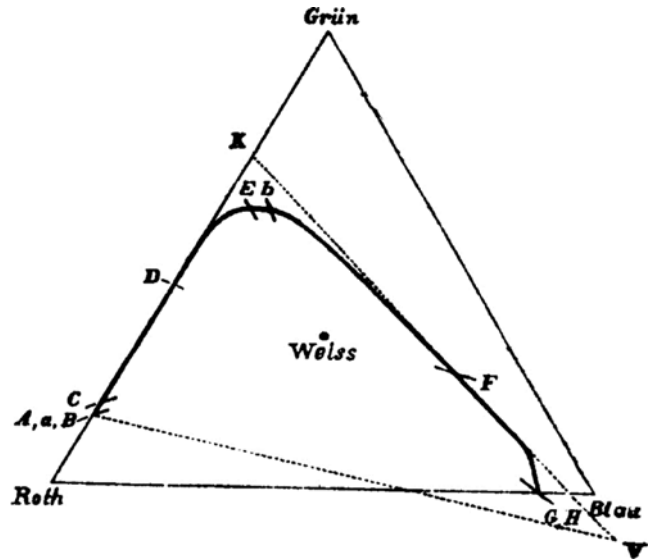
Helmholtz was born on August 31, 1821 in Potsdam near Berlin, Germany, into a well-educated family. His father, an educator, taught him the classical languages and introduced him to philosophy. Helmholtz studied medicine in Berlin under physiologist Johannes Müller, at the same time attending lectures in physics and mathematics. Müller was an adherent of the then broadly accepted philosophical nativism. Helmholtz spent much of his life finding objective arguments against nativism and in favor of empiricism. After spending several years in military service in Potsdam and a stint as associate professor of physiology at the Prussian University, he became a full professor of anatomy and physiology in Bonn. Three years later, he moved to the University of Heidelberg and in 1871 became professor of physics at the University of Berlin, where he remained until his retirement. Helmholtz had a broad range of interests, from astronomy to physics, physiology, and the relationship between the physical world and human sensory perception. In 1847 he wrote an important article on the conservation of physical force [1]. In 1850 Helmholtz invented the ophthalmoscope for inspecting the interior of eyes for medical purposes and revolutionized the practice of ophthalmology [2]. Among Helmholtz's many students were H. R. Hertz, M. Planck, A. A. Michelson, W. Wundt, and W. Kohlrausch (the codiscoverer of the Helmholtz-Kohlrausch effect).

Major Accomplishments/Contributions

In sensory perception, he specifically covered hearing and, most extensively, vision. His most important publication is *Handbuch der physiologischen Optik* (Treatise on Physiological Optics), a 900+-page work whose first edition was published in 1867 when he was in Heidelberg [3]. Its main sections are (1) anatomical description of the eye; (2) physiological optics; (3) dioptrics of the eye; (4) sensations of vision, including simple colors and compound colors; (5) intensity and duration of the sensation of light; (6) contrast (including colored shadows); and (7) duplicity theory. A second edition authored by Helmholtz

**Helmholtz, Hermann
Ludwig von,**

Fig. 1 Chromaticity diagram based on König and Dieterici's fundamental color sensitivity functions determined in Helmholtz's laboratory. Loci of spectral color stimuli are on the curved solid line. Letters indicate loci of Fraunhofer lines in the spectrum



H

was published posthumously in 1896, supervised by Helmholtz' assistant Arthur König who also added over 7800 literature references [4]. A third edition was published in 1909, consisting of the second edition, with contributions by A. Gullstrand, J. von Kries, and W. Nagel. It was translated into English by J. P. C Southhall and published with an additional contribution by the Optical Society of America [5].

Helmholtz built his trichromatic theory of color vision on Newton's findings, Young's theory of three sensor types in the eye, Maxwell's experimental findings of color mixture, and his own extensive experimental data concerning various aspects of color vision. He developed a spectral light mixture apparatus (see portrait) with which, after initial unsatisfactory results causing the mathematician H. Grassmann to challenge his data, he established the first detailed quantitative data for the relationship between color stimuli and percepts. Mathematical models of his findings and concepts were a standard procedure in his work. Under his guidance, his assistants König and Dieterici developed in 1885 a first quantitative version of a trichromatic chromaticity diagram (Fig. 1) [4].

While Helmholtz developed an experimentally supported general psychophysical theory of vision, he was also concerned about and experimented with many additional phenomena of the human visual system. He is the first to mention color

constancy, saying that "we always start out forming a judgment about the colors of bodies, eliminating the difference of illumination by which the body is revealed to us" [6]. The color vision theory of Helmholtz, based on Young's conjecture of three sensor types, is known as the Young-Helmholtz theory of trichromatic color vision.

His rival Ewald Hering developed a theory of color vision essentially based on psychology to which he gave a presumed physiological basis. Both views attracted followers [7]. Today, the controversy has been narrowed but still remains. Interestingly, their expressed epistemological views are not as might be expected: Helmholtz viewed lights and colors as symbols. In 1852 he said: "Light and color percepts are only symbols for the relations of reality; with the latter they have as little and as much similarity or relationship as the name of a person or its written form with the person itself" [8]. Hering, on the other hand, believed that, in perception, our access to real objects is a direct one.

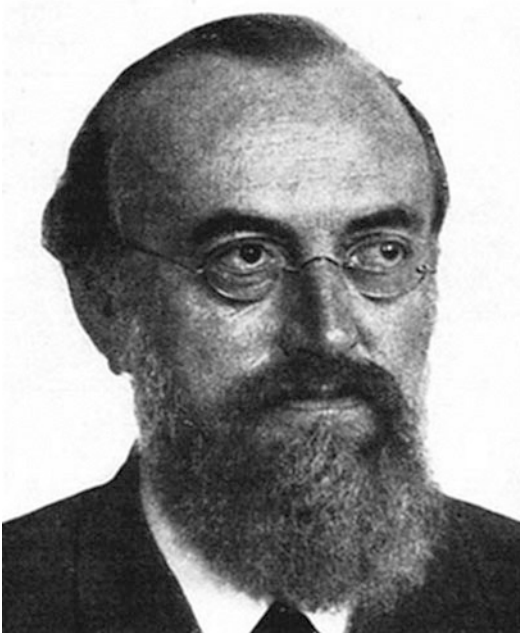
References

1. von Helmholtz, H.: Über die Erhaltung der Kraft. Reimer, Berlin (1847)
2. von Helmholtz, H.: Beschreibung eines Augenspiegels zur Untersuchung der Netzhaut im lebenden Auge. Förstner, Berlin (1851)

3. von Helmholtz, H.: Handbuch der physiologischen Optik. Voss, Leipzig (1867)
4. Von Helmholtz, H.: Handbuch der physiologischen Optik, 2nd edn., edited and with contributions of A. P. König. Voss, Hamburg (1896)
5. von Helmholtz, H.: Handbuch der physiologischen Optik, 3 edn., with contributions by A. Gullstrand, J. von Kries, W. Nagel. Voss, Hamburg (1909). English translation with an additional contribution by C. Ladd-Franklin published by the Optical Society of America as Helmholtz's treatise on physiological optics in 1924.
6. von Helmholtz, H.: Handbuch der physiologischen Optik, 3 edn., with contributions by A. Gullstrand, J. von Kries, W. Nagel. Voss, Hamburg (1909). English translation with an additional contribution by C. Ladd-Franklin published by the Optical Society of America as Helmholtz's treatise on physiological optics in 1924. English translation, part 2, p. 286/7
7. Turner, R.S.: In the Eye's Mind: Vision and the Helmholtz-Hering Controversy. Princeton University Press, Princeton (1994)
8. von Helmholtz, H.: Über die Natur der menschlichen Sinnesempfindungen. In: Wissenschaftliche Abhandlungen, vol. 2. Barth, Leipzig (1883)

Hering, Karl Ewald Konstantin

Rolf G. Kuehni
Charlotte, NC, USA

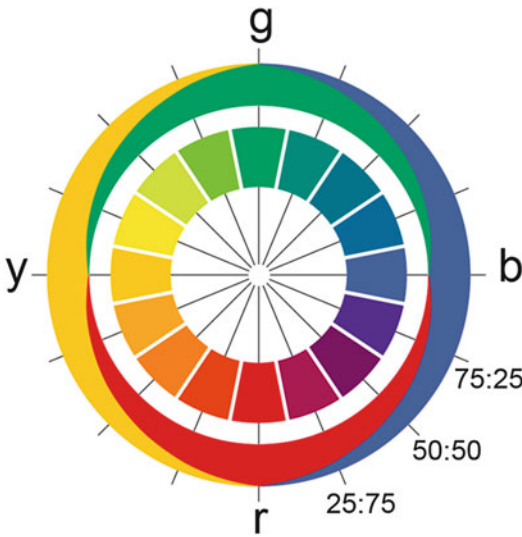


Biography

Karl Ewald Konstantin Hering was born on August 5, 1834, in Alt-Gersdorf, Saxony, son of a Lutheran pastor and his wife. He studied medicine at the University of Leipzig, obtaining an MD in 1860. For the next five years, he practiced medicine in Leipzig and pursued personal interests in vision on the side, publishing five *Beiträge zur Physiologie (Contributions to Physiology)* between 1861 and 1864 [1]. He was married in 1863, and he and his wife had a son, Heinrich Ewald, born in 1863. In 1865 he was appointed professor of physiology at the Josephinum Academy in Vienna. In 1870 he became the successor of J. E. Purkinje as professor of physiology at the University of Prague where he remained for 25 years, studying among other things the electrical actions of nerves and muscles, as well as the perception of light and color vision. In 1895 he was invited to join the University of Leipzig where he remained until his retirement in 1915. During his lifetime Hering assumed the role of anti-Helmholtz, scientifically and philosophically battling with him in regard to several subject matters [see, e.g., 2]. Hering died on January 26, 1918, in Leipzig.

Major Accomplishments/Contributions

Hering developed a theory, alternative in detail to that of Helmholtz, in regard to spatial perception based on images in two eyes. Publication of and the response to Helmholtz's first edition of the *Handbuch der Physiologischen Optik* in 1867 [3] provided a basis for Hering to consider in detail the issues of color perception and develop his own different theory. The result was the presentation in 1874 of six extended contributions to the Imperial Academy in Vienna, published in book format in 1878 as *Zur Lehre vom Lichtsinne (On the Theory of the Light Sense)* [4]. Hering considered the Young-Helmholtz theory to give too much weight to physics and not enough to the perceptual aspects of color. In response he developed an opponent-color theory postulating three



Hering, Karl Ewald Konstantin, Fig. 1 Superimposed images of the conceptual mixture of the four hue fundamentals in different ratios and the resulting hue circle

opponent pairs of fundamental colors (*Urfarben*), yellow-blue, red-green, and white-black, building on earlier ideas by H. Aubert and E. Mach. The hues are arranged according to simple principles in a hue circle (Fig. 1) (derived from two images in Ref. [4]. Image Tilo Hauke), the complete arrangement of all colors of a given hue being in a triangle, with white, black, and full color at the corners and “veiled” colors filling the interior. Hering called the result, in form of a double cone, the “natural color system.” He also proposed dissimilation/assimilation processes in the eye/brain as physiological basis of the opponent system, in opposition to the Young-Helmholtz theory involving three fundamentals. A commercial version of the Natural Color System atlas was introduced in 1979 by the Scandinavian Colour Institute [5].

Today, the perceptual aspects of Hering's system continue to be considered essentially valid, with its physiological part having fallen short of reality. Despite many attempts, a psychophysical model of the four hues presumed fundamental and their mixtures that also meet other components of the colorimetric system is still lacking at this time, a major issue being the fact that while mean

unique yellow and blue stimuli are essentially complementary, unique green and unique red stimuli are far from it, complicating any model with the purpose of representing a perceptually meaningful and at the same time colorimetrically valid model.

References

1. Hering, E.: Beiträge zur Physiologie. Engelmann, Leipzig (1861)
2. Turner, R.S.: In the Eye's Mind, Vision and the Helmholtz-Hering Controversy. Princeton University Press, Princeton (1994)
3. von Helmholtz, H.: Handbuch der physiologischen Optik. Voss, Leipzig (1867)
4. Hering, E.: Zur Lehre vom Lichtsinne. Gerold, Wien (1878). Translation: Outlines of a theory of the light sense. Hurvich, L. M., Jameson, D., trans. Harvard University Press, Cambridge, MA (1964)
5. Scandinavian Colour Institute: Swedish Natural Colour System (NCS)

Hertzsprung-Russell Diagram

Michael H. Brill

Datacolor, Lawrenceville, NJ, USA

Definition

A scatter graph of stars showing the relationship between each star's absolute magnitude and its estimated surface temperature, or between optical and perceptual proxies for these quantities.

Absolute Magnitude and Temperature Scales

The absolute magnitude and temperature can be described as follows:

A star's absolute magnitude is the attenuation (in factors $10^{-0.4}$) of the star Vega's power (as received at 10 pc viewing distance [32.6 light-years]) to equal that of the star (also

corrected to 10 pc). The convention that dimmer stars have higher magnitude is a historical precedent that dates from Hipparchus (c. 190 BCE–c. 120 BCE), whose system of stellar magnitudes was based on visual assessment.

A star's surface temperature is estimated in one of three ways: by the observed color (an old way), by a comparison of two sensor outputs such as blue and violet (a newer way), or by a model prediction of the temperature of a black-body radiator with the same radiation power per unit star-surface area (the most modern way). The third way requires independent inference of the star's radius but assumes the star has an emissivity of 1. To acknowledge the lack of compensation for the true emissivity, the temperature on an H-R diagram is called "effective temperature."

Coordinate Conventions

Here are the coordinate conventions of the H-R diagram: Temperature increases from right to left, and magnitude (i.e., dimness) increases from bottom to top. Hence dim red stars (red giants) appear near the upper right of the diagram, and bluish bright stars (such as white dwarfs) appear in the lower left. A long cluster of stars called the main sequence extends from the upper left to the lower right. Higher mass stars occur at the upper left of this sequence, and the Sun appears approximately in the middle. The main sequence is composed of stars that are currently dominated by hydrogen that is fusing into helium. According to currently accepted theory of stellar evolution, such stars will eventually migrate either to the red giants or white dwarf domain of the H-R diagram.

H-R diagrams are often depicted in color, either pseudocolor with a thermal code to show the temperature or coded according to star categories such as cluster membership.

History

The H-R diagram was originated by Danish astronomer Ejnar Hertzsprung and American astronomer Henry Norris Russell ([https://en.](https://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram)

[wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram](https://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram)). Originally the diagram was based on visual estimation of magnitude and color, and it was a research tool to help characterize stellar evolution before the mechanism of nuclear fusion was understood. After about the 1930s, the H-R diagram became based on objective measurements but was used less as a research tool and more as a way to illustrate the theoretically predicted evolution of stars through trajectories in the diagram.

Cross-References

- [Color-Magnitude Diagrams](#)
- [Galaxy Color Magnitude Diagram](#)

References

1. Gribbin, J.: *Companion to the Cosmos*. Little, Brown, NY (1996)
2. Friedman H.: *The Astronomer's Universe: Stars, Galaxies, and Cosmos*. W. W. Norton, NY (1990)
3. Moore, P.: *Atlas of the Universe*. Cambridge U. Press, NY (1998)
4. Pasachoff, J. M.: *A Brief View of Astronomy*. Saunders College Publ., NY (1986)
5. Ronin C. A.: *The Natural History of the Universe: From the Big Bang to the End of Time*. MacMillan, NY (1991)
6. Shapley, A. E.: *Physical Properties of Galaxies from $z = 2 - 4$* . <http://arxiv.org/abs/1107.5060v2>, accessed 26 Feb. 2013
7. Spence, P. (ed.): *The Universe Revealed*. Cambridge U. Press, NY (1998)

High- and Low-Pressure Sodium Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Sodium gas discharge lamps](#)

Modified and reproduced from [1] by permission of Philips Lighting, Eindhoven, the Netherlands. © 2012 Koninklijke Philips Electronics N.V.

Definition

Lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a sodium vapor that is contained in a transparent bulb.

Sodium Gas Discharge Lamps

Sodium gas discharge lamps can be distinguished in two main groups: low-pressure sodium lamps in which the gas pressure is very low (0.7 Pa or 7.10^{-6} atm) and high-pressure sodium lamps in which the gas pressure is a factor 10,000–100,000 times higher (10–100 kPa); [2–4]. Because their construction and performance is in many ways different, they are dealt with in two different sections of this entry.

Low-Pressure Sodium Lamps

Low-pressure sodium lamps belong to the group of high-intensity discharge (HID) lamps, because they are available in high-light output (and thus high-luminous-intensity) versions. All low-pressure gas discharge lamps have in common the fact that they are long. Low-pressure sodium lamps are highly efficient lamps with a good lifetime but no color rendition at all. Their application is therefore restricted to those situations where color rendering is of no importance, as, for example, on motorways, on railway-marshaling yards, and in some security-lighting

situations. Low-pressure sodium lamps are sometimes also referred to as LPS lamps.

Working Principle

The gas discharge principle of low-pressure sodium lamps is similar to that of low-pressure mercury lamps (see that chapter for general details). In low-pressure sodium lamps, the discharge takes place in vaporized sodium. The low-pressure sodium discharge emits monochromatic radiation in the visible range. Therefore, unlike low-pressure mercury lamps, they do not need fluorescent powders to convert the wavelength of the radiation. The monochromatic (single wavelength) radiation is the reason that color rendering is nonexistent. The wavelength of the monochromatic radiation is 589 nm (yellowish light), which is very close to the wavelength for which the eye has its maximum sensitivity (Fig. 1). It is mainly for this reason that the lamp has such an extremely high-luminous efficacy (up to 190 lm/W system efficacy). Like all gas discharge lamps (with very few exceptions), a low-pressure sodium lamp cannot be operated without a ballast to limit the current flowing through it.

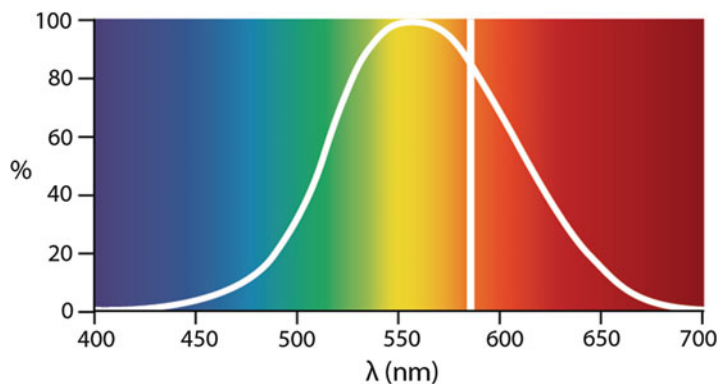
Materials and Lamp Construction

The main parts of a low-pressure sodium lamp are (Fig. 2):

- Discharge tube
- Fill gas
- Electrodes
- Outer bulb with inner coating
- Lamp cap

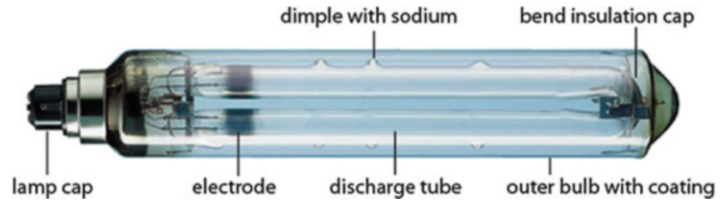
High- and Low-Pressure Sodium Lamp,

Fig. 1 Relative eye-sensitivity curve and the monochromatic line of 589 nm of the low-pressure sodium spectrum



High- and Low-Pressure Sodium Lamp,

Fig. 2 Principle parts of a low-pressure sodium-discharge lamp [1]



Discharge Tube

Just as with the low-pressure mercury, fluorescent tube, and lamp, the power dissipated in the low-pressure sodium lamp largely determines the length of the discharge tube. Especially for the higher wattages (and thus lumen packages), the unfolded length has to be really long. To reduce the actual length of the lamp, the discharge tube of low-pressure sodium lamps is therefore always U shaped. Nevertheless, the highest lumen packages still require a lamp length of about 1.2 m. The U-shaped discharge tube is made of sodium-resistant glass and contains a number of small dimples, or hollows, where the sodium is deposited as a liquid during manufacture. After ignition, the discharge first takes place through the inert gas mixture. As the temperature in the tube gradually increases, some of the sodium in the dimples vaporizes and takes over the discharge, which then emits the monochromatic radiation. At switch-off, the sodium condenses and again collects at the dimples, these being the coldest spots in the tube. Without the dimples the sodium would, after some switch-on switch-off cycles, gradually condense along the whole inner tube wall, decreasing light transmission considerably.

Fill Gas

The inert gas mixture of neon and argon, called the “Penning mixture,” acts as a starting gas and buffer gas (to protect the electrodes). During start-up, the discharge only takes place in this gas, which is why a low-pressure sodium lamp radiates deep-red light for some 10 min during start-up (Fig. 3).

Electrodes

Most modern low-pressure sodium lamps have cold-start electrodes. These consist of a triple-coiled tungsten wire, so that they can hold a large quantity of emitter material.



High- and Low-Pressure Sodium Lamp, Fig. 3 Immediately after switch-on, a low-pressure sodium lamp emits just a little reddish light, which gradually changes to the familiar yellow sodium light when the lamp has fully warmed up [1]

Outer Bulb

The optimum sodium-vapor pressure is reached when the temperature of the wall of the discharge tube is maintained at 260 °C. To reach this temperature efficiently, the U-shaped discharge tube is contained in an evacuated outer-glass tube. To further increase the thermal insulation, this outer-glass tube is coated on its inner surface with an interference layer that reflects infrared radiation but transmits visible radiation. In this way, most of the heat radiation is reflected back into the discharge tube, so maintaining the tube at the desired temperature, while visible radiation is transmitted



High- and Low-Pressure Sodium Lamp, Fig. 4 Immediately after switch-on, a low-pressure sodium lamp emits just a little reddish light, which gradually changes to the familiar yellow sodium light when the lamp has fully warmed up

through the layer. Early low-pressure sodium lamps did not have an integrated outer bulb with infrared reflecting coating. They used a separate double-wall evacuated tube (“thermos bottle”) which was reused after failure of the gas discharge tube at the end of its life (Fig. 4).

Apart from the dimples, the bend of the U-shaped discharge tube also forms a cold spot where sodium can condense and accumulate. The bend is therefore insulated by a heat-reflecting metal cap.

Lamp Cap

All low-pressure sodium lamps are provided with a bayonet-type lamp cap. This allows the discharge tube to be accurately positioned. This is critical, because the light distribution of a low-pressure sodium luminaire is dependent on the position of the U-shaped discharge tube.

Properties

Energy Balance

A low-pressure sodium lamp emits approximately 40 % of the input power in the form of visible radiation. This is the highest percentage of all gas discharge lamps. The remaining part of the input power is lost in the form of heat.

System Luminous Efficacy

The luminous efficacy of the system is strongly dependent on the wattage of the lamp and ranges from 70 to 190 lm/W, for low and high wattages, respectively.

Lumen-Package Range

Low-pressure sodium lamps are available in the range from approximately 2,000 to 30,000 lm (corresponding wattage range, 18–180 W).

Color Characteristics

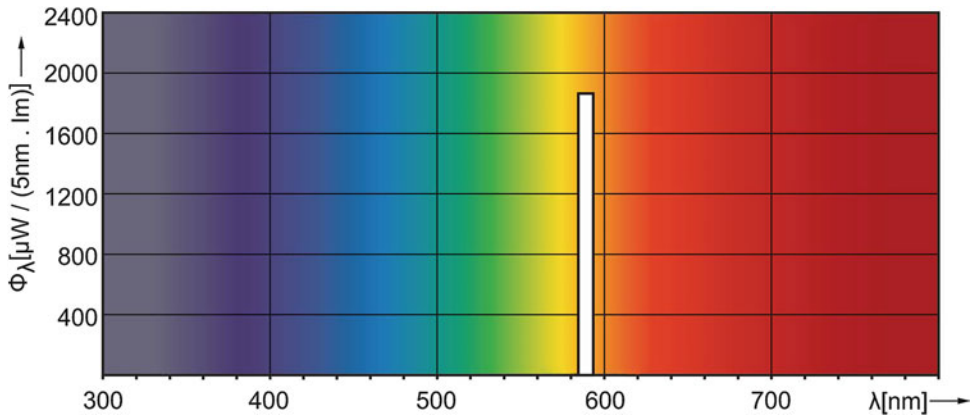
As mentioned before, low-pressure sodium lamps emit monochromatic light in the yellowish part of the spectrum (Fig. 5). Color rendering is therefore nonexistent ($R_a = 0$). The correlated color temperature is around 1,700 K.

Lamp Life

Apart from the normal cause of failure in gas discharge lamps (viz., electrode emitter exhaustion), low-pressure sodium lamps may also fail because of cracks or leaks in the long discharge tube or outer bulb. This may especially be the case in environments where there are strong vibrations as, for example, may occur in poorly designed road-lighting luminaires during strong winds. Leakage in the outer bulb disrupts the thermal isolation, which in turn means that not enough sodium will vaporize. As a consequence, the discharge will take place in the starting-gas mixture, so emitting only the corresponding deep-red light. Economic lamp life is around 12,000 h (based on a 20 % mortality rate). Some versions make use of special getter material to maintain a high vacuum, resulting in fewer failures during the economic lifetime of the lamp: the economic lamp lifetime is about 15,000 h (again, mortality rate of 20 %).

Lamp-Lumen Depreciation

Lumen depreciation occurs through blackening of the discharge tube by scattering of the emitter material of the electrodes and by discoloration of the glass caused by the sodium. Depending on the type of control gear used, these effects are partially counteracted by a slow and gradual increase in the power dissipated in the lamp.



High- and Low-Pressure Sodium Lamp, Fig. 5 Spectral energy distribution of low-pressure sodium lamps [1]

Burning Position

Electrodes and lead-in wires coming into contact with condensed sodium can eventually suffer damage. To prevent this from happening, low-pressure sodium lamps have restrictions as to their burning position. Base-down burning positions in particular have to be avoided. The restrictions are less critical for lower-wattage (viz., smaller) lamps, simply because they contain less sodium.

Run-Up and Reignition

As has already been explained, the sodium needs time to vaporize while the discharge takes place in the starting-gas mixture. The warming-up process takes about 10 min. Nearly all low-pressure sodium lamps reignite immediately. The exceptions are the highest wattage lamps (131 and 180 W), which restrike after 10 min.

Dimming

Low-pressure sodium lamps cannot be dimmed. Dimming would decrease the lamp temperature so that not enough sodium would remain in the vapor state to maintain the sodium discharge.

Ambient-Temperature Sensitivity

The good thermal insulation afforded by the outer bulb ensures that lamp performance is almost independent of ambient temperature. Also, thanks to the starting-gas mixture, starting too is almost independent of ambient temperature.

Mains-Voltage Variations

The variations in lamp current and lamp voltage as a consequence of a change in mains supply voltage tend to cancel each other out, the net result being that the lamp wattage, and, to a certain extent, the luminous flux remain practically constant over a wide range.

Product Range

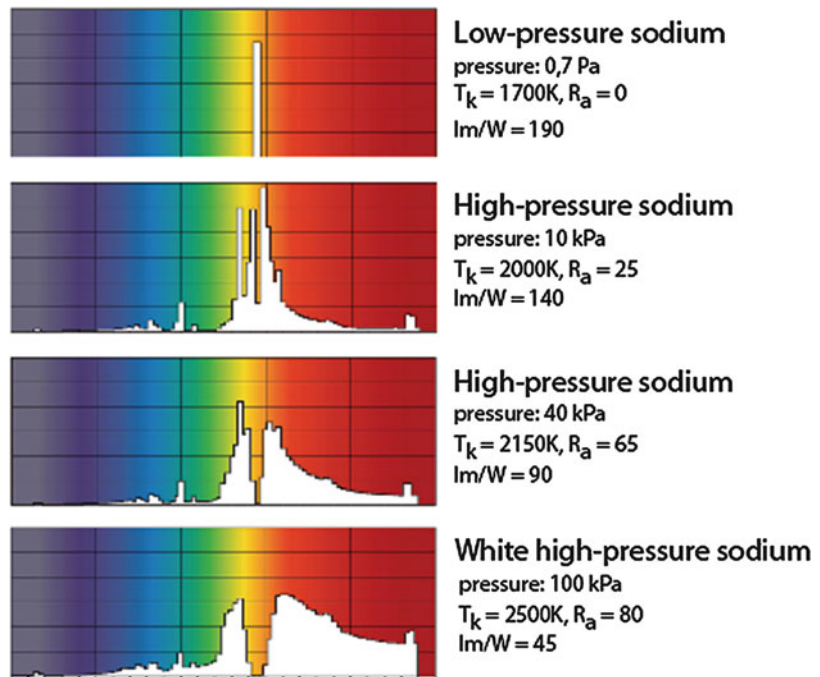
Low-pressure sodium lamps are available in two versions, each with a different balance between efficacy and lumen output. There is a high-lumen-output version with a 10–15 % lower efficacy and a high-efficacy version with a 20 % lower lumen output, respectively. The high-efficacy lamps can be operated on HF electronic control gear, which further increases their efficacy by between 15 % and 35%. The length of low-pressure sodium lamp increases considerably with wattage: the 18 W version (the lowest wattage available) has a length of 22 cm, while the longest lamps (131 and 180 W versions) have a length of 112 cm.

High-Pressure Sodium Lamps

High-pressure sodium gas discharge lamps belong as well as low-pressure sodium lamps to the group of HID lamps, because they are available in high-light output (and thus high-luminous-intensity) versions. High-pressure sodium lamps

High- and Low-Pressure Sodium Lamp,

Fig. 6 Effect of sodium-vapor pressure on the spectral power distribution of different sodium gas discharge lamps [1]



H

are also referred to as HPS lamps. High-pressure sodium lamps, in common with all high-pressure discharge lamps, are relatively compact. By increasing the vapor pressure in a sodium lamp, the color rendering improves and the color appearance changes from yellow to yellow-white, albeit at the cost of a decrease in efficacy. However, the resulting efficacy is more than double that of a high-pressure mercury lamp. At its introduction in the late 1960s, a very efficient alternative was thus obtained for the many high-pressure mercury lamps employed at that time in road lighting. Today, road-lighting installations all over the world very often use high-pressure sodium lamps, although for some installations LED solutions have become an alternative.

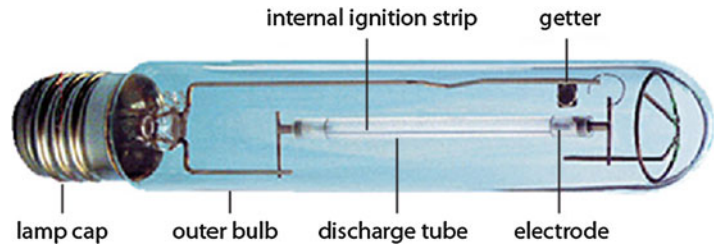
By further increasing the sodium pressure, the color quality of the light improves to such an extent that the light becomes real white. These so-called white high-pressure sodium lamps have a lower efficacy but sometimes offer an acceptable alternative to halogen and compact metal halide lamps for accent lighting.

Working Principle

It has been shown that with low-pressure sodium lamps at the low working pressure of that discharge, a single, monochromatic line of light at a wavelength of 589 nm is emitted. With increasing pressure, the radiation in the core of the discharge is absorbed by the cooler surrounding gas and reemitted in the form of radiation not of the 589 nm line but with wavelengths slightly smaller and slightly larger than 589 nm. So, the 589 nm line gradually disappears (called self-absorption), while in the wavelength area to the left and right of that value, more and more light is emitted (broadening of the spectrum). The phenomenon of self-absorption and spectrum broadening is illustrated in Fig. 6, which shows examples of sodium lamps with different operating vapor pressures. The phenomenon is accompanied by a loss of efficacy. At the operating pressure of a normal high-pressure sodium lamp, the lamp has a yellow-white color appearance (2,000 K) and a moderate color-rendering index R_a of approximately 25 at an efficacy, for the higher wattages, of some 140 lm/W. With further increase in operating pressure, the same process continues, viz.,

High- and Low-Pressure Sodium Lamp,

Fig. 7 Main parts of a high-pressure sodium lamp (example, tubular version) [1]



widening of the spectrum at the cost of efficacy. With lamps that operate at a four-times-higher pressure the color rendering improves to “fairly good” ($R_a = 65$) at an efficacy of around 90 lm/W. A version with an operating pressure ten times higher than that of the standard high-pressure sodium lamp is also produced: white HPS. This version has an R_a of 80 and radiates white light with a color temperature of 2,500 K at an efficacy of around 45 lm/W.

Materials and Construction

The main parts of a high-pressure sodium lamp are (Fig. 7):

- Discharge tube
- Fill gas
- Outer bulb
- Electrodes
- Lamp caps
- Getter

Discharge Tube

Sodium at high temperatures has the tendency to migrate slowly through quartz. Quartz can therefore not be used as material for discharge tubes of high-pressure sodium lamps. Discharge tubes made of ceramic material withstand sodium at very high operating temperatures. This is why, right from their introduction in 1966, high-pressure sodium lamps have translucent, tubular-shaped, ceramic discharge tubes. Since the ceramic tube is produced by sintering minuscule aluminum-oxide particles, one cannot look straight through the wall, although the light transmittance is more than 90 % (translucent instead of transparent material).

Fill Gas

Sodium is introduced into the gas discharge tube as a sodium-mercury amalgam composition, which partially vaporizes when the lamp reaches its operating temperature. The sodium vapor is responsible for excitation and subsequent light radiation, while the mercury gas acts to regulate the voltage of the lamp and reduces thermal losses (buffer gas). The starter gas normally added is xenon. An exception to this is found in the case of low-wattage lamps, which have a built-in igniter in the form of a bimetal switch and a neon-argon mixture – the Penning mixture also used in fluorescent lamps – as starting gas. By increasing the pressure of the starting-gas xenon, the luminous efficacy of the lamp increases by about 20 %. However, to ensure proper ignition at this higher starting-gas pressure, an auxiliary ignition wire has to be added very close to the discharge tube. Some versions of high-pressure sodium lamps therefore have an internal ignition strip (see Fig. 7). From an environmental point of view, it can be desirable to have high-pressure sodium lamps that do not contain mercury. Such lamps are, in fact, available. In these mercury-free high-pressure sodium lamps, xenon is used not only as a starting gas but also as the buffer gas that regulates the voltage and reduces thermal losses.

Electrodes

The electrodes employed in high-pressure sodium lamps are basically the same as those found in high-pressure mercury lamps. They consist of a core of tungsten rod with a tungsten coil (impregnated with emissive material) wound around it.

High- and Low-Pressure Sodium Lamp,

Fig. 8 Different outer bulbs and lamp caps of high-pressure sodium lamps



Outer Bulb

To thermally insulate the gas discharge tube and to protect its components from oxidation, an outer bulb is employed. The outer bulb of standard high-pressure sodium lamps is either ovoid or tubular (T version) in shape (Fig. 8). The internal wall of the ovoid bulb is usually coated with a diffusing powder. The coated versions were introduced so as to obtain the same light-emitting area as in normal ovoid, fluorescent-powder-coated, high-pressure mercury lamps. In this way the coated ovoid high-pressure sodium lamps can be used with the same luminaire optics as those developed for high-pressure mercury lamps. This was especially important at the original introduction of high-pressure sodium lamps, when they were replacing many existing high-pressure mercury lamps that were then being used in many road-lighting installations. Note that the coating in high-pressure sodium lamps is of the diffusing, nonfluorescent type. Since high-pressure sodium lamps produce practically no UV radiation, there is no point in using fluorescent powder. White high-pressure sodium lamps are usually only available in the tubular form. The glass used for the outer bulb of high-pressure sodium lamps for wattages of more than 100 W is hard glass.

Lamp Caps

The lamp caps employed for normal high-pressure sodium lamps are of the Edison screw type. The white high-pressure sodium lamps have a special bi-pin cap to ensure exact positioning in a luminaire.

Getter

During the process in which the lamp is evacuated, it is impossible to remove all traces of air and water vapor. During the operation of the lamp, minuscule particles evaporate from the glass and metals in the tube. All these traces would lead to an unacceptably short life. To remove them a getter is added that absorbs these traces. The getter is usually in the form of a small piece of solid material (see Fig. 7).

Properties

Energy Balance

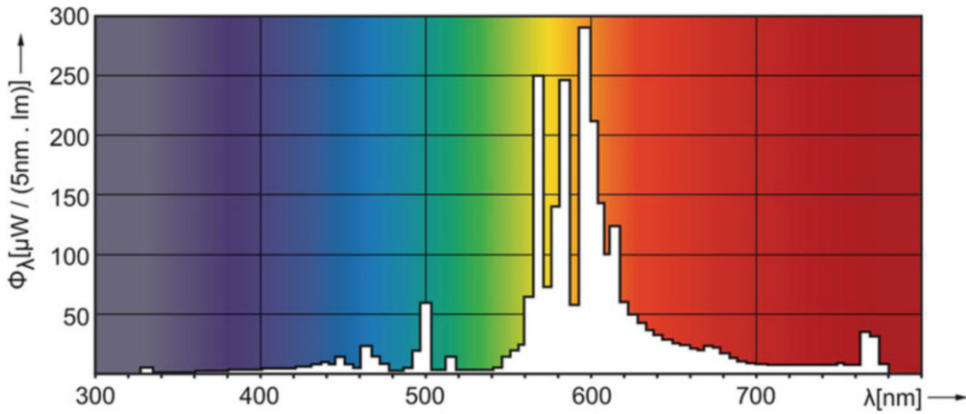
Some 30 % of the input power of a middle-range type of high-pressure sodium lamp is emitted in the form of visible radiation. Compare this with the 40 % of low-pressure sodium lamps and the 17 % of high-pressure mercury lamps.

System Luminous Efficacy

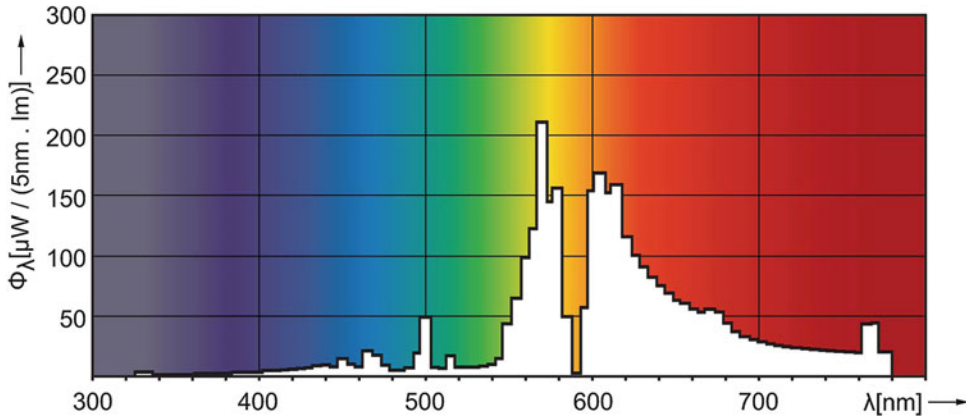
The efficacy of the compact white high-pressure sodium lamp varies between approximately 30 and 45 lm/W. The high-pressure sodium lamp with color-rendering index of 60 has an efficacy between 75 and 90 lm/W, and the normal high-pressure sodium lamps (with R_a around 25) an efficacy of between some 80 and 140 lm/W. The higher the wattage, the higher the efficacy.

Lumen-Package Range

Compact white high-pressure sodium lamps are available in lumen packages from some 1,500 to 5,000 lm (corresponding wattage range of 35–100 W). Normal high-pressure sodium lamps



High- and Low-Pressure Sodium Lamp, Fig. 9 Spectral energy distribution of a normal high-pressure sodium lamp: T_k 2,000 K and R_a 25 [1]



High- and Low-Pressure Sodium Lamp, Fig. 10 Spectral energy distribution of a color improved high-pressure sodium lamp: T_k 2,150 K and R_a 65 [1]

are being produced in the approximate range of 4,000–150,000 lm (corresponding wattage range of 50–1,000 W).

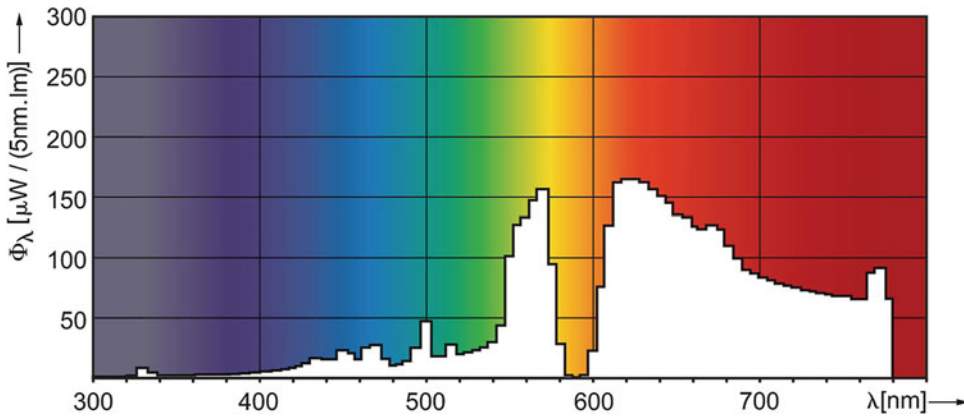
Color Characteristics

As with all gas discharge lamps, the high-pressure sodium lamp spectrum is discontinuous. Figures 9, 10, and 11 show the spectra of a normal high-pressure sodium lamp, a color improved lamp, and a white lamp. The color temperature of these versions ranges from 2,000 to 2,500 K and the color-rendering index from 25 to 80. Since the spectrum of all versions is relatively strong in the red wavelength area, the rendition of human

faces is often experienced as being somewhat flattering. Of course, for indoor lighting the color rendering of the normal high-pressure sodium lamp is far from adequate. For road lighting it is experienced as being acceptable.

Lamp Life

The lamp voltage of a high-pressure sodium lamp increases gradually with life. The chief cause of lamp failure is that the lamp voltage rises higher than the voltage output of the ballast, causing the lamp to extinguish. When this happens, the lamp cools down and the pressure in the lamp decreases so that the igniter can ignite the lamp again. After



High- and Low-Pressure Sodium Lamp, Fig. 11 Spectral energy distribution of a white high-pressure sodium lamp: T_k 2,500 K and R_a 80 [1]

H

some minutes, the lamp voltage again increases too much and the lamp extinguishes again. So, the normal end of life of a high-pressure sodium lamp is accompanied by this so-called cycling effect. For those situations where this on-off cycling might be disturbing or would damage the ballast, special “self-stopping” gear is available that stops igniting the lamp once the high voltage at end of life is reached.

Normal high-pressure sodium lamps have an economic life of up to some 20,000 h (20 % mortality). The lifetime of white high-pressure sodium lamps is not determined by the moment of actual failure of the lamps but by the onset of too large a color shift of the light, which is caused by the gradual increase of lamp voltage. To greatly increase their lifetime, white high-pressure sodium lamps therefore employ an electronic voltage stabilizer integrated into their control gear. The economic lifetime of compact white high-pressure sodium lamps, depending on type, lies between some 8,000 and 12,000 h (based on a 20 % too-large color shift).

Lamp-Lumen Depreciation

For the compact white lamps, lumen-depreciation values vary between some 20 % and 25 % (after 10,000 h). The normal high-pressure sodium lamps have a much smaller lumen depreciation of between approximately 5 % and 10 % (after 20,000 h).

Run-Up and Reignition

The high-pressure sodium lamp must be ignited by a high-voltage pulse, typically 1.8–5 kV. After ignition, the color of the light is initially white (discharge in the starting gas), changing to yellowish after some 20 s as the sodium amalgam gradually vaporizes and the vapor pressure rises; until after some 3–5 min, the nominal pressure and full light output are reached. Reignition of the hot lamp requires the lamp to cool down for about 1 min to allow the pressure to decrease to a point where the ignition pulse can again ionize the sodium atoms.

Dimming

All high-pressure sodium lamps can be dimmed to a certain extent, depending on the type of dimming equipment used. Lower wattages (100–150 W) can be dimmed with special electronic gear, which allows for dimming to 20 %. Higher lamp wattages can be dimmed by including an extra inductive coil (ballast) in the ballast circuit. Lamp color remains virtually constant and lifetime is not affected.

Ambient-Temperature Sensitivity

The behavior of high-pressure sodium lamps during temperature variations differs from that of other discharge lamps because of the excess of amalgam used in the lamp. Although the outer bulb offers some degree of thermal isolation, the

lamp manufacturer's specifications should be followed in the design of luminaires as far as its effect on lamp temperature is concerned.

Mains-Voltage Variation

A 5 % mains-voltage variation has a 15 % effect on the light output of high-pressure sodium lamps. The same 5 % mains-voltage variation has a 5 % effect on lamp voltage.

Product Range

As shown, high-pressure sodium lamps are available in three basic types: standard high-pressure sodium, an improved color version of high-pressure sodium, and a compact, white-light, high-pressure sodium lamp. The standard high-pressure sodium lamp is available in two forms: the standard one and an extra-high-efficacy version (with internal ignition strip). Most high-pressure sodium lamps contain a very small quantity of mercury, but there is also a version that is completely free of mercury.

For ease of replacement of high-pressure mercury lamps in existing installations with more efficient high-pressure sodium types, a special type of high-pressure sodium lamp has been developed. This type uses a neon-argon (Penning) mixture as starting gas and is fitted with an ignition coil surrounding the discharge tube. These features allow the lamp to be operated on a standard high-pressure mercury-lamp ballast.

Cross-References

- [High-Pressure Mercury Lamp](#)
- [Metal Halide Lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Course book, Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: *Lamps and Lighting*, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *Illuminating Engineering Society of North America*, IESNA, New York, 10th edn (2011)
4. De Groot, J., Van Vliet, J.: *The High-Pressure Sodium Lamp*. Kluwer Technische Boeken, Deventer/Antwerpen (1986)

High Dynamic Range Imaging

Tania Pouli

Technicolor, Cesson-Sévigné, France

Synonyms

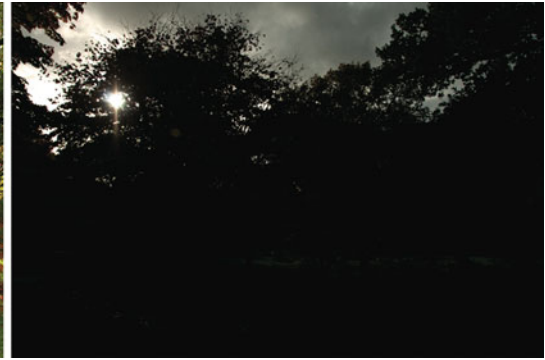
[HDR](#); [HDRI](#); [Wide-dynamic-range imaging](#)

Definition

High-dynamic-range imaging (HDRI or HDR) is a collection of hardware and software technologies that allow the capture, processing, and display of image and video content containing a wide range of intensities between the darkest and brightest areas of an image. Lighting conditions in both natural and man-made environments can range from starlight to artificial illumination to bright sunlight. Traditional imaging techniques typically store information using one byte per pixel for each channel, allowing for 256 distinct steps per channel. Consequently, they can only represent a narrow range of this illumination, often resulting in over- or undersaturated regions in the image. In contrast, HDR technologies represent image content using floating-point numbers and thus both allow for a much larger number of intensity steps to be encoded and reduce the distance between consecutive steps in image intensity.

Overview

Natural environments contain a vast range of illumination that extends from bright sunlight to starlight. In fact, lighting conditions in natural scenes can extend over more than 14 orders of magnitude. Human eyes are not able to perceive this enormous range simultaneously, but mechanisms in the visual system allow adaptation to the prevailing lighting conditions and surroundings. A recent study has evaluated the simultaneous range of the human visual system under extended



High Dynamic Range Imaging, Fig. 1 Both photographs capture the same scene but using different exposure levels. The foliage and colors of the tree are well exposed in the *left image* but the sky has been overexposed and has no details visible. The *right image* on the other hand shows

the details of the cloudy sky much better but at the expense of most information in the rest of the image. Problems such as these can easily be overcome with the use of high-dynamic-range technologies

H

luminance levels and has found that under specific circumstances, it can reach (and theoretically exceed) 3.7 orders of magnitude [1].

Traditionally, digital images are stored using one byte (8 bits) per pixel to represent the values for each of the three channels, allowing for 256 distinct levels between the darkest and brightest value within each channel. Consequently, standard imaging techniques can capture and store roughly 2 orders of magnitude between black and white, separated by distinct steps. Any information exceeding the available range appears over- or underexposed, while values between intensity levels are mapped to one of the 256 available levels, introducing steps into smooth gradients.

This is a familiar problem for photographers, as many cases exist where this range cannot adequately capture the full scene. For instance, when capturing the interior of a room, if there is a window with bright sunlight outdoors, the region covered by that window would be significantly brighter than the interior. The photographer would then have to choose whether to expose for the room interior or the window to be captured correctly, losing the remaining information of the scene. Such a solution could be sufficient in an artistic context; however, when images are intended as a realistic and complete representation of a scene, the photographer's choices and limitations of the equipment can affect the results.

Useful information may not be captured or images may include bias, as the photographer needs to select an exposure level. Figure 1 shows two photographs of the same scene exemplifying this problem.

High-dynamic-range imaging (HDRI) is a collection of tools and algorithms that allow the capture, storage, processing, and display of images with a wider dynamic range than that afforded by traditional imaging techniques. High-dynamic-range (HDR) images use floating-point numbers to represent values in pixels, leading to two main differences compared to traditional, 8-bit images. First, a much wider range of values can be encoded, far exceeding the dynamic range occurring in nature. Second, the use of floating-point representation means HDR images effectively have no quantization since consecutive steps in intensity can be much smaller.

Imaging Pipeline

These inherent differences between HDR and traditional imagery, in addition to allowing more information to be captured and opening up many possibilities in terms of applications, also mean that all stages of an imaging pipeline need to be modified. From the capture of HDR images and video, to storage solutions, to processing, and finally to display, many existing algorithms



High Dynamic Range Imaging, Fig. 2 An HDR image (here shown tonemapped) constructed by combining the seven exposures shown on the *left*. Each individual exposure cannot fully capture all the detail in the scene

make assumptions that do not apply to images and video content with a wider dynamic range.

Capture

To capture the full dynamic range of a scene such as the one shown in Fig. 1, a camera would need to be able to capture both bright and darker areas simultaneously. Although this is not currently possible using a single camera sensor, an HDR image can be constructed by capturing a scene at multiple exposure levels, which are combined into a single image [2].

Typically, multiple exposures are captured sequentially, with all camera settings apart from the exposure time fixed to minimize changes (many recent cameras include “Auto-Exposure Bracketing” settings to simplify this process). One example of a series of exposures is shown in Fig. 2. Although this is often an effective solution, as consecutive exposures are taken at slightly different times, fast-changing scenes (e.g., moving people or tree branches) can lead to misaligned exposures or ghosting artifacts [3]. To counter these issues, differently exposed versions of the scene can be captured concurrently

using specialized cameras, where multiple sensors record the same scene at different exposure levels [4].

When the HDR image itself is not required, multiple exposures can be directly combined into an 8-bit image, bypassing the floating-point representation. This process is known as *exposure fusion* [5].

Storage

A direct consequence the higher fidelity afforded by HDR is the increased storage requirement. Existing image file encodings are not readily amenable to the storage of HDR images as they often rely on the well-defined range of values encountered in traditional imagery. As such, several new file formats and extensions to existing formats have been developed for the storage of HDR images [3]. Video formats, although a more recent area of study, have also successfully been developed [6].

Display

The final stage within an HDR image pipeline is the display of still and video content, which can be

achieved in two different ways. A number of display devices capable of an extended dynamic range have been developed, albeit mostly at a prototype level. The most common solution combines a spatially varying backlight with an LCD front panel, extending the dynamic range compared to conventional displays both toward the dark and the bright ends of the spectrum. Solutions exist that combine a DLP projector, placed behind an LCD panel [7], or alternatively, a matrix of LED light sources can replace the projector [8].

Although HDR display technologies are becoming more feasible, they are still only capable of displaying a fraction of the dynamic range available in nature. As such, to display images of arbitrary dynamic ranges on hardware with more restricted capabilities (including commercially available monitors), algorithms that can compress the range of the image to that of the display are necessary. Many such techniques exist and are known as *tonemapping operators*, which will be discussed in the following section.

Tone Reproduction

Different devices are capable of displaying varying dynamic ranges. To ensure that HDR images and content are compatible both with existing, more limited displays and with future ones, solutions allowing the HDR content to be mapped to the capabilities of a particular display are necessary. Since the output of such a mapping should still be useful for visual consumption, the form of this mapping is an important consideration.

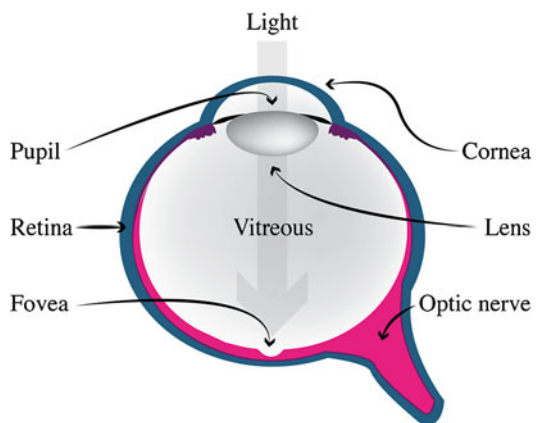
In the simplest scenario, one could linearly scale an HDR image to the given display range. Such a solution though would not be adequate for many scenes. Luminance values in HDR images are not linearly distributed across the available range. Instead, the distribution of luminance values in HDR imagery is highly kurtotic: sources and highlights only cover a small number of pixels in the image but are significantly brighter than the majority of the scene, while most pixels represent a relatively small range of intensities. In practice, the consequence of this is that a

nonlinear mapping is necessary to ensure that sufficient detail in the scene remains visible.

The aim of tonemapping techniques can then be seen as the compression of the range of an HDR image such that it can be displayed on devices of a lower dynamic range, while preserving the visible detail in the scene [9]. Often, this is also combined with a transition from floating point to integer steps in intensity. Compressing the dynamic range of an image leads inevitably to the loss of some information. Consequently, tonemapping operators (TMOs) need to decide which information to keep and which to discard. The human visual system faces a similar challenge, as it is only capable of seeing a considerably smaller range than that available in nature. It should not come as a surprise then that dynamic range reduction is often inspired by aspects of human vision. Before discussing the different tonemapping techniques and their respective merit, a better understanding of the human eye and the relevant processes to tone reproduction is necessary.

Perceptual Background

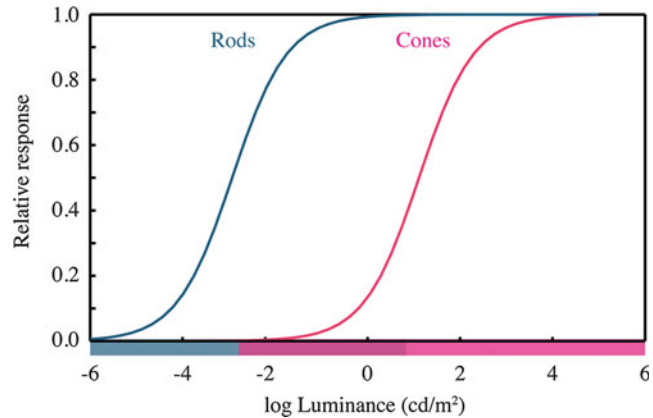
Human eyes are one of the means for sensing the world. Figure 3 shows an annotated cross section of the human eye. Light enters the eye through the pupil, travels through the ocular media, and finally hits the retina. The retina is a layer of neural sensors lining the back of the eye, which



High Dynamic Range Imaging, Fig. 3 A simplified cross section of the human eye

High Dynamic Range Imaging,

Fig. 4 Photoreceptors respond nonlinearly to increasing luminance levels. The Naka-Rushton equation models this nonlinearity, producing S-shaped curves when plotted in a log-linear scale, known as sigmoid



transduce light into a signal that is transmitted to the brain. Although several layers of neurons are present at the retina, it is the photoreceptors that are sensitive to light. Specifically, rod photoreceptors are sensitive to low light conditions, while cone photoreceptors are sensitive to brighter conditions.

One of the most important properties of the human visual system with respect to dynamic range reduction is its ability to adapt to a wide range of illumination. This effect is experienced on a daily basis: it is sufficient to walk into a dark room from bright sunlight for the eyes to adjust to see under the new conditions. Effectively, the human visual system adapts to the prevailing lighting conditions so that contrasts in the scene can still be distinguished [10].

It is not a single process within the human visual system that is responsible for this effect, but rather a combination of factors. One familiar to most is the adjustment of the diameter of the pupil depending on the illumination. Another set of mechanisms occurring in the photoreceptors, which prevents them from saturating while adapting to the enormous range of illumination in nature, is their nonlinear response to illumination. This nonlinearity was first studied by Naka and Rushton [11] and provides a very effective mechanism for the compression of dynamic range. Intensities in the middle of the range lead to approximately logarithmic responses, while as the intensity is increased, the response tails off to a maximum. After that maximum is reached, further increases in intensity will not lead to a

corresponding increase in response of the photoreceptors, ensuring that saturation will not occur [12].

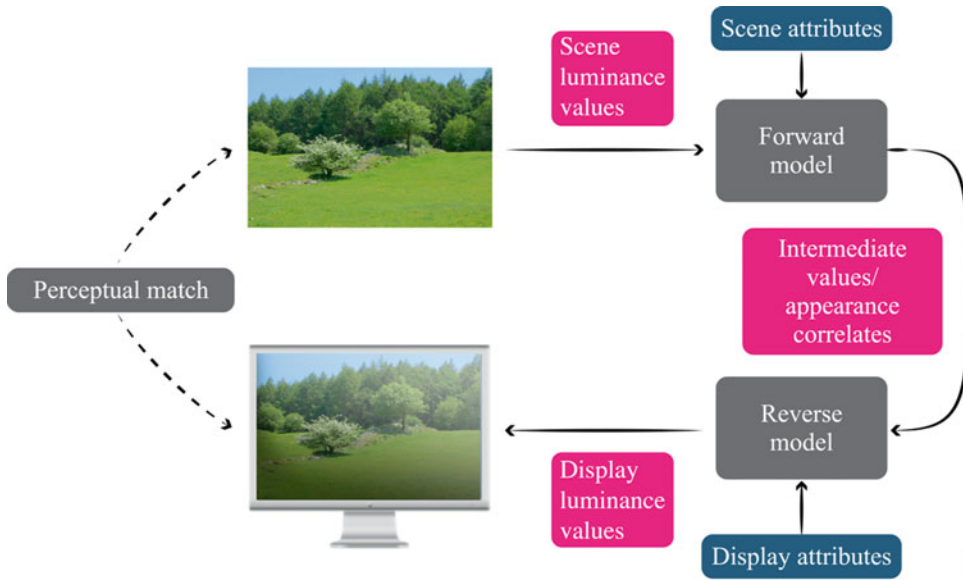
Figure 4 shows the nonlinear response of photoreceptors to increasing luminance values, which can be modeled with what is known as the *Naka-Rushton equation*:

$$\frac{V}{V_{\max}} = \frac{I^n}{I^n + \sigma^n} \quad (1)$$

where V is the response at an intensity I , V_{\max} is the peak response at saturation, and σ is the intensity necessary for the half-maximum response. The exponent n controls the slope of the function and is generally reported to be in the range of 0.7–1.0 [12]. This functional form is known as a *sigmoid* because it forms an S-shaped curve when plotted on log-linear axes. This aspect of light adaptation of the visual system has played an instrumental role in dynamic range compression, which will be discussed in the following section.

Perceptually Motivated Tone Reproduction

Many tonemapping operators use models of photoreceptor adaptation to effectively reduce the dynamic range of input HDR images. Typically this involves a forward step where the dynamic range of the image is compressed, often taking into account the lighting conditions and appearance of the scene, followed by an inverse step that considers the parameters of the display where the image will be viewed, as shown in Fig. 5. If the forward step is derived from a model of the human



High Dynamic Range Imaging, Fig. 5 The typical steps in the tonemapping process

visual system, it transforms the image from luminance values to perceived values, essentially modeling what the original scene would be perceived as if the viewer were present. The reverse step then takes the compressed (now perceived) values back to luminance values, allowing them to be correctly displayed [3]. Note however that many tonemapping operators do not employ the reverse step in order to maximize compression.

To approximate the compression afforded by photoreceptors, several functions have been proposed. At their simplest, linear scaling functions have been shown to be effective if the scaling factor is chosen appropriately, depending on the properties of the image and the intended display. Although such a linear compression scheme can create realistic results in less extreme cases, some information will be clamped for the image to fit within the displayable range. A better approximation of light adaptation behavior of photoreceptors can be achieved using a logarithmic mapping. Further, to better adapt this mapping to the image content, the base of the logarithms can be modified on a per-pixel basis in order to appropriately adjust the compression [3]. A perceptual evaluation of many of these methods can be found here [13]. More akin to photoreceptor

behavior are sigmoid models, which have been used extensively in tone reproduction [9, 14, 15].

More complete models of photoreceptor physiology and light adaptation have been repeatedly used in the tone reproduction literature, offering perceptually accurate reproductions but at the cost of increased computational complexity. An early and fairly complete model that incorporates luminance, color, and pattern representations to create a perceptually accurate compressed image employs a multi-scale processing scheme [16]. Another notable example that lends itself to HDR compression is the *retinex* model, which considers aspects of both the retina and the brain in order to model complex perceptual phenomena such as color constancy [17].

Multi-scale Tone Reproduction

Features in images occur at different scales, and many tone reproduction techniques employ mechanisms that treat images at multiple scales, improving the appearance of the final result as a different amount of compression can be applied to different features in the image. This ensures that high-contrast features are sufficiently compressed, while local edges (e.g., texture) are not proportionally scaled as that would reduce their

visibility. To decompose the image into different scales, a form of edge-preserving filtering is typically used, such as the *bilateral filter* [3].

Display and Viewing-Adaptive Tonemapping

Most tone reproduction algorithms discussed so far aim to compress the range of a given image to a displayable range, typically between 0 and 255, but with no consideration of any further properties of the device or the environment where the image will be viewed. Although many detailed models exist that take into account aspects of the environment where the image was captured as well as where and how it will be displayed, such models tend to require additional input that is usually not available in a typical tonemapping setting.

In addition to the parameters of the viewing environment, different display devices offer varying capabilities that need to be considered. Mantiuk et al. [18] recently formulated tonemapping as an optimization problem that strives to reduce the dynamic range of an image while minimizing the visible distortions that will inevitably be introduced in the process. To detect visible distortions, a model of the human visual system is used, and to ensure that the resulting image is suitable for the target device, a display model is incorporated. Such a scheme allows for images to be tonemapped even for unconventional devices and viewing conditions.

Dynamic Range Expansion

Tonemapping techniques discussed so far compress the dynamic range of images to prepare them for displays with more limited capabilities. With the advent of HDR displays, the inverse problem arises. To fully take advantage of the extended dynamic range available on such a display, it may be desirable to expand the intensity range of traditional image and video content, a process known as *dynamic range expansion* or *inverse tonemapping*. This can be achieved by applying an expansion function to the intensity values of the image to map them to the range of the display. Both nonlinear (such the inverse of

Eqs. 2 and 4) and linear schemes have been proposed [19], but a recent psychophysical study suggested that linear expansion leads to the most plausible results [20]. Alternatively, an observation that has been exploited in inverse tonemapping is that in many scenes, luminance distribution is highly kurtotic, with most pixels covering a relatively small part of the range, while most dynamic range is allocated to light sources and highlights. To simulate this when expanding the dynamic range of images, bright areas of the image can be enhanced more [21, 22].

A final issue to be considered in dynamic range expansion is that of under- or overexposed areas in the image. Even in a well-exposed image, light sources and highlights are likely clip to white, leading to loss of detail in those areas. When expanding such images, it is desirable to reintroduce some of the lost detail in these areas, which can be achieved through *hallucination* techniques, where overexposed regions are filled in using information from surrounding areas of the image [23].

Color Management in Tone Reproduction

Most of the tone reproduction operators discussed so far operate on a luminance representation of the image. Typically, a tonemapping operator will compress the world luminance L_w , which is computed from the RGB values in the image, to a display luminance L_d . To form the resulting tonemapped image, the RGB values will have to be scaled in such a way that the ratio between the three channels is kept the same before and after the compression [14]. This can be achieved as follows:

$$\begin{bmatrix} R_d \\ G_d \\ B_d \end{bmatrix} = \begin{bmatrix} L_d \frac{R_w}{L_w} \\ L_d \frac{G_w}{L_w} \\ L_d \frac{B_w}{L_w} \end{bmatrix} \quad (2)$$

This process preserves the ratios between channels but it is not sufficient to maintain the

appearance of colors in the image. Saturation depends on both chromatic information and intensity values, and thus, luminance compression through tonemapping can lead to an oversaturated appearance (or undersaturated in the case of dynamic range expansion). To control the saturation of the tonemapped image, an exponent s can be used when reconstructing the three color channels from the adjusted luminance [24] (here shown only for the red channel):

$$R_d = L_d \left(\frac{R_w}{L_w} \right)^s \quad (3)$$

Although this approach is frequently employed as part of tonemapping operators, it may cause undesirable luminance shifts. Alternatively, the three channels can be recomputed from the adjusted luminance as follows:

$$R_d = L_d \left(\left(\frac{R_w}{L_w} - 1.0 \right) s + 1.0 \right) L_d \quad (4)$$

Both solutions offer simple saturation control but rely on manual selection for the parameter s . To automate the parameter selection for global tonemapping operators, a recent series of psychophysical experiments has derived a link between the tonemapping curve and the saturation parameter for both correction formulas [25]. Despite this improvement though, such post hoc corrections cannot guarantee that the color appearance of the image will be preserved.

If additional information is available about both the image itself and the target viewing environment and device, accurate treatment of luminance in conjunction with color is possible using *color appearance models* (CAMs). Most such models are designed with accurate color reproduction as their primary aim and operate on a limited range of intensities. The increasing availability of HDR devices and content, however, has led to the development of several models capable of handling content of extended dynamic ranges, effectively combining tone reproduction and color appearance modeling. In contrast to the operators discussed so far, CAMs process each

channel of the image separately rather than only compress luminance [26]. Typically, such models employ both a forward and an inverse step, taking the scene and viewing parameters into account, although it was recently shown that the inverse step may be bypassed, while still accurately reproducing color appearance [27].

Cross-References

- [Adaptation](#)
- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIECAM02](#)
- [CIELAB](#)
- [Color Constancy](#)
- [Color Spaces](#)
- [Image Quality](#)
- [Psychological Color Space and Color Terms](#)
- [Unique Hues](#)

References

1. Kunkel, T., Reinhard, E.: A reassessment of the simultaneous dynamic range of the human visual system. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization. ACM, New York (2010)
2. Debevec, P., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: Proceedings of the SIGGRAPH '97: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. ACM, New York (1997)
3. Reinhard, E., Ward, G., Pattanaik, S., Debevec, P., Heidrich, W., Myszkowski, K.: High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting. Morgan Kaufmann Publishers, San Francisco (2010)
4. Tocci, M.D., Kiser, C., Tocci, N., Sen, P.: A versatile HDR video production system. ACM Trans. Graph. **30**(4): 41 (2011), 41–41: 10
5. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion: a simple and practical alternative to high dynamic range photography. Comput. Graph. Forum **28**(1), 161–171 (2009)
6. Myszkowski, K., Mantiuk, R., Krawczyk, G.: High dynamic range video. Synthesis Lect. Comput. Graph. Animation **2**(1), 1–158 (2008)
7. Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., Vorozcovs, A.: High dynamic range display systems. ACM Trans. Graph. **23**(3), 760–768 (2004)

8. Seetzen, H., Whitehead, L.A., Ward, G.: 54.2: A High Dynamic Range Display Using Low and High Resolution Modulators. SID Symp. Dig. Tech. Pap. **34**, 1450–1453 (2003)
9. Tumblin, J., Rushmeier, H.: Tone reproduction for computer generated images. IEEE Comput. Graph. Appl. **13**(6), 42–48 (1993)
10. Wandell, B.A.: Foundations of Vision. Sinauer Associates, Sunderland (1995)
11. Naka, K.I., Rushton, W.A.H.: S-potentials from luminosity units in the retina of fish (Cyprinidae). J. Physiol. **185**, 587–599 (1966)
12. Dowling, J.E.: The Retina: An Approachable Part of the Brain. Belknap, Cambridge (1987)
13. Čadík, M., Wimmer, M., Neumann, L., Artusi, A.: Evaluation of HDR tone mapping methods using essential perceptual attributes. Comput. Graph. **32**(3), 330–349 (2008)
14. Schlick, C.: Quantization techniques for the visualization of high dynamic range pictures. In: Proceedings of the Photorealistic Rendering Techniques. Springer, Berlin/Heidelberg/New York (1994)
15. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. ACM Trans. Graph. **21**(3), 267–276 (2002)
16. Pattanaik, S.N., Ferwerda, J.A., Fairchild, M.D., Greenberg, D.P.: A multiscale model of adaptation and spatial vision for realistic image display. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '98), pp. 287–298. ACM, New York (1998). doi: 10.1145/280814.280922
17. McCann, J.J., Rizzi, A.: The Art and Science of HDR Imaging. Wiley, Hoboken (2011)
18. Mantiuk, R., Daly, S., Kerofsky, L.: Display adaptive tone mapping. ACM Trans. Graph. **27**(3), 68 (2008)
19. Banterle, F., Artusi, A., Debattista, K., Chalmers, A.: Advanced High Dynamic Range Imaging: Theory and Practice. AK Peters/CRC Press, Natick (2011)
20. Oğuz Akyüz, A., Fleming, R., Riecke, B.E., Reinhard, E., Bühlhoff, H.H.: Do HDR displays support LDR content?: a psychophysical evaluation. ACM Trans. Graph. **26**(3), Article 38 (2007). doi:10.1145/1276377.1276425
21. Didyk, P., Mantiuk, R., Hein, M., Seidel, H-P.: Enhancement of bright video features for HDR displays. In: Computer Graphics Forum, vol. 27, no. 4, pp. 1265–1274. Blackwell (2008)
22. Rempel, A.G., Trentacoste, M., Seetzen, H., Young, H.D., Heidrich, W., Whitehead, L., Ward, G.: Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs. ACM Trans. Graph. **26**(3), 39 (2007)
23. Wang, L., Wei, L.-Y., Zhou, K., Guo, B., Shum, H.-Y.: High dynamic range image hallucination. In: Proceedings of the 18th Eurographics Conference on Rendering Techniques, pp. 321–326. Eurographics Association (2007)
24. Tumblin, J., Hodgins, J.K., Guenter, B.K.: Two methods for display of high contrast images. ACM Trans. Graph. **18**(1), 56–94 (1999)
25. Mantiuk, R., Mantiuk, R., Tomaszewska, A., Heidrich, W.: Color correction for tone mapping. Comput. Graph. Forum (Proc. Eurographics '09) **28**(2), 193–202 (2009)
26. Fairchild, M.D.: Color Appearance Models. Addison-Wesley, Reading (2005)
27. Reinhard, E., Pouli, T., Kunkel, T., Long, B., Ballestad, A., Damberg, G.: Calibrated image appearance reproduction. ACM Trans. Graph. **31**(6), 201 (2012), 201–201: 211

High-Pressure Mercury Gas-Discharge Lamp

► High-Pressure Mercury Lamp

High-Pressure Mercury Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

High-pressure mercury gas-discharge lamp;
High-pressure mercury-vapor lamp

Definition

Lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a high-pressure mercury vapor that is contained in a transparent bulb. In some versions, fluorescent powder is applied that converts the ultraviolet part, which is emitted together with visible light, into visible light to improve the color quality of the light.

Modified and reproduced from [1] by permission of Philips Lighting, Eindhoven, The Netherlands. © 2012 Koninklijke Philips Electronics N.V.

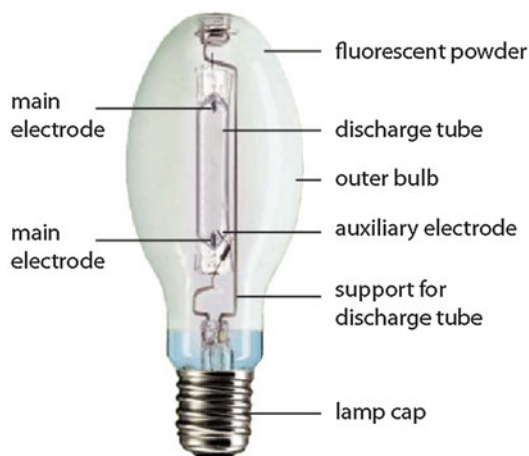
High-Pressure Mercury Gas-Discharge Lamps

High-pressure mercury lamps belong to the group of high-intensity discharge (HID) lamps, because they are available in high-lumen output (and thus high-luminous-intensity) versions. High-pressure mercury lamps are available in versions where the discharge takes place in vaporized mercury only and in versions in which metal halides are added so that the discharge takes place in mercury vapor and in vaporized metals from the metal-halide components [2, 3]. These latter types are called metal-halide lamps. Given the special operating principle and construction of metal-halide lamps, they are dealt with in a separate entry “Metal-halide lamp.”

High-pressure mercury lamps, like all high-pressure discharge lamps, are compact compared to low-pressure discharge lamps. They have a moderate efficacy and moderate color rendering. With their cool-white light, they were, until the 1970s of the last century, extensively used in road lighting, especially in built-up areas. Since the introduction of the more efficient high-pressure sodium lamps in the late 1960s, these lamps have in many cases replaced high-pressure mercury lamps.

Working Principle

The gas-discharge principle of high-pressure mercury lamps is similar to that of all other gas-discharge lamps. In high-pressure mercury lamps, the discharge takes place in vaporized mercury at a pressure of around 10^6 Pa (10 atm). The spectrum of the radiation is a line spectrum with emissions in the long-wave UV region and in the visible region at the yellow, green, blue, and violet wavelengths. The lamp without fluorescent powder lacks red in its spectrum and has a bluish-white color appearance and very poor color rendering. In most high-pressure mercury lamps, fluorescent powders are used to improve the color quality by converting a large part of the (small) UV component into visible radiation, predominantly in the red end of the spectrum. The result is cool-white light of moderate color rendering and improved efficacy. Like all gas-discharge lamps (with very few



High-Pressure Mercury Lamp, Fig. 1 Principle parts of a high-pressure mercury gas-discharge lamp [1]

H

exceptions), a high-pressure mercury lamp cannot be operated without a ballast to limit the current flowing through it.

Materials and Construction

The main parts of a high-pressure mercury lamp are (Fig. 1):

- Discharge tube
- Fill gas
- Electrodes
- Outer bulb (often with fluorescent coating)
- Lamp cap

Discharge Tube

In view of the high operating temperature, quartz is used for the discharge tube because it has a higher melting temperature than glass.

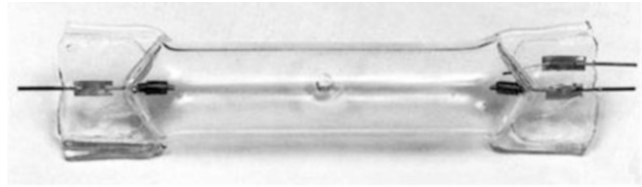
Fill Gas

The discharge tube contains a small quantity of mercury and an inert gas filling.

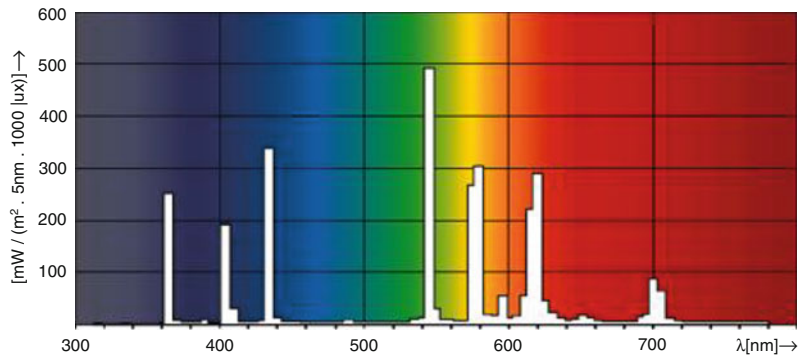
Electrodes

The main electrodes consist of a core of tungsten rod with a tungsten coil (impregnated with emissive material) wound around it. To aid starting, a normal high-pressure mercury lamp has not only an inert gas but also an auxiliary electrode. Because of this, a normal mercury lamp does not

High-Pressure Mercury Lamp, Fig. 2 Discharge tube showing the auxiliary electrode and two main electrodes



High-Pressure Mercury Lamp, Fig. 3 Spectral energy distribution of a high-pressure mercury lamp with fluorescent powder [1]



need an external igniter. The auxiliary electrode simply consists of a tungsten wire positioned very close to one of the main electrodes (Fig. 2).

Outer Bulb

An outer bulb (usually ovoid in shape) with an inert gas filling isolates the gas-discharge tube so that changes in ambient temperature have no influence on its proper functioning. It also protects the lamp components from corrosion at the high operating temperatures involved. For the smaller wattage lamps, with their lower operating temperatures, normal glass is used, while for the other types, hard glass is used.

Fluorescent Powder

As has already been mentioned, high-pressure mercury lamps usually employ fluorescent powder to improve the color quality of the light emitted. The powder is provided as a coating on the inner surface of the outer bulb. Different fluorescent coatings are used to obtain different lamp types with different color qualities and lamp efficacies.

Lamp Cap

Lamp caps are of the Edison-screw type, with the wattage of the lamp determining their size (E27 and E40).

Properties

Energy Balance

Approximately 17 % of the input power is emitted in the form of visible radiation. Compare this with the 28 % of a tubular fluorescent lamp and the 30 % of a high-pressure sodium lamp.

Luminous Efficacy

Luminous efficacy varies with lamp wattage and with the color quality of the lamp from some 35 to 60 lm/W.

Lumen-Package Range

High-pressure mercury lamps are produced in lumen packages between some 2,000 and 60,000 lm (corresponding wattages between 50 and 1,000 W).

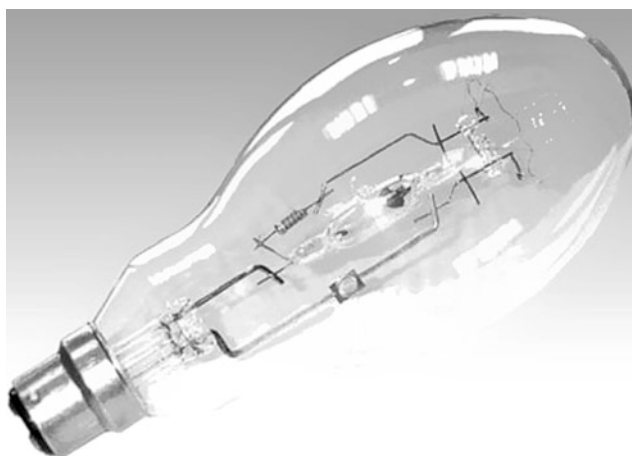
Color Characteristics

As has already been mentioned, high-pressure mercury lamps have a line spectrum (Fig. 3). The two lines in the red part of the spectrum are obtained by conversion of ultraviolet radiation by the fluorescent powder. The color characteristics are dependent on the composition and quality of the fluorescent powders used. Different compositions and qualities are used to produce lamps with

High-Pressure Mercury Lamp, Fig. 4 Different wattage ovoid and reflector type of high-pressure mercury lamps [1]



High-Pressure Mercury Lamp, Fig. 5 Blended light lamp with high-pressure mercury gas-discharge tube and tungsten incandescent filament combined in one outer bulb. To show the construction, a clear lamp is shown, although most blended light lamps use a fluorescent powder coating on the interior of the outer bulb [1]



H

color temperatures between some 3,500 and 4,500 K, with color-rendering index (R_a) values of around 60 for high-quality versions and around 40 for ordinary versions.

Lamp Life

As with most gas-discharge lamps, lamp life is determined by emitter exhaustion of the electrodes. Economic life varies according to type between 10,000 and 15,000 h (20 % mortality).

Lamp-Lumen Depreciation

Lamp-lumen depreciation is caused by evaporation and scattering of electrode material (lamp blackening) and by the gradual decrease in the activity of the fluorescent powder. The point at which 20 % lumen depreciation occurs lies at around 10,000–15,000 h.

Run-Up and Re-ignition

The run-up time of a high-pressure mercury lamp to its full temperature and corresponding nominal mercury pressure is some 4 min. The hot lamp will not restart until it has cooled sufficiently to lower the vapor pressure to the point at which restrike with the voltage available is possible. The re-ignition time is in the order of 5 min.

Dimming

High-pressure mercury lamps cannot be dimmed.

Mains-Voltage Variations

A 5 % variation in the mains voltage changes both lamp current and light output by 10 %. Overvoltage decreases lamp life and increases lamp depreciation because of the correspondingly higher current.

Product Range

High-pressure mercury lamps are available in an ordinary version with poor color rendering (R_a of around 40) and in so-called comfort versions with an improved color rendering of around 60. The bulb is ovoid in shape and increases in size with increase in wattage (Fig. 4). Versions without fluorescent powder are hardly, anymore, produced. Reflector lamp versions are produced with a cone-shaped outer bulb and an internal reflective coating on the front (Fig. 4, right).

There is one version of the high-pressure mercury lamp, the “blended light lamp,” that does not need an external ballast. The ballast has simply been built into the lamp itself in the form of a tungsten filament (Fig. 5). The lamp can be connected direct to the mains. The light from the mercury discharge and that from the heated filament blend together (hence the name blended light lamp). The color characteristics of this lamp are therefore better than those of a normal high-pressure mercury lamp, but this comes at the cost of a considerably lower efficacy.

Cross-References

- [Metal Halide Lamp](#)
- [Phosphors and Fluorescent Powders](#)
- [Tubular and Compact Fluorescent Lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Course book, Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn Illuminating Engineering Society of North America, IESNA, New York (2011)

High-Pressure Mercury-Vapor Lamp

- [High-Pressure Mercury Lamp](#)

Highway Lighting

- [Road Lighting](#)

History, Use and Performance

- [CIELAB](#)

Hue Difference, Delta H

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

Hue Opponency

- [Color Vision, Opponent Theory](#)

Hue Sequence

- [Color Circle](#)

Ibn Rushd (Averroes)

Eric Kirchner

Kirchner Publications, Leiden, The Netherlands



Abu'l-Walid Muhammad bin Ahmad Ibn Rushd (Latinized name: Averroes) was born in Cordoba (Spain, 1126) and died in Marrakech (Morocco, 1198). His works range from philosophy, astronomy, and medicine to religion. He would be the most influential of Aristotle's medieval commentators, writing comments, and corrections on all the Aristotelian works available to him [1].

Scholars in Christian Europe would refer to him as “the Commentator [of Aristotle].” The revival of Aristotelism in twelfth century Europe was mainly based on Ibn Rushd's commentaries. But in the Islamic world, Ibn Rushd's defense of rationalist philosophy would be less influential than the religious Asharite philosophy that emphasized divine manifestations in nature, as advocated by al-Ghazali.

Ibn Rushd tried to formulate a theory of light and color, as consistent with Aristotle and his followers as possible. Like Ibn Sina had done before, Ibn Rushd provided many arguments why the so-called extramission theories from Ptolemy and Euclid are absurd, and vision cannot be caused by visual rays emerging from the eye. And similar to Ibn Sina, Ibn Rushd explained vision in terms of Forms transmitted from the visible object to the eye, i.e., an intromission theory [2].

Ibn Rushd maintained that colors exist even when they are not perceived, and that light is necessary for colors to be visible [3]. Like Ibn al-Haytham had done before, Ibn Rushd considered light as playing an active role in color vision, thus breaking with the traditional Aristotelian view. But unlike Ibn al-Haytham, Ibn Rushd did agree with Aristotle that also the medium played an active role in color vision. Apparently, Ibn Rushd was not aware of the theory on vision that had been proposed by Ibn al-Haytham, successfully combining the various classical theories into one mathematical–physical intromission theory of light, color and vision.

Ibn Rushd explained the different species of color as consisting of various mixtures of bodies of much or little transparency with bodies of much or little luminosity. Since every material is composed of the four elements, with only water and air being transparent and only fire being luminous, the color of a material can be attributed to the relative amounts of the elements. This explanation was later adopted by Theodoric of Freiberg (p. 172 of Ref. [4]), who explained it in terms of four principles (much or little transparency, much or little luminosity) of colors.

Ibn Rushd did not agree with Ibn Sina's criticism of the Aristotelian ideas on color mixing. In his *Jawāmi' al-āthār al-'ulwiyya* (Short Commentary on the Meteorology: 74, 19–76, 18), Ibn Rushd defended the Aristotelian view by arguing that it referred to color mixing in a qualitative sense, but not in a quantitative sense [5, 6]. For example, Ibn Rushd argued that green is formed by mixing the yellow that exists in light red with the black that is in purple. From a modern point of view, this makes the approach philosophical rather than scientific.

Also in his description of the rainbow, Ibn Rushd closely followed Aristotle: the rainbow forms by reflection of sunlight on individual rain drops that reflect light and transmit color [4].

Ibn Rushd's ideas on color would be influential on later scholars. In his works on colors (*De coloribus*) and on the rainbow (*De iride*), Theodoric of Freiberg (d. 1318) makes clear that his theory on the formation of colors is based on Ibn Rushd's color theory (p. 47 of Ref. [7]; p. 11 of [8]). Theodoric literally quotes the Latin translation of Ibn Rushd's *Tractato de Sensu et Sensato* (p. 35n2 of Ref. [7]) [9]. His description of the colors of the rainbow, and also his four principles by which color is explained were all explicitly taken from Ibn Rushd's work.

References

1. Lindberg, D.C.: Theories of Vision from al-Kindi to Kepler. University of Chicago Press, Chicago (1976)
2. Sabra, A.I.: Optics, islamic. In: Strayer, J.R. (ed.) Dictionary of the Middle Ages, pp. 240–247. Scribner's sons, New York (1989)
3. Morabia, A.: Lawn. In: Bosworth, C.E. (ed.) Encyclopaedia of Islam. Brill, Leiden (1991)
4. Wallace, W.A.: The Scientific Methodology of Theodoric of Freiberg – A Case Study of the Relationship Between Science and Philosophy. University Press, Frimbourg (1959)
5. van Campen, M.: De regenboog bij de Arabieren. University of Utrecht, Utrecht MSc Thesis, 95 (1988)
6. Ibn Rushd: Short commentary: Kitab al-athar al-'ulwiyya. In: Osmaniya, D.M. (ed.) Rasa'il Ibn Rushd, vol. 4. Hyderabad, (1947). Another edition is: Abu Wafia, S.F., 'Abd ar-Raziq, S.A. (eds.) Kitab al-atar al-'ulwiyya. Epitome Meteorologica, Cairo (1994). (Latin translation: Aristotle, Aristotelis omnia quae extant opera, vols. V. and VI/2)
7. Krebs, E.: Meister Dietrich – Theodoricus Teutonicus de Vriberg – Sein Leben, seine Werke, seine Wissenschaft. Aschendorffsche Verlagsbuchhandlung, Münster (1906)
8. Würschmidt, J.: Dietrich von Freiberg – über den Regenbogen und die durch Strahlen erzeugten Eindrücke. Aschendorffsche Verlagsbuchhandlung, Münster (1914)
9. Cordubensis, A.: De sensu et sensato. In: Ledyard Shields, A. (ed.) Compendia librorum Aristotelis qui parva Naturalia vocantur, pp. 14–16. Mediaeval Academy of America, Cambridge, MA (1949)

Ibn Sahl, Abu Sa'd al-'Ala'

Eric Kirchner¹ and Seyed Hossein Amirshahi²

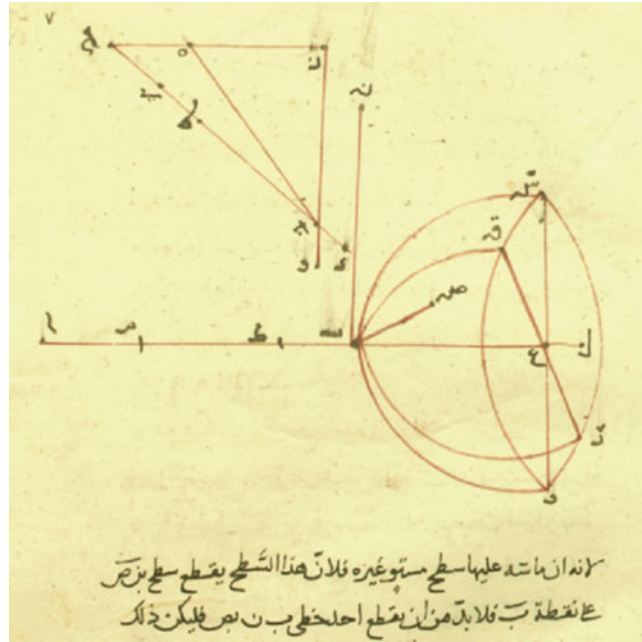
¹Kirchner Publications, Leiden, The Netherlands

²AUT Textile Department, Amirkabir University of Technology, Tehran Polytechnic, Tehran, Iran

Abu Sa'd al-'Ala' Ibn Sahl lived and worked as a geometer at the Abbasid (circa 940 - 1000) court in Baghdad. He wrote important works on geometric optics, mathematics, and astronomy. Nowhere in his extant works, Ibn Sahl dealt with (or even mentions) vision [1].

While investigating the transparency of the heavenly spheres that occur in Aristotelian cosmology, Ibn Sahl decided to study Ptolemy's classical work *Optics*, written in the second century AD. This made Ibn Sahl to be the first in the Arabic sources to have read and correctly understood Ptolemy's theory of refraction [1]. Ibn Sahl utilized this theory in an entirely original way for

With the triangles in the top diagram, Ibn Sahl illustrated the law that is currently known as Snell's law. © Iranian National Library, Tehran. Manuscript MS 687



constructing burning instruments such as lenses and glass spheres by means of refraction.

In the year 984, Ibn Sahl wrote the treatise “On Burning Mirrors and Lenses.” In this work, he investigated the optimum (nonspherical) shape of lenses and mirrors to focus light at a given distance. Ibn Sahl “appears to be the first in history to engage in research on burning lenses” [2]. He subsequently treated the parabolic mirror, the ellipsoidal mirror, the plano-convex lens, and the biconvex lens. His calculations on geometric aberration predate similar calculations done by Descartes in the 1620s. Ibn Sahl carried out these calculations both for light sources at an (almost) infinite distance such as the sun and for light sources at finite distances [3].

In these calculations, Ibn Sahl needed a law of refraction. Intriguingly, for this he used a law that is geometrically equivalent to Snell's law that would be rediscovered in 1621 by Willebrord Snellius and in 1602 by Thomas Harriot. The sine law of refraction was thus discovered by Ibn Sahl [4, 5].

The illustration shows a page from Ibn Sahl's manuscript. In the top part of the figure, two overlapping triangles are shown. The hypotenuse of the external triangle represents the

direction of incident light, while the hypotenuse of the internal triangle represents the direction of refracted light inside a transparent medium. By demanding that these two direction vectors intersect in one point, constructing the direction of refracted light with this figure is geometrically equivalent to keeping the ratio of sines of incident and refracted angles constant. This is Snell's law of refraction.

Ibn Sahl used this law of refraction several times in the treatise, but without explicitly stating it as a law [4, 5]. Indeed, the concept of natural law did not exist at the time. Ibn Sahl used his law as if it was a mathematical relation that was well known. He repeatedly applied this relation, utilizing the fact that the ratio between the sines of incoming and refracted angles is constant. He made no reference to the fact that this constant depends on material-dependent properties, i.e., what is now known as their refractive index.

Ibn Sahl's treatise was later used by Ibn al-Haytham in his investigations of refraction. Interestingly, Ibn al-Haytham apparently did not recognize the law of refraction as used by Ibn Sahl. Instead, Ibn al-Haytham started his own experimental investigation into finding a law of refraction.

References

1. Sabra, A.I.: Ibn al-Haytham's revolutionary project in optics: the achievement and the obstacle. In: Hogendijk, J.P., Sabra, A.I. (eds.) *The Enterprise of Science in Islam*, pp. 85–118. MIT Press, Cambridge, MA (2003)
2. Rashed, R.: Geometrical optics. In: Rashed, R. (ed.) *Encyclopedia of the History of Arabic Science*, vol. 2, pp. 643–671. Routledge, London (1996)
3. Berggren, L.: Ibn Sahl: Abu Sa'd al-'Ala'ibn Sahl. In: Hockey, T. (ed.) *The Biographical Encyclopedia of Astronomers*, p. 567. Springer, New York (2007)
4. Rashed, R.: Géométrie et dioptrique au Xe siècle: Ibn Sahl, al-Quhi et Ibn al-Haytham. *Les Belles Lettres*, Paris (1993)
5. Rashed, R.: A pioneer in anaclasses: Ibn Sahl on burning mirrors and lenses. *Isis* **81**, 464–491 (1990)

ICC L*a*b*

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

ICCLAB

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

Ideal Observer Models of Color Category Learning

► [Bayesian Approaches to Color Category Learning](#)

Illuminance Meter

Bor-Jiunn Wen
Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

Synonyms

[Luxmeter](#)

Definition

In photometry, an illuminance meter is a device that measures the total luminous flux incident on a surface, per unit area. It is a measure of how much the incident light illuminates the surface, wavelength-weighted by the luminosity function to correlate with human brightness perception.

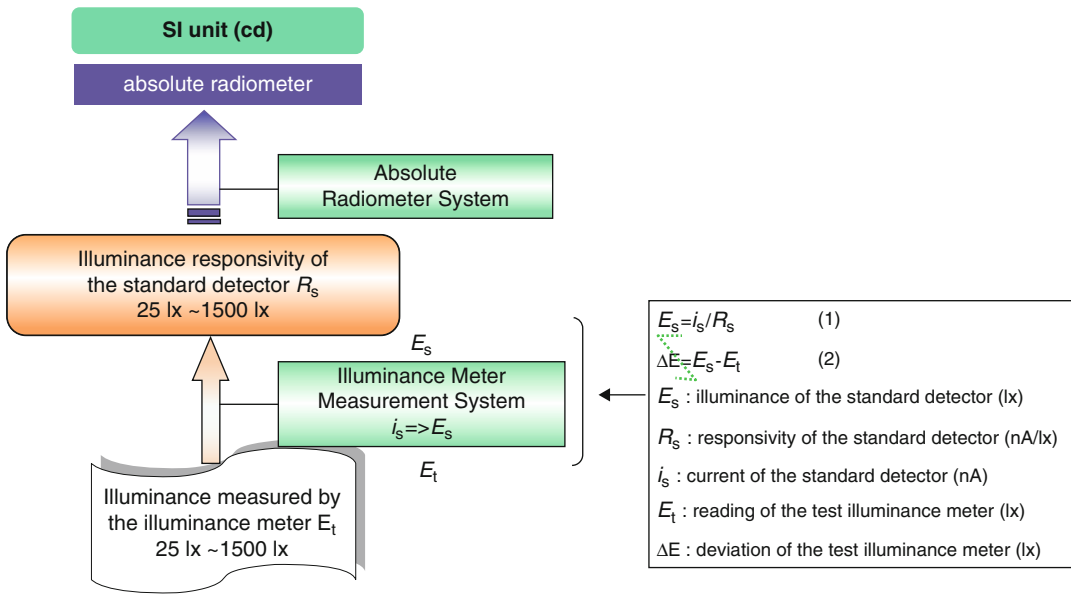
Introduction

There are many units for defining illuminance. The International System of Units defines using the unit of lux (lx) or lumens per square meter ($\text{cd} \cdot \text{sr} \cdot \text{m}^{-2}$), which is most frequently used. In the centimeter/gram/second system, the unit of illuminance is the phot, which is equal to 10,000 lx. The foot-candle is a non-metric unit of illuminance that is also used in photography. Illuminance is sometimes called brightness, but this leads to confusion with other uses of the word, such as brightness perception. Hence, “brightness” should never be used for quantitative description, but rather for describing the visual perceptions of light or color.

The human eye is capable of seeing within a range of more than two trillion-fold. The presence of white objects is somewhat discernible under starlight, at 5×10^{-5} lx, while at the bright end, we can easily read large text at 10^8 lx, or at about 1,000 times that of direct sunlight, although this can be very uncomfortable and cause long-lasting after-images.

Application 1

In addition to an illuminance meter, illuminance color spectral meters including spectrometer, spectrophotometer, spectrograph or spectroscopy are also important for measuring properties of light over a specific portion of the electromagnetic spectrum, typically used in spectroscopic analysis to identify materials [1]. The parameter measured is most often the light's intensity, but could also be, for instance, the polarization state. The independent variable is usually the wavelength of the



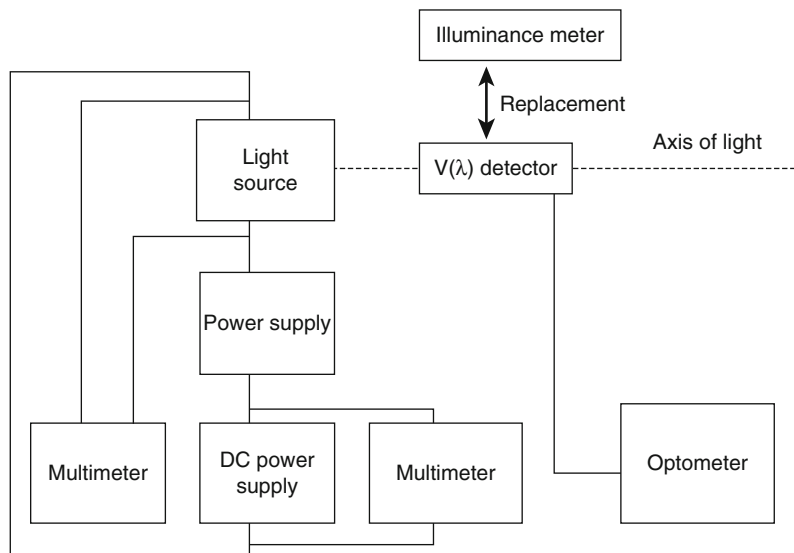
Illuminance Meter, Fig. 1 Traceability chain of illuminance

light or a unit directly proportional to the photon energy, such as wavenumber or electron volts, which has a reciprocal relationship with wavelength. A spectrometer is used in spectroscopy for producing spectral lines and measuring their wavelengths and intensities. "Spectrometer" is a term applied to instruments that operate over a very wide range of wavelengths, from gamma rays and X-rays into the far infrared. If the instrument is designed to measure the spectrum in absolute units rather than relative units, then it is typically called a spectrophotometer. The majority of spectrophotometers are used in spectral regions near the visible spectrum. A spectrograph is an instrument that separates an incoming wave into a frequency spectrum. There are several kinds of machines, referred to as spectrographs, depending on the precise nature of the waves. Spectroscopes are often used in astronomy and some branches of chemistry. Early spectroscopes were simply prisms with graduations marking wavelengths of light. Modern spectroscopes generally include a diffraction grating, a movable slit, and some kind of photodetector. It is all automated and controlled by a computer.

Application 2

In general, much of computer vision, image processing, and imaging is predicated on the assumption that there is a single prevailing illuminant lighting a scene. However, often multiple lights are present. Common examples include outdoor scenes with cast and attached shadows, indoor office environments that are typically lit by skylight and artificial illumination, and the spot-lighting used in commercial premises and galleries. Relative to these mixed lighting conditions, many imaging algorithms based on the single light assumption can fail. Therefore, Finlayson et al. [2] proposed a method of detecting multiple illuminants in images based on the chromogenic theory. By forcing the lights to be examined pair-wise without insisting on the accuracy of the illuminant estimation, and by processing the results on the basis of region rather than pixels, the research obtained very accurate results over a variety of illuminations. Accordingly, the results are remarkably good, especially considering the simplicity and speed of the method.

Illuminance Meter,
Fig. 2 Schematic diagram
 of illuminance meter
 calibration



Calibration

In the color measurement field, calibration is desired to ensure that the system is closely monitored in good working condition. The traceability chain of illuminance is presented in Fig. 1 [3]. An illuminance meter calibration set-up is shown in Fig. 2 [3]. First, warm up the multimeters. Check that the power of the device under test (DUT) is sufficient. Check the connections of all the instruments as shown in Fig. 1. They should be well connected, including the polarity of the lamp. Second, align the lamp filament to the axis of the light source and fix the position of the DUT. Third, turn on the power supply after the multimeters have already warmed up. Ramp up the lamp current to operation value. Execute calibration after the standard lamp has stabilized. Fourth, read the output current of the $V(\lambda)$ [4] (the spectral luminous efficiency function for photopic vision) detector, and calculate the illuminance of the position of the $V(\lambda)$ detector. Finally, replace the $V(\lambda)$ detector with the test sample at the same position, and record the test reading on the recording card.

References

1. Butler, L.R.P., Laqua, K.: Nomenclature, symbols, units and their usage in spectrochemical analysis-IX. Instrumentation for the spectral dispersion and isolation of optical radiation. *Pure Appl. Chem.* **67**(10), 1725–1744 (1995)
2. Finlayson, G., Fredembach, C., Drew, M.S.: Detecting illumination in images. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on Computer Vision*, pp. 1–8, 14–21 Oct 2007
3. Calibration Procedure for Illuminance Meter of Absolute Radiometer System, 07-3-80-0086, 3rd edn. Center for Measurement Standards/Industrial Technology Research Institute (2002)
4. Ohta, N., Robertson, A.R.: *Colorimetry Fundamentals and Applications*. John Wiley & Sons, Inc. (2005)

Illusory Colors

► [Fechner's Colors and Behnam's Top](#)

Image Burn-In

► [Afterimage](#)

Cross-References

► [Instrument: Spectrophotometer](#)
 ► [Spectrometer](#)

Image Fidelity

► [Image Quality](#)

Image Quality

Herzog Robert
Max-Planck Institute for Informatics,
Saarbrücken, Germany

Synonyms

[Image fidelity](#); [Image similarity](#)

Definition

Image quality is commonly characterized as the perceived image degradation with respect to an ideal undistorted image.

Overview

For many applications in research and industry, there is a constant need for quality assessment of images (e.g., computer graphics, image compression, camera manufactures, medical imaging).

Image quality cannot be formalized in general since it plays a different role depending on the application. For example, in *lossy compression* and *streaming*, the perceived quality-to-bit rate ratio with respect to a reference image needs to be maximized, whereas in computer vision, forensic, and medicine, the image quality is driven by *task performance* (i.e., how much semantic information is conveyed in the image). In photo-realistic image synthesis (e.g., 3D computer games, movies), image quality can be regarded as a measure of realism (photography versus rendered image). Moreover, in art and in particular in photography, image quality is strongly influenced by content and salience (see “► [Image Fidelity](#)”).

Apart from photographic lens distortions, image quality is largely affected by common factors such as noise, blurring, ringing (sharpening), aliasing, dynamic range compression, color quantization, image resolution, and application-specific artifacts (e.g., JPEG compression, rendering artifacts).

Image Quality Assessment

Image quality assessment can be divided into *subjective quality assessment* and *objective quality assessment*.

In subjective quality assessment, human observers independently and subjectively quantify the quality of an image by assigning it a single number that represents the quality score. The average is referred to as the *mean opinion score* (MOS). Objective quality assessment refers to the task of automatically evaluating the quality of images in a way human observers would do.

With today's new technologies, the spectrum of image quality degradations has widened including distortions coming from ► [3D stereoscopy](#), ► [high-dynamic-range \(HDR\) imaging](#), 3D image synthesis, and light-field imaging.

Image Quality Metrics

Automatic image quality metrics (IQM) have been developed for the task of objective image quality assessment. Although primarily designed for analyzing still image quality, IQMs have also been exploited for the task of video quality control. However, the human perception of video quality requires proper spatiotemporal modeling of the human visual system (HVS) including also the perception of temporal artifacts like flickering, which is covered by special *video quality metrics*. IQMs can be classified as *full-reference*, *reduced-reference*, and *no-reference* image quality metrics (the latter often referred to as *blind IQM*).

Full-Reference Quality Metrics

Full-reference image quality metrics aim at measuring the strength of general image distortions given two corresponding images, a distorted and a reference (undistorted) one. The given reference image simplifies the detection of universal distortions and their spatial localization in the image but, on the other hand, requires a relatively precise alignment of the two images.

When predicting a single image quality score similar to the subjective MOS, a two-stage approach is pursued. First, a local distortions measure is computed to obtain a quality map. Second,

a spatial pooling algorithm is employed to integrate the local distortions into a single number. Most IQM ignore the *chroma* component and focus only on *luminance* for which the human visual system is most sensitive to.

Full-reference quality metrics can be further divided into *physical* IQMs, *perceptual* IQMs, and *structural* and *information-theoretic* IQMs.

Physical IQMs have a clear physical meaning, are fast to compute, and are easy to differentiate making them preferred candidates for mathematical optimization techniques (energy minimization). Common physical IQMs are *mean-square error* (MSE), *peak signal-to-noise ratio* (PSNR), *mean absolute error*, and its generalization the *Minkowski error metric* using the l_p norm:

$$E_M = \left(\sum_{i=1}^N |x_i - y_i|^p \right)^{1/p},$$

where x_i, y_i are the values of the i th pixel in the distorted and reference image respectively and p is a positive real number.

Physical IQMs consider an image as a collection of N independent pixels x_i thereby ignoring any correlation among them. As shown in Fig. 1, the human-perceived error with respect to a reference image varies locally depending on the brightness, frequency, and correlation of the reference image content and the error signal.

Therefore, *perceptual image quality metrics* have been developed in order to better predict the perceived image quality (image similarity), which they achieve by simulating the early pathways of the human visual system (HVS) modeling effects such as *light adaptation*, *luminance masking* (Weber's law), *contrast sensitivity function* (CSF), and *contrast masking* (also known as *visual masking*). For this reason they are also referred to as *bottom-up* quality measures. Various perceptual models for image quality were proposed: the visual difference predictor (VDP) [5], the Lubin model [3], HDR-VDP [2], Watson's DCT model [4], JPEG 2000 [6], etc. Many of these models are relatively complex and computationally intensive but predict the human-

perceived distortions reliably near the visibility threshold (super-threshold). However, because they are mostly tuned to experimentally acquired visibility thresholds measured on simple stimuli, the prediction of strongly visible distortions (supra-threshold) is questionable and remains still an open research field.

To overcome the complexity of HVS-based metrics, structure-based metrics (e.g., SSIM [7]) have been derived that do not explicitly model the HVS but merely compute a relatively simple statistical measure of image similarity based on local pixel-value correlation and contrast.

Information-theoretic image quality metrics predict the perceived image quality using a statistical approach. They estimate how much information is shared between the reference and distorted image assuming the distorted image is transmitted over a noisy communication channel. Exploiting entropy-based methods known from signal coding and compression, the perceived quality of an image is computed from the *mutual information* of the reference and distorted image. A known information-theoretic quality metric is, for example, the *visual information fidelity* (VIF) index [8].

No-Reference Image Quality Metrics

No-reference (blind) image quality metrics pursue a top-down approach for predicting the quality of a single image. Thereby, the HVS is treated as a black box, and the quality of an image is predicted utilizing classification or regression techniques that operate on distortion-specific image features (e.g., gradient histograms, wavelet coefficients). No-reference IQMs require careful modeling of the image features and need many image examples for training a supervised learning algorithm (e.g., a support vector machine) [9]. Therefore, no-reference IQMs do not generalize well and are typically tuned to detect only one or a few specific image distortions (e.g., JPEG blockiness, noise, blur) where the success of the prediction primarily depends on the complexity of the type of distortion. Typical no-reference IQMs are based on features computed from ► [Natural Image Statistics](#).

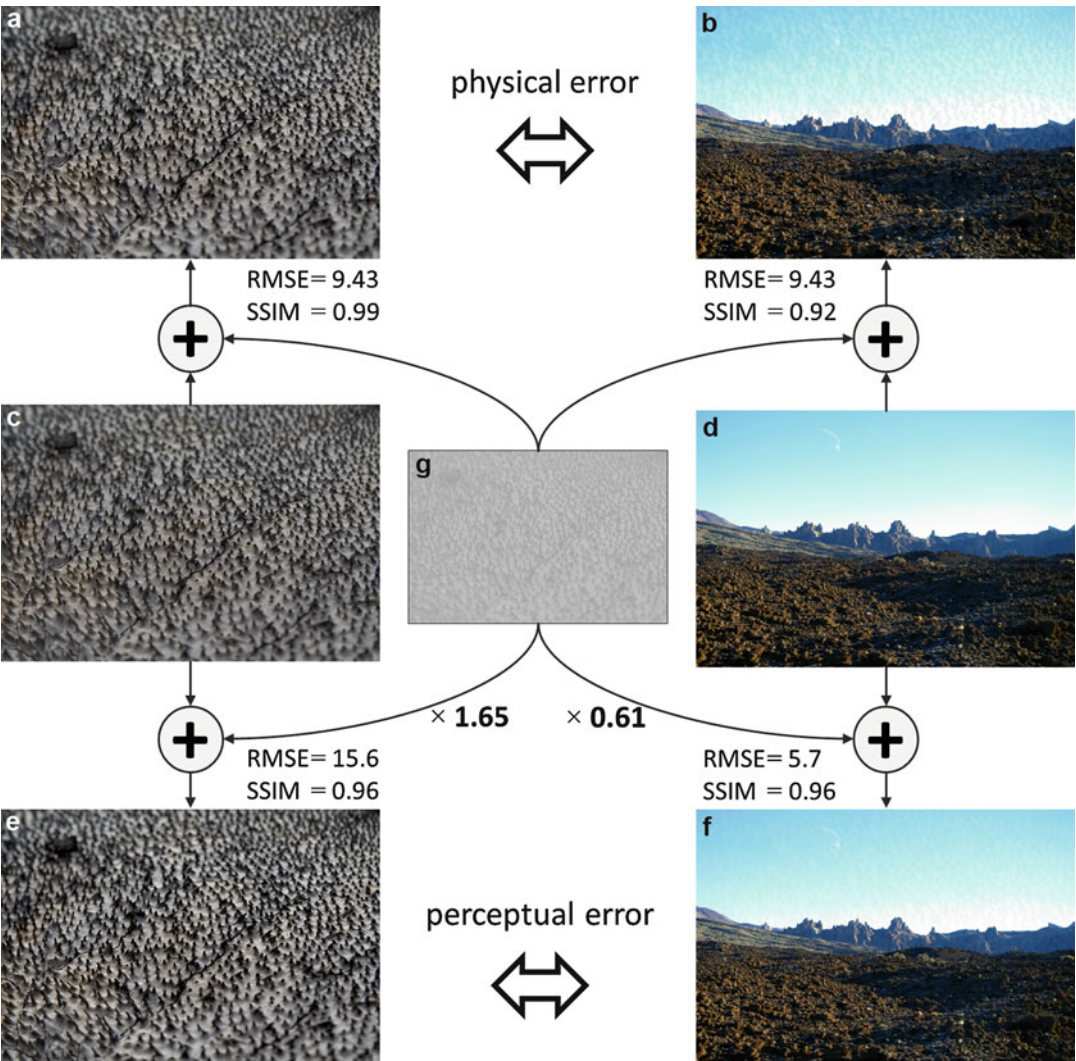


Image Quality, Fig. 1 Nonlinearity of the perception of image quality: images (a) and (b) have exactly the same root-mean-square error (RMSE) with respect to the original images (c) and (d), respectively. However, the distortions in image (b) are visually disturbing while they are almost invisible in image (a). This perceptual effect is referred to as *masking*. Note that the distortions in image (b) are more

salient in dark regions and regions with frequency content (e.g., sky) different from the error signal (g). Images (e) and (d) exhibit approximately the same perceived image quality as predicted by the structural similarity (SSIM) index metric [7]. Please note that the distortions perceived in the images depend on the display size of the images

Reduced-Reference Image Quality Metrics

In a streaming scenario, full-reference IQMs cannot directly be applied for predicting the image quality on the client side since only the server knows the reference and the distorted image. On the other hand, no-reference IQMs are less reliable and in general also more expensive to compute than full-reference IQM. Therefore, a new

paradigm for quality control has emerged, which is called reduced-reference IQM. Reduced-reference IQMs sparsely encode auxiliary information about the reference image that is streamed together with the lossily encoded image and enables the client to faithfully compute a quality score for the decoded image. The more data about the reference image is streamed, the better the

client can predict the image quality degradation. Thus, reduced-reference IQMs trade quality information for bit rate and can be regarded as a compromise between full-reference IQM and no-reference IQM.

Cross-References

- [Chromatic Contrast Sensitivity](#)
- [Image Fidelity](#)
- [Image Quality](#)
- [Light Distribution](#)
- [Luminance Meter](#)

References

1. Wang, Z., Bovik, A.C.: Modern Image Quality Assessment. Morgan & Claypool (2006)
2. Mantiuk, R., Kim K.J., Rempel A.G., Heidrich W.: HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.* **30**, 40:1–40:14 (2011)
3. Lubin, J.: A visual discrimination model for imaging system design and evaluation. In *Vision Models for Target Detection and Recognition*, 245–283, World Scientific (1995)
4. Watson, A.: DCT quantization matrices visually optimized for individual images. In *Human Vision, Visual Processing, and Digital Display IV*, **1913**, 202–216 (1993)
5. Daly, S.: The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In *Digital Image and Human Vision*, Cambridge, MA: MIT Press, 179–206 (1993)
6. Zeng W., Daly, S., Lei, S.: Visual optimization tools in JPEG 2000. In *IEEE International Conference on Image Processing*, 37–40 (2000)
7. Wang Z., Bovik, A.C., Sheik, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**, 600–612 (2004)
8. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. In *IEEE Transactions on Image Processing* **15**, 430–444 (2006)
9. Cortes, C., Vapnik V.: Support-vector network. In *Machine Learning* **20**, 273–297 (1995)

Image Similarity

- [Image Quality](#)

Impressionism

Georges Roque

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

Definition

Impressionism is considered as the first vanguard movement of Occidental art history; it opens modernity in art. Its starting point is a group of artists who had in common a rejection of academic art. They organized eight collective exhibitions between 1874 and 1886. Most of the artists only took part in some of these exhibitions, like Cézanne. The painters considered as more “typically” impressionist are above all Claude Monet, Pierre-Auguste Renoir, and Camille Pissarro.

Overview

The term “impressionist” was coined by a critic as a nickname in order to criticize the movement, as he was struck by a painting presented by Monet in the first 1874 collective exhibition and entitled *Impression, soleil levant* (*Impression, Sunrise*). The novelty of Impressionism is twofold. First, it has to do with the subject matter, often taken from everyday life (train stations, boulevards, scenes of leisure, etc.), corresponding so to Baudelaire’s conception of modernity. And second, it is technical: the way of rendering light, color, atmosphere, the nature of the brushstrokes, the innovative composition, and the impression of unfinished and incompleteness are among the main features of Impressionism.

Theory and Practice

According to the standard account of Impressionism (largely widespread by one of his best specialists, John Rewald), the Impressionists didn’t need any theory, since they trusted their eye. For sure, some of the statements made by Monet and

Renoir seem to reinforce such a conception. However, the issue of the relationship between Impressionism and color science is more complex. The first part of this paper is devoted to general features of Impressionism regarding color and the second one to the conception of color that can be found in Monet, Renoir, and Pissarro.

First of all, it must be noted that, unlike Neoimpressionist painters, the Impressionists didn't write "theoretical" essays, so that they have not conveyed a general idea of their principles. Furthermore, what they had in common was a rejection of academic art more than a clear doctrine. For this reason, the practice of the group members differs a lot from each other and depends more on personal skills than on shared beliefs, even though they had in common some general aims, like the ideal of modernity or the fact of preferring the rendering of ever-changing moments and atmospheres instead of permanent features. Yet, this doesn't mean that they had an "empirical" practice. Even though Monet liked to say, "I have already had a horror of theories" [1, p. 586], it should be recalled that such a statement doesn't mean that Monet rejected all theories but only dogmatic submission of practice to a theory prior to it, as did Bonnard and Matisse, too.

An often-quoted passage is usually given as an example of Monet's "empirical" practice: "When you go out to paint, try to forget what objects you have before you, a tree, a house, a field or whatever. Merely think here is a little square of blue, here a oblong of pink, here a streak of yellow, and paint it just as it looks to you, the exact color and shape, until it gives your own naïve impression of the scene before you" [2, p. 35]. This passage recalls Ruskin's famous conception of an "innocent eye": "The whole technical power of painting depends on our recovery of what may be called the *innocence of the eye*: that is to say, of a sort of childish perception of these flat stains of colour, merely as such, without consciousness of what they signify" [3, p. 27].

Monet – who praised a lot Ruskin's *Elements of Drawing* – coincided with Ruskin in the attempt to retrieving the "innocent" eye, that is to say, trying to paint things not as they are known but as they are seen. The problem, however, is that

there is no "purely" visual perception: the so-called innocent eye also depends on an implicit or explicit knowledge about the perceptual ► [appearance](#) of what is seen, knowledge that, in turn, depends on cultural background. Not surprisingly, the fact of seeing out in the world not "objects" but stains of colors corresponds to what scientists said at the time: perception of objects is the result of a psychophysiological process in the brain [4, p. 427]. Similar ideas were already present in the field of literature and art criticism. In his obituary of Delacroix, Baudelaire wrote that there are neither lines nor colors in nature, for they are created by man [5, p. 752]. It must be added that the vocabulary of artists is also borrowed from science. From this point of view, "impression" corresponds to the empiricist view according to which in visual perception it is the first impression not yet reworked by thought. The new meaning given by Helmholtz to the concepts of "impression" and "sensation" had been popularized in France through Taine's *De l'Intelligence* (1870); artists, however, often used them as if they were equivalent.

Optical Mixture

It is generally assumed that the Impressionists achieved the particular and luminous effect of their paintings through a particular technique, known as "division of the tone" and consisting in breaking down colors into their constituents. Duranty, one of the first critics of the movement, explained indeed, "Proceeding from intuition to intuition, they [the Impressionists] have little by little succeeded in breaking down sunlight into its rays, its elements and to reconstitute its unity by means of the general harmony of the iridescent color which they spread on their canvases [...]. The most learned physicist could find nothing to criticize in their analyses of light. . ." [2, p. 4].

It could be thought that critics were just trying to find a way to give a scientific legitimacy to a young movement highly criticized, arguing that their way of painting was "scientific," which in fact is not true [6]. However, besides similar comments made by other critics, there is a more

precise text based on conversations with Pissarro and that the painter never disowned:

About that time [1870–1871], M. Pissarro, after a holiday at Plette in Britain, went to London with M. Claude Monet. During their ten months stay in this city, they sent canvases to the National Academy, which then condemned the painters' audacity. Nevertheless, on British soil, their eyes had been educated by Turner, who had for a long time restricted himself solely to the colors of the prism. In their study of his work, they found confirmation of theories already discussed in private and realized in individual essays that had not yet been publicly shown: *the law of complementary colors and its natural end, the division of the tone*.

This stay in England hastened the Impressionist's evolution. Back in Paris, MM. Pissarro and Monet made themselves the exegetes of the new technique. Quickly, their friends, prepared by previous attempts, recognized the superiority of the retinal color mixture over the obviously darker mixture that occurs on the palette. The optical reconstitution of the complementary colors divided on the canvas finally gave them those blond lights so patiently sought.

Impressionism, stemming from precise theories, soon emerged with the brilliance of its luminous and vibrating harmony. [7]

This text shows at least that Pissarro and Monet attempted something considerably beyond “empirical” practice of painting: they needed theoretical explanations grounded on the scientific knowledge available in their time in order to confirm their practice. Indeed, the “law of complementary colors” refers to the theory of the French chemist Michel-Eugène Chevreul, who was very influential at the time and is directly or indirectly the main source of the Impressionist color theory [8]. According to Charles Blanc, a very important art historian (he was also director of the Paris École des Beaux-Arts), whose book *Grammar of Painting and Engraving* was avidly read by the artists at the time, “it is the mutual exaltation of juxtaposed complementary colors that Mr. Chevreul called ‘the law of simultaneous contrast of colors’” [9, p. 562]. If two ► [complementary colors](#) are juxtaposed, indeed, they enhance each other, and Impressionist painters were eager to use bright colors. For this reason, they took seriously the ► [harmony](#) of complementary colors. They also thought that they would obtain a more luminous result when juxtaposing dots of

colors (and particularly dots of complementary colors) that were supposed to fuse in the eye; for this reason, the “division” of the tone is directly related to the law of complementary contrast. What has been called “indigomania,” due to a pervasive use of violet and blue in Impressionist canvases, has a lot to do with this issue [8, pp. 34–36]. Georges Guérout (one of the translators of Helmholtz into French) gave indeed an explanation for the use of violet and blue, in particular for shadows: “The *plein air* school is right to put blue or violet in the shadows. For yellow, whose complement is violet, dominates much in the sun, and, by contrast, violet dominates in the shadow” [10, p. 176].

Artificial Pigments

Besides harmony of ► [complementary colors](#), another general feature of Impressionism is the use of artificial pigments that had been recently introduced in the color market [11, pp. 51–67, 12, pp. 136–155]. Furthermore, they liked to use them pure, as they come from the tube, i.e., without mixing them. Such a use has been deeply criticized at the time, as pure hues were not considered as “true” artistic colors. According to this dated conception, colors, in order to be accepted as artistic, have to be prepared and mixed. The reason for the Impressionist practice regarding pigments is related to their modernism: painters were proud to use newly created synthetic pigments instead of the old ones; furthermore, using them pure was also another way of being modern, as it entailed a strong critique of the traditional view according to which painters had to prepare their own colors.

Renoir, Monet, and Pissarro

After this general overview, a look at some painters in particular could confirm, despite what is generally assumed, that they were interested in color theory and applied it to their works. From this point of view, artists' statements don't have to be taken literally. Renoir once explained that he

painted “like a child,” and he added, “I don’t have neither rules nor methods” [13, p. 13]. However, a closer look at his writing shows that he was aware of the fact that the brightness of a hue depends on the other colors contiguous to it. When Julie Manet asked him advices as how to improve a still life on a white tablecloth, Renoir explained to her that “one has to give white its intensity thanks to the value of what surrounds it, and not by adding more white to it” [13, p. 223]. Now, this is exactly one of the six advices Chevreul gave painters: white placed beside a color heightens its tone [14, § 336].

Exactly the same warning can be made about Monet’s statements. And in his case, too, relevant information confirms his interest for color theory. In an interview given in 1888 to an English newspaper, he explained, “Color owes its brightness to force of contrast rather than to its inherent qualities [...] ► [primary colors](#) look brightest when they are brought into contrast with their complementaries” [11, p. 88]. This statement explains his well-known taste for the field of poppies he painted very often, as they gave him an excellent opportunity to juxtapose green surfaces with patches of red, enhancing so the two ► [complementary colors](#). Another source of information is the visit the painter Louis Anquetin made to Monet in order to inquire into his conception of color. According to Émile Bernard’s account of this visit, Anquetin was disappointed, realizing that Monet “only knew a very little part of what he was coming to ask him, contenting himself with the theory of the complementary colors, applied with the palette of the seven prismatic colors” [15, p. 112]. This information confirms that Monet used the harmony of complementary colors.

Now, what about Pissarro? According to Geffroy, he “was always passionately and learnedly interested in the theories of lights and colors, fervent in the search for processes, always ready for research and realization” [16, p. 263]. Not surprisingly, his contribution to color is also remarkable, in particular concerning frames. As a critic recalled, “At the 1877 Exhibition, M. Pissarro, applying in its rigorous logic the law of the complementary colors, set his canvases into white frames that, without influencing the

colors, left the tones with their exact values” [7]. Indeed, Chevreul had actually drawn attention to the influence of the colors of the frame on the colors of the framed work and had argued against the use of gilded frames [14, §565]. Since Chevreul was probably the only scientist to have paid attention to this problem at the time, Pissarro could not have learned it from any other source. Pissarro liked the idea and used white frames for his paintings, after convincing his dealer to do so. In 1881, for the Sixth Impressionist Exhibition, Pissarro innovated again: he tinted the stretchers with the complementaries of the color dominating in the painting, which was another original way of enhancing the colors of his paintings.

To conclude, even though in some statements Impressionist painters said they had a purely empirical conception of color and were not interested in theory, there is enough evidence showing that they were using recipes allowing them to enhance their colors. For this aim, Chevreul’s law of simultaneous contrast, as understood by Blanc and others, provided excellent tools and advices. Closer technical analysis of Impressionist paintings confirm their frequent use of juxtaposed ► [complementary colors](#) [11, 12].

Cross-References

- [Anchoring Theory of Lightness](#)
- [Appearance](#)
- [Chevreul, Michel-Eugène](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Complementary Colors](#)
- [Helmholtz, Hermann Ludwig von](#)
- [Neo-impressionism](#)
- [Primary Colors](#)

References

1. Monet, C.: Letter to Charteris (June 21, 1926), quoted by Rewald, J.: *The History of Impressionism*, 4th edn. Museum of Modern Art, New York (1973)
2. Nochlin, L.: *Impressionism and Post-Impressionism 1874–1904. Sources and Documents*. Prentice Hall, Englewood Cliffs (1966)

3. Ruskin, J.: *The Elements of Drawing*. Dover, New York (1971)
4. von Helmholtz, H.: *Handbuch der physiologischen Optik*. Leopold Voss, Leipzig (1867)
5. Baudelaire, C.: *Oeuvres complètes II*. Gallimard, Paris (1976)
6. Carson Webster, J.: The technique of impressionism: a reappraisal. *College Art J.* **IV**(1), 3–22 (1944)
7. Lecomte, G.: Camille Pissarro. *Les Hommes d'aujourd'hui* **VIII**(366), (1890)
8. Roque, G.: Chevreul and impressionism: a reappraisal. *Art Bull.* **LXXXVIII**(1), 25–39 (1996)
9. Blanc, C.: *Grammaire des arts du dessin*. Renouard, Paris (1880). The chapter on color has been translated into English in Taylor, J.C. (ed.): *Nineteenth-Century Theories of Art*. University of California Press, Berkeley (1989)
10. Guérault, G.: *Formes, couleurs et mouvements*. *Gazette des Beaux-Arts* **XXV**, 165–179 (1882)
11. Bomford, D., et al.: *Art in the Making. Impressionism*. The National Gallery, London (1990)
12. Callen, A.: *The Art of Impressionism. Painting Technique & the Making of Modernity*. Yale University Press, New Haven/London (2000)
13. Renoir, P.-A.: *Écrits, entretiens et lettres sur l'art*. Éditions de l'Amateur, Paris (2002)
14. Chevreul, M.-E.: *De la loi du contraste simultané des couleurs...* Pitois-Levrault, Paris (1839). The latest English translation is: *The Principles of Harmony and Contrast of Colors and their Applications to the Arts* (1855). Kessinger Publishing LLC, Whitefish (2009)
15. Bernard, E.: Louis Anquetin. In: *Gazette des Beaux-Arts* 76th Year, vol. XI(part 2), pp. 108–121 (1934)
16. Geffroy, G.: Monet, sa vie, son oeuvre. *Macula*, Paris (1980) (1st edn. 1924)

Inappropriate Use of color

► Color Pollution

Incandescence

Wout van Bommel
Nuenen, The Netherlands

Definition

Emission of optical radiation by the process of thermal radiation.

Thermal radiation is electromagnetic radiation originating in the thermal motions of particles of matter (atoms, molecules, ions, etc.). All matter above the temperature of absolute zero emits thermal radiation. Objects around room temperature emit mainly infrared radiation. Objects at temperatures higher than some 800 K also emit electromagnetic radiation in the visible area.

Thermal Light Radiators

Sun that generates light by its hot plasma with a temperature of around 6,000 K

Glowing coal and red hot horseshoe with a temperature of around 1,000 K

Torches, oil lamps, candles, and gas lamps with temperatures of around 2,000 K

Carbon-arc lamps with temperatures of up to 5,000 K

Old-fashioned chemical flash bulbs with temperatures of around 4,000 K

Incandescent lamps with temperatures of 2,700 K

Halogen incandescent lamps with temperatures of 2,900–3,200 K

Black body radiator

Black Body Radiator

A black body radiator is an ideal thermal radiator that absorbs completely all incident radiation, whatever the wavelength, the direction of incidence, or the polarization. It is also referred to as Planckian radiator. The research on and description of the behavior of thermal radiators is based on this ideal thermal radiator. Black body radiation is also the basis for the correlated color temperature of gas discharge and solid state light sources.

The thermal radiation of a black body radiator depends solely on the temperature of the material (T) and is completely described by three laws: the law of Planck, the law of Stefan-Boltzmann, and the displacement law of Wien. The law of Planck defines the continuous spectrum of the black body radiator. The law of Stefan-Boltzmann states that the total radiant flux increases proportional to T^4 .

Wien's displacement law defines how the peak of the spectrum shifts to shorter wavelengths with increase of T (i.e., from the infrared region of the spectrum via the red part of the visible spectrum to the blue part of the visible spectrum).

Incandescent material used for actual lamps such as tungsten filaments of incandescent and halogen lamps behaves similar as the black body radiator but with some important differences.

Characteristics of Commercial Incandescent Lamps

The efficacy of incandescent lamps is with some 5–30 lm/W (dependent on type and wattage) low compared to that of gas discharge lamps and solid state LED light sources with values between 50 and 200 lm/W. The average lifetime with values between 1,000 and 3,000 h is short relative to the economic lifetime of gas discharge lamps and LED lamps with values varying, with type, from some 10,000–10,000 h. The temperature of the incandescent filament is limited to some 3,200 K because at higher temperatures the material will evaporate quickly or even start to burn. The consequence is that the continuous spectrum of both the incandescent and halogen lamps has its peak in the red part of the spectrum. The peak with the halogen type is slightly more to shorter wavelengths. The color of the light is warm-white (color temperature between 2,700 and 3,200 K). Since the spectrum is continuous, the color rendering index is 100.

Cross-References

- [Carbon arc lamp](#)
- [Combustion lamp](#)
- [Halogen lamp](#)
- [Incandescent lamp](#)

Incandescent Bulb

- [Incandescent Lamp](#)

Incandescent Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Incandescent bulb](#); [Tungsten filament lamp](#)

Definition

Lamp that produces light as a result of an electrical current through a tungsten wire, contained in a transparent bulb, which heats the metal to incandescence.

Incandescent Lamps

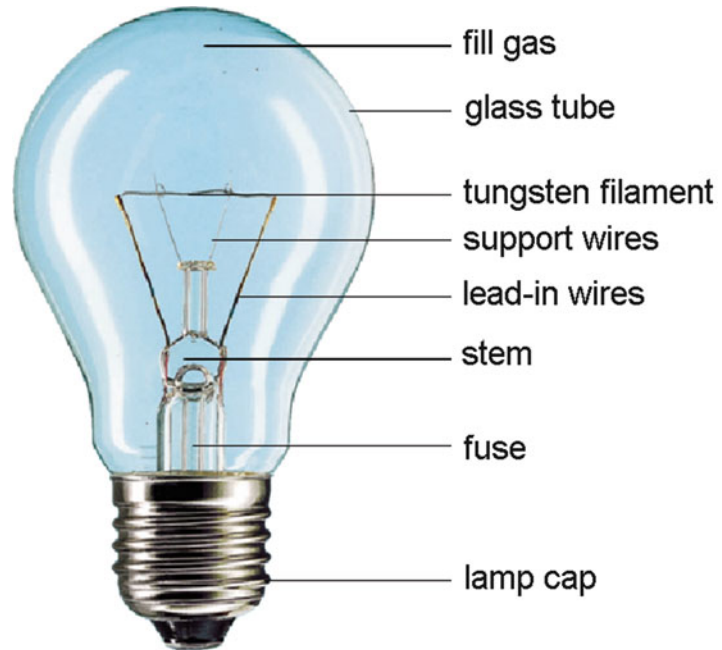
There exists a wide variety of incandescent lamps, ranging from the most common types (usually referred to as general lighting service, GLS, lamps) to reflector lamps, colored lamps, and halogen incandescent lamps [1, 2]. Given the special operating principle and construction of halogen incandescent lamps, these will be dealt with separately under “halogen, halogen lamp.” The most important application area of normal incandescent lamps is home lighting. Because of the low efficacy and short lifetime of incandescent lamps, the more efficient and longer-life gas discharge and solid-state lamp alternatives have become increasingly more important for home lighting. For reasons of sustainability, governments in some parts of the world have taken the decision (e.g., Australia and the European Union), or are considering taking the decision, to ban GLS incandescent lamps.

Working Principle

The operating principle of the incandescent lamp is extremely simple. An electric current is passed through a thin wire of comparatively high resistance so as to heat this to incandescence. The wire is usually heated to a temperature of between

Incandescent Lamp,

Fig. 1 The main parts of an incandescent lamp [3]



2700 K and 2800 K, at which temperature it emits warm-white light. The wire is placed inside a glass bulb, which is either at a vacuum or contains an inert gas.

Materials and Construction

The main parts of an incandescent lamp are (Fig. 1):

- Glass bulb
- Filament
- Filament support consisting of a glass stem, lead-in wires (also serving as a fuse), and support wires
- Bulb fill gas
- Lamp cap

Filament

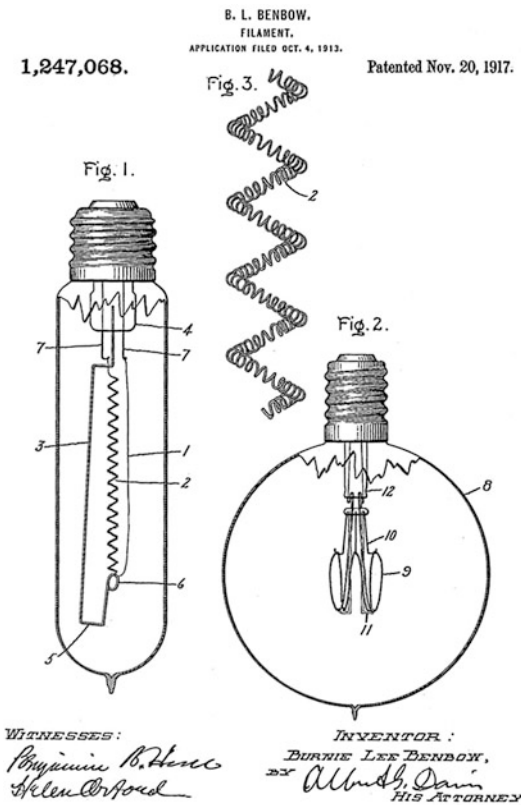
With very few exceptions, filaments for incandescent lamps are made of tungsten, a metal that has the advantage of a high melting point combined with a relatively low vapor pressure, even at very high temperatures. This low vapor pressure means a low evaporation rate of the tungsten material and thus a relatively long life of the coil. The tungsten wire is generally wound into a single or double coil for reasons of compactness and heat

conservation and concentration around the coil. The double coil (Figs. 2 and 3), which conserves and concentrates the heat better than does a single coil, was patented in 1917. It was the last fundamental improvement of incandescent lamps as far as operating principle is concerned.

Bulb

The bulbs of GLS lamps are made of soda-lime glass, the most common and cheapest type of glass available. For lamps that must withstand high temperatures or temperature shocks, more resistant glasses are used. The bulb of an incandescent lamp can be of different sizes and may take different shapes, depending on the application. Also depending on the application, the bulb may undergo various treatments. The most commonly employed treatments are:

- Frosting: this is done by etching the inside of the bulb. The etching of the surface results in a satin finish and moderate diffusion of the light with hardly any reduction in transmittance.
- Opalizing: this is achieved by coating the inside of the bulb with a special powder. It provides better diffusion of the light at the cost of slightly greater light absorption.



Incandescent Lamp, Fig. 2 Patent of 1917 of the double-coiled tungsten wire

- Mirror coating: reflector lamps receive an internal mirror coating. Silver mirrors are produced by evaporation of aluminum under vacuum.

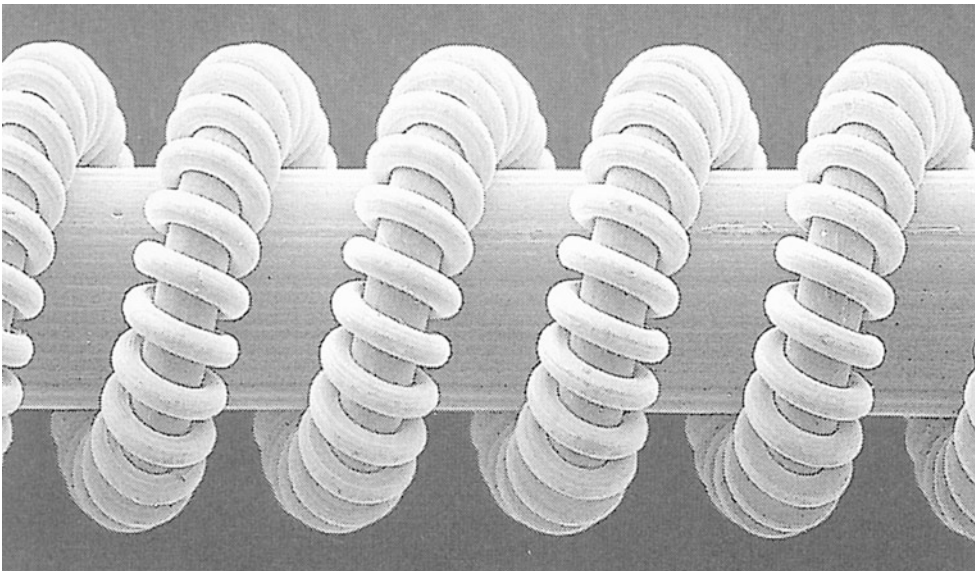
Fill Gas

Two types of incandescent lamp are available on the market; vacuum lamps and those with a fill gas. The great majority of incandescent lamps contain a fill gas, the main purpose of which is to reduce the evaporation rate of the filament and thus increase the lamp life. The fill gas is a mixture of argon and nitrogen. In incandescent car lamps, krypton is usually employed (it is more expensive, but results in a higher luminous efficacy).

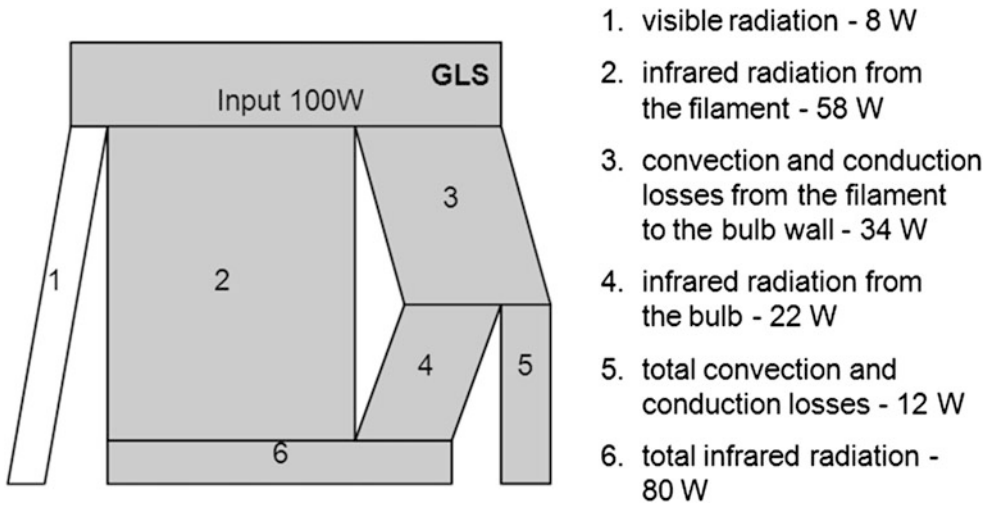
Even small quantities of oxygen or water vapor could shorten the lifetime of the tungsten wire through corrosion. To remove even the slightest traces of these, a so-called getter is added. The getter absorbs oxygen and water.

Lamp Cap

The lamp cap connects the lamp to the power-supply socket. There are two different standardized cap types: the screw cap (also called the Edison cap) and the bayonet cap. Both are employed in various sizes:



Incandescent Lamp, Fig. 3 Winding of the tungsten wire into a double coil around a core, which is removed later on in the manufacturing process [3]



Incandescent Lamp, Fig. 4 Energy balance of a 100 W GLS lamp [3]

- Edison screw type: IEC designation – E10, E14, E27, E40. In the USA names rather than IEC designations are used: Candelabra (E12), Intermediate (E17), Medium (E26), and Mogul (E39).
- Bayonet type: IEC designation B15, B22.

The number indicates the diameter of the lamp cap in millimeters.

Properties

Energy Balance

Figure 4 shows the energy balance of a 100 W GLS lamp. It shows that only approximately 8 % of the input power is emitted in the form of visible radiation. The rest is lost as heat (by conduction, convection, and infrared radiation).

Luminous Efficacy

The luminous efficacy of a standard GLS lamp varies, depending on its wattage, between 8 and 15 lm/W. For a standard 75 W lamp, the efficacy is around 12 lm/W. The lower the wattage, the lower the efficacy.

Compared with gas discharge and solid-state light sources, the efficacy of incandescent lamps is very low indeed. For reasons of sustainability, governments in some parts of the world (e.g., the

European Union) have taken the decision, or are considering taking the decision, to ban GLS incandescent lamps.

Lumen-Package Range

Common types of GLS lamps are available in the range from some 100 to 1500 lumen (corresponding wattage range approx. 15–200 W). There are, of course, many special lamps with both lower and higher lumen packages: think, for example, of bicycle and torch lamps on the one hand and light-house lamps on the other.

Color Characteristics

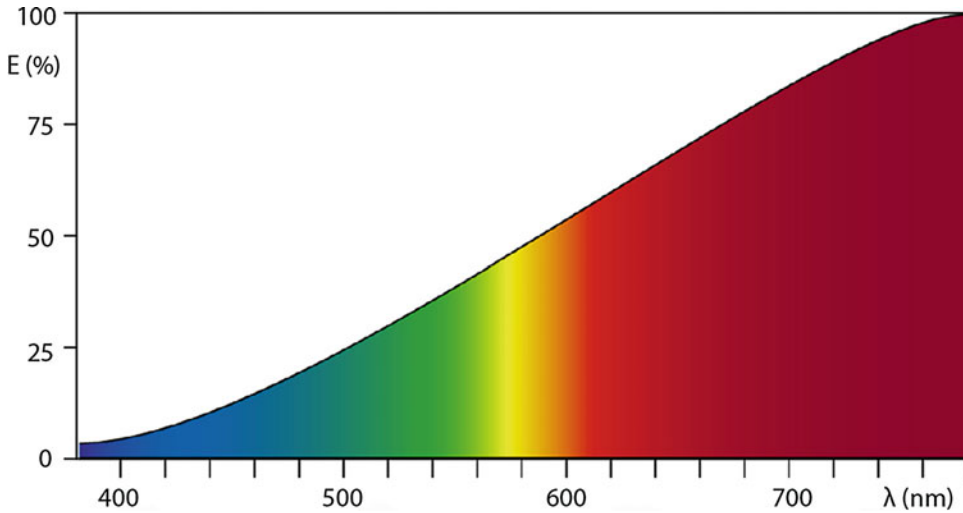
Incandescent lamps have a continuous spectrum radiating more energy at longer wavelengths (red) than at shorter wavelengths (blue) (Fig. 5).

Their color rendering index is 100, and standard GLS lamps have a color temperature of between 2700 K and 2800 K (warm-white light).

Colored opal lamps for party or festive lighting are available, as are lamps producing a special tint of “white” light (e.g., flame, terracotta, beige).

Lamp Life

Lamp manufacturers can, in principle, balance the life of an incandescent lamp against its luminous efficacy: higher lifetime, lower efficacy. However,



Incandescent Lamp, Fig. 5 Relative spectral energy distribution of an incandescent lamp [3]

by international agreement, the lamp industry has standardized the average rated life of standard GLS lamp at 1000 h (750 h in the USA). Nevertheless, for special applications, lamps are available where the balance between lifetime and efficacy is different. Examples are photo and film-studio lamps and torch bulbs on the one hand with their shorter life but higher efficacy (and higher color temperature) and beacon lamps on the other with their longer life but lower efficacy.

Lamp-Lumen Depreciation

An important reason why the light output of an incandescent lamp decreases with time is that tungsten evaporates from the filament and settles on the bulb wall, blackening it. At the end of the rated average life (1000 h), the lumen depreciation can amount to as much as 15–25 %.

Run-up and Re-ignition

Incandescent lamps give their full light output immediately after being switched on, and after being switched off, they re-ignite again immediately.

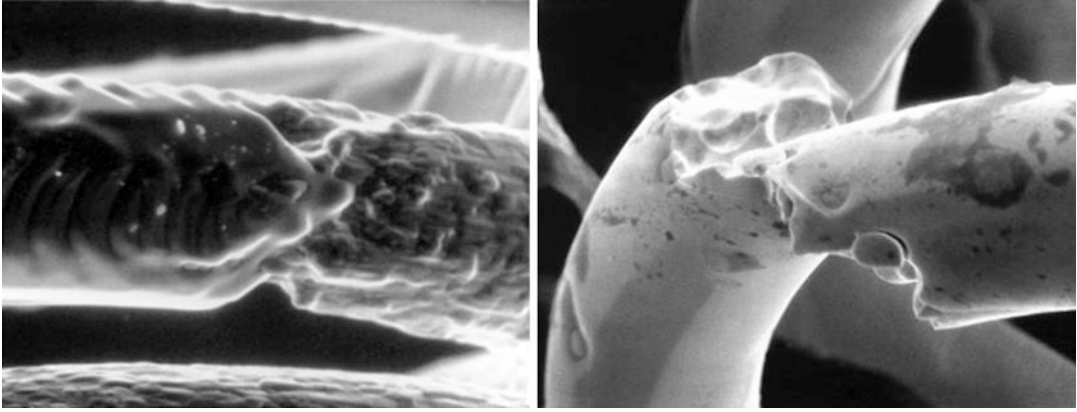
Switching

Frequent switching is not normally detrimental to lamp life, but when the filament has become

critically thin through age, the mechanical strain caused by the rapid temperature change as a result of switching will be sufficient to cause its breakdown (Fig. 6). This is the reason why incandescent lamps approaching the end of their life usually fail the moment they are switched on or off.

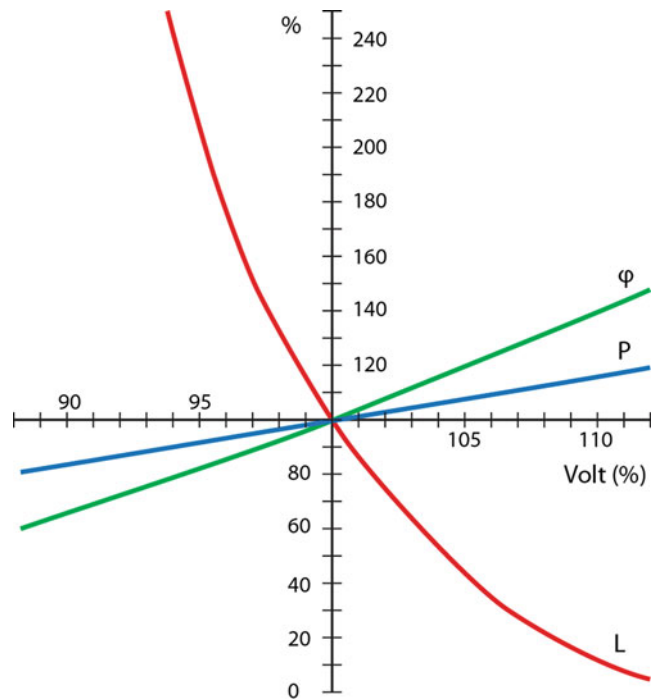
Dimming

Normal incandescent lamps can be dimmed without restriction. Dimming by reducing the supply voltage to the lamp with an adjustable resistance in series with the lamp is technically easy but does not result in energy saving. A relatively expensive adjustable transformer (potentiometer) results in energy saving, but not very much. Today the usual method to dim incandescent lamps is by thyristor dimmers that cut part of the AC current during each of the 50 (Europe) or 60 (USA) cycles. During the time that the current is cut, no power is dissipated, so that indeed energy is saved. These systems are called phase-cutting systems. They are small enough to fit into a wall switch box or cord switch. A dimmed incandescent lamp will have a lower filament temperature, which results in a longer life and lower color temperature (warm-colored light). The lamp efficacy while dimming decreases, which means that the percentage of power saved is less than the percentage reduction in light output.



Incandescent Lamp, Fig. 6 Thin spot in the filament eventually resulting in breakage of the filament [3]

Incandescent Lamp, Fig. 7 Diagram indicating the effects of mains voltage variation on lifetime (L), luminous flux (ϕ), and dissipated power (P)



Mains Voltage Variations

An overvoltage of even a few percent results in a drastically reduced lamp life. For example, a permanent overvoltage of 5 % reduces lamp life by 50 % (Fig. 7).

Product Range

The product range of incandescent lamps comprises not only general lighting service lamps

(in clear, frosted, opal, and colored versions) but also reflector lamps, tubular lamps, decorative lamps, beacon lamps, vehicle lamps (bicycles, cars, trains), signal lamps, and studio and theater lamps. Figures 8 and 9 show some examples of these lamps. All these lamps have blown-glass bulbs (blown from a single “blob” of glass), with the exception of the pressed-glass reflector lamp (PAR). The latter is molded in two pieces, which



Incandescent Lamp, Fig. 8 Examples of general lighting service (GLS) incandescent lamp types



Incandescent Lamp, Fig. 9 Examples of refractor/reflector incandescent lamps

are sealed together – they are also sometimes called sealed-beam lamps. One advantage of this construction is that the mirror in the rear and the refractor on the front can be accurately shaped to increase the beam quality and efficiency. Another advantage is its high mechanical and thermal strength, allowing the lamp to be used outdoors without a protecting luminaire.

References

1. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
2. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook, 10th edn, Illuminating Engineering Society of North America, New York (2011)
3. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Course Book. Philips Lighting, Eindhoven (2012)

Cross-References

► [Incandescence](#)

Indirect Illumination

► [Global Illumination](#)

Individual Differences

► Environmental Influences on Color Vision

Indoor Lighting

► Interior Lighting

Induction Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

Electrodeless low-pressure mercury lamps

Definition

Lamps that produce light as a result of an electrical discharge, generated by an induction coil, in a low-pressure mercury vapor that is contained in a transparent tube or vessel whose inside is coated with fluorescent powder that converts the ultraviolet part of the emitted radiation from the discharge in visible light.

Induction Lamps

Induction lamps, like fluorescent lamps, belong to the family of low-pressure mercury gas discharge lamps. Unlike other discharge lamps they have no electrodes, which is why they are also called “electrodeless lamps” [2, 3]. The consequence of

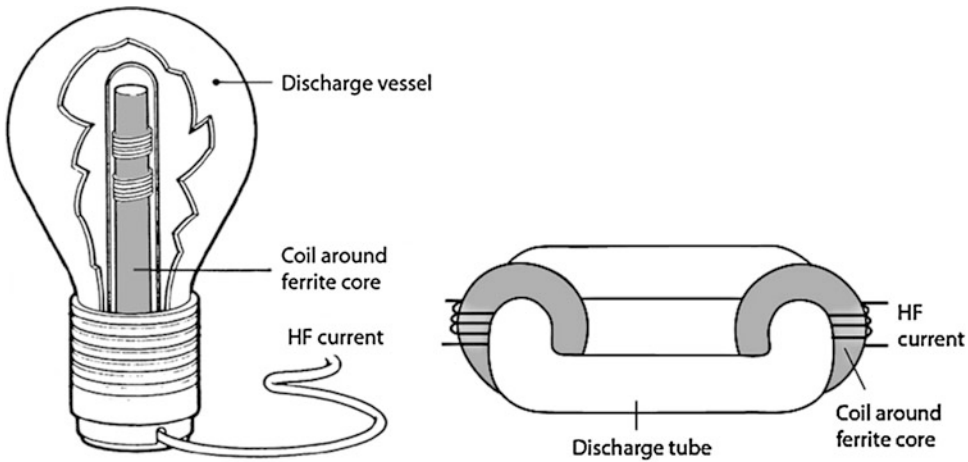
having no electrodes is a very long economic life of around 60,000–75,000 h. This long life is also the main feature of induction lamps. They find their application in situations where lamp replacement is near impossible or very expensive.

Working Principle

In an induction lamp, the free-running electrons needed for the gas discharge are obtained by winding an induction coil around a ferrite core placed in or around a discharge vessel (Fig. 1). The induction coil is connected to a high-frequency power source and acts like a primary winding in a transformer. In a transformer, an alternating current in a primary winding around an iron core creates an alternating magnetic field that in turn initiates an alternating current in a secondary coil wound around the primary coil. In the discharge vessel or tube, the mercury gas surrounding the ferrite core acts as the secondary coil (since mercury is a metal). The secondary current initiated in the mercury consists of free-running electrons around the ferrite core (Fig. 2). As in a normal fluorescent gas discharge, these free-running electrons ionize and excite other mercury atoms, which results in the emission of the same radiation as in a tubular fluorescent lamp. Since the vessel’s interior is coated with the same fluorescent powder as in normal fluorescent lamps, the same type of light is obtained. The high frequency of 2.65 MHz is generated by an electronic circuit.

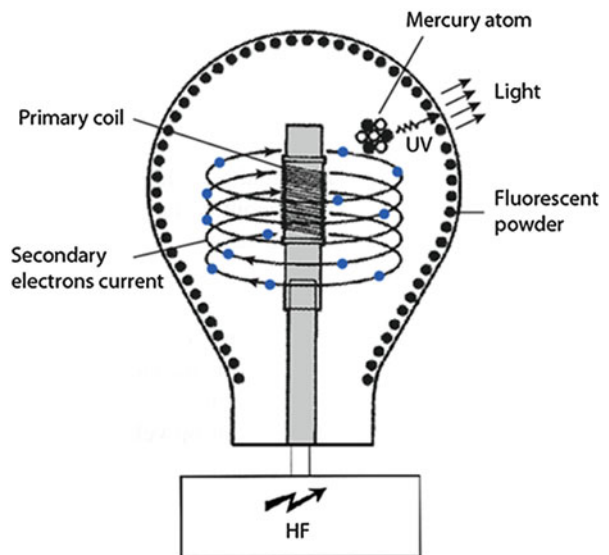
Materials and Construction

The discharge vessel or tube is made of the same type of glass as normal fluorescent tubes and has, in the case of the internal coil, a cylindrical glass cavity in which the ferrite core with coil (also called “antenna”) is positioned (Fig. 3). Special precautions are taken in the design of the lamp to limit excessive heat buildup around the antenna. The high-frequency power generator is connected to the antenna with a coax cable. The cable forms part of the electronic circuit and is therefore an integral part of the system (Fig. 4). The version with the coil with iron core around the outside of the discharge tube has a different, planar shape than the pear-shaped version.



Induction Lamp, Fig. 1 The principal parts of an induction lamp. *Left*, induction coil inside the discharge vessel; *right*, induction coils around outer part of discharge tube

Induction Lamp, Fig. 2 Principle of an induction lamp [1]



Properties

Energy Balance

Around 17 % of the input power of an induction lamp is radiated as visible light. The remaining part is lost as heat in the power generator, in the antenna, and in the discharge.

System Luminous Efficacy

The luminous efficacy of induction lamps is smaller than that of normal fluorescent lamps. It

varies, depending on the wattage, between roughly 65 and 75 lm/W.

Lumen-Package Range

Induction lamps are available in a lumen-package range from some 3,500 to 12,000 lm (corresponding wattage range 55–165 W).

Color Characteristics

The fluorescent coating is of the same composition as that in normal fluorescent tubes, so the

color characteristics are also the same. See chapter “► [Tubular and Compact Fluorescent Lamp](#).”

Lamp Life

The lamp life of most induction lamps is extremely long. Based on a mortality rate of 20 %, induction lamps have a lifetime of between

60,000 and 75,000 h. This is almost 7 years of continuous operation, day and night.

Lamp Price

Induction lamps are expensive. The high lamp price has to be balanced against the extremely long lifetime of these lamps. This means that where lamp replacement is difficult, very expensive, or even impossible, the balance will certainly be in favor of the longer life of these lamps.

Lamp-Lumen Depreciation

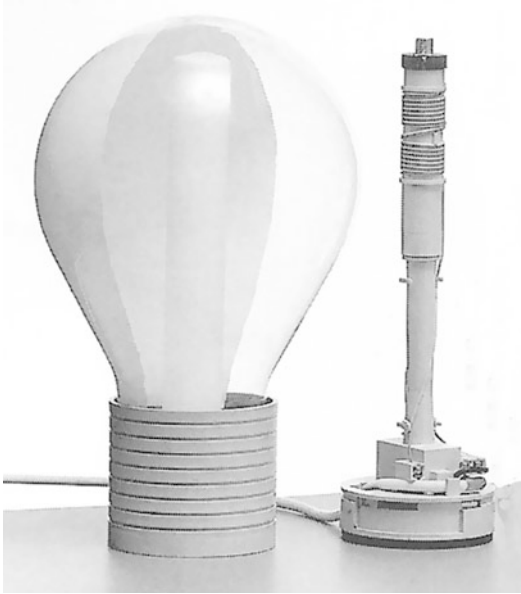
Lumen depreciation in induction lamps is determined by a decrease in the activity of the fluorescent powder. At 60,000 h the depreciation is around 25 %, which is why this lifetime figure is usually quoted. The actual life of induction lamps is usually much longer.

Run-Up and Reignition

A high-voltage ignition pulse produced by the HF generator ignites the lamp within 5 s, after which it emits its full light output within 1 min. Thanks to the high-voltage ignition pulse, hot reignition of the lamp is no problem.

Dimming

Most standard versions of induction lamps are not dimmable. However, some special versions, which are dimmable to 50 %, are available.



Induction Lamp, Fig. 3 Photograph of induction lamp vessel and antenna. This lamp is only partly covered with fluorescent powder in order to show the cylindrical cavity in which the coil is positioned [1]



Induction Lamp, Fig. 4 Pear-shaped induction lamp with internal coil (*left*) and plane-shaped lamp with external coils

Ambient-Temperature Sensitivity

Amalgam is used instead of pure mercury to keep the sensitivity to ambient temperatures within practical limits. Lamp position sometimes plays a role in this as well. It is therefore important to follow the recommendations from the lamp manufacturer with respect to the luminaire design.

Product Range

As has been shown, induction lamps are available in versions with internal and external antenna. The corresponding shapes are pear and plane shaped. They are normally available in color-rendering types 800 and 900, with color temperatures of 3,000 and 4,000 K.

Cross-References

- ▶ [Phosphors and Fluorescent Powders](#)
- ▶ [Tubular and Compact Fluorescent Lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Course Book. Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: *Lamps and Lighting*, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *IES Handbook*. 10th edn. Illuminating Engineering Society of North America, IESNA, New York (2011)

Infant Color Categories

Anna Franklin

Psychology, University of Surrey, Surrey, UK

Synonyms

[Prelinguistic color categorization/categorical perception](#)

Definition

A response to color in infancy that indicates that infants can divide the continuum of color into discrete groups.

Key Principles and Concepts

Although humans can discriminate millions of colors, language typically refers to color using a limited number of discrete categories (e.g., *red*, *green*, *blue*). Color categories are not only present in language (color terms), they are also present in “thought.” For example, at least under some circumstances, adults’ perceptual or cognitive judgments of color (e.g., color memory, search, or similarity judgments) are affected by whether colors come from the same or different linguistic color categories [see Hanley]. There has been considerable debate on the origin of color categories and on the extent to which color categories are arbitrarily constructed through language (see ref. [1] for review). To establish whether there is a nonlinguistic route to color categorization, prelinguistic infants’ response to color has been investigated. This has provided converging evidence that infants’ perceptual or cognitive judgments of color can also be affected by whether colors come from the same or different adult color categories. This suggests that color categories are present even in infancy and that even infants, on some level, can parse the continuum of color into discrete groups.

Methods and Examples

Investigations of adults’ categorical response to color have used a range of experimental tasks that provide behavioral or electrophysiological measures of color memory, color search, or chromatic change detection. In these studies, chromatic differences for same- and different-category color pairs are equated, and an influence of the categorical relationship of colors on task performance is assessed. An equivalent approach has been taken to investigate infant color categories. Many of the studies of infant color categories have used the well-established method of *habituation* to assess the influence of categories on infants’ response to chromatic novelty. This method involves the repeated presentation of a stimulus (or set of stimuli) until infants’ looking time at the stimulus declines (habituation). Then, during a test phase,

novel and original stimuli are presented and looking time is recorded. Greater looking to the novel stimulus relative to the original stimulus during the test phase (novelty preference) indicates that the infant is, on some level, treating novel and original stimuli as different. A lack of longer looking at the novel stimulus relative to the original indicates that the infant is, on some level, treating novel and original stimuli as equivalent. The degree of novelty preference can therefore be compared when novel and original stimuli from the same or different categories. Greater different-than same-category novelty preference when chromatic differences are equated would indicate an influence of categories on infants' perception or cognition. In addition, categorization can be broadly defined as the treatment of discriminably different stimuli as equivalent (e.g., [2]). Therefore, a lack of novelty preference when infants can discriminate between novel and familiar stimuli becomes a clear marker of categorization. This method has provided overwhelming evidence for infant categories, for a wide range of categories such as phonemes, facial expressions, animals, spatial locations, and orientation (e.g., see [2]).

For color, multiple studies have found a categorical response in infants at 4 months using the habituation method (e.g., [3–6]). Bornstein et al. [3] tested infants at 4 months using monochromatic lights, with the difference between the original and novel stimulus equated in wavelength. Infants responded to a change in stimulus that crossed blue-green, yellow-green, and yellow-red category boundaries, but not to a change in stimulus within these categories. There was one exception to this: when infants were habituated to a red wavelength, their interest in the original red stimulus continued at test, most likely due to the salience of red in infancy. Apart from this, infants' recognition of novelty in hue was completely categorical, and Bornstein et al. concluded that infants categorize blue, green, red and yellow hues at 4 months. Two further studies also provide evidence of a categorical response to color in infants, finding that 4-month-olds fail to recognize the novelty of a color when original and novel colors are from the same blue or green category [4] or from the

same red or blue categories after a 5 min delay between habituation and test [5]. In addition, a fourth study indicates that 4-month-old infant's response to chromatic novelty is also categorical across blue-green, red-pink, and blue-purple category boundaries, when same- and different-category colors are equated in color metrics such as the Munsell color system and CIE ($L^*u^*v^*$, 1976) color space [6].

Therefore, a categorical response to color in infancy has now been found in multiple habituation studies, for over 25 different stimulus pairs taken from three different color metrics, and with different versions of the habituation task. In all of these studies, the same-category chromatic differences are known to be discriminable to infants at 4 months. Infants at 4 months are able to discriminate same-category chromatic differences of an equivalent size when tasks do not involve color memory (e.g., color search: [7, 8]), and the same-category chromatic differences were all much greater than chromatic discrimination thresholds at 4 months. In fact, some of the same-category stimulus separations in Bornstein et al. investigation were as much as 20 times larger than the estimated just-noticeable difference for a 4-month-old (see [6]). Even when a luminance difference is intentionally added to maximize the difference between novel and original stimuli [6], 4-month-old infants still fail to respond to the novelty of a color from the same category as the original. Infants' pattern of responding to color on habituation tasks therefore fits the classic definition of categorization that "discriminably different stimuli are treated as equivalent" (e.g., [2]). These studies provide a parallel to adult studies which have used an X-AB task. The X-AB task requires participants to encode an original color and then distinguish it from a novel foil after a delay. Adults are poorer at identifying the original color when the original and foil are from the same category than when they are from different categories. Infants' pattern of response on habituation tasks is equivalent to this, as they are also poorer at distinguishing between original and novel same- than different-category colors at test. In fact, infants are even more categorical in their response than adults as, at least

under some conditions, they fail to distinguish between the original and novel same-category colors at all.

Another method which has been used to investigate both adult and infant color categories is the event-related potential (ERP) technique. This involves measuring electrical activity from the scalp, which is then time locked to an event such as stimulus onset. Event-related potentials have been measured on a visual oddball task which is commonly used in ERP studies of both adult and infant categorization. The visual oddball task involves the frequent presentation of one stimulus (the standard) interspersed with the infrequent presentation of other stimuli (the deviants), and ERP components for standard and deviant stimuli are compared for evidence of change detection. For adults, the categorical relationship of standard and deviant colors affects the latency and amplitude of several ERP components elicited in response to the stimuli. For example, different-category deviants elicit ERP components that peak earlier or are of greater amplitude than same-category deviants, indicating quicker and greater detection of different- than same-category chromatic change (see [9] for review). Category effects in ERPs on a visual oddball task have also been found in infancy [9]. Even at 7-months, the categorical relationship of standard and deviant colors affected infants' ERP components. Category effects were found in ERP components thought to index recognition memory (slow waves) and attentional allocation (negative central). For both, the different-category deviant elicited an ERP component that was more negative than that elicited by the standard color, yet the same-category deviant did not. These findings indicate that, within the context of the frequent repetition of a color, an infrequent different-category color will elicit greater attentional allocation and will be seen as more novel than the frequent color, whereas a same-category infrequent color will not. As in the habituation studies, the same-category colors were discriminable to infants at 7 months, and therefore the equivalence of the same-category deviant and standard in ERP components of attention and memory indicates electrophysiological markers

of infant color categorization. Due to the time course of the negative central ERP component, Clifford et al. study also suggests that infants register the categorical status of a color from as early as 250 ms poststimulus onset. These findings are in line with other ERP studies of categorical responding in infancy, such as phoneme categorization and category learning of cats and dogs (see [9]).

Chromatic search tasks have also been used to investigate infant color categories. In adult studies that have used these tasks, participants are required to search for a colored target that is either from the same or a different category to the colored distracters or colored background. When same- and different-category colors are equated in CIE ($L^*u^*v^*$, 1976) or Munsell color metrics, at least under some conditions, adults are faster or more accurate at searching for targets amongst different- than same-category distracters or backgrounds. The influence of color categories on infants' color search has been investigated by measuring 4–6-month-old eye movements to colored targets presented on colored backgrounds. Infants are significantly faster at fixating a colored target [7] and at initiating an eye movement to a colored target [8] when presented on a different- than same-category colored background. In these infant studies, the chromatic difference between the target and background was equated for same- and different-category color pairs in CIE $L^*u^*v^*$ color space yet are also closely equated when stimuli are converted to the MacLeod-Boynton cone excitation color space. Franklin et al. [8] found this effect of categories on infants' target detection when the target was presented to the left visual field, but not to the right visual field, potentially suggesting that there is a right hemisphere lateralization of categorical perception of color in infancy. This right hemisphere lateralization contrasts with reports of left hemisphere lateralization in the adults' category effect (see [1]), and it appears as though a right-to-left hemisphere switch occurs around the time of color term acquisition [10]. However, these effects have so far only been tested for in infants across the blue-green category boundary, and recent studies also report a lack of hemispheric lateralization for the

effect in adults under some conditions (e.g., [11]). Further research is therefore needed to establish the nature of these effects in both infants and adults and the conditions for the pattern of hemispheric lateralization.

Negative Findings

There is only one published study which fails to find a categorical novelty response in infants [12], and that study tested for a categorical response across the red-orange category boundary using a task which determined whether a novel color “popped out” amongst six instances of a color to which infants had been familiarized. This could indicate that infants do not respond categorically across the red-orange boundary or that the novelty response is not categorical under conditions of “pop out.” There may well be conditions under which infants’ responding is not categorical, and as for adults, the relative strength of the categorical and physical code may depend on cognitive demands. However, it has been argued that an unintended category boundary for one of the same-category stimulus pairs, and the use of an inappropriate illuminant, renders the findings of Gerhardstein et al. unsafe [13]. A second unpublished study has been cited as evidence for a lack of replication of an infant category effect across the blue-green boundary. However, the habituation task that was used failed to detect any response to color in the infants at all, most likely due to particular parameters of the habituation task (e.g., a change from central to peripheral stimulus presentation at test could have detracted attention from the change in color). Therefore, as there was no response to color at all, a categorical effect could not be assessed (see [14] for further discussion). These negative findings, which should be considered as unsafe, are outweighed by the converging behavioral and electrophysiological evidence for a categorical response to color in infancy. The finding of infant color categories has now been replicated across multiple color category boundaries, using multiple tasks and color metrics, and by several independent research teams.

Alternative Explanations

As outlined above, there is converging evidence that, at least under some conditions, infants’ response to chromatic novelty or recognition memory can be completely categorical. Potential alternative explanations for infants’ response to color in these studies have been considered and ruled out. First, *a priori* hue preferences are unable to account for the findings as studies have controlled for these (see [14]). For example, same-category stimulus pairs have been sampled from both sides of the category boundary (e.g., [3, 6]), *a priori* preferences have been recorded and shown not to account for infants’ response (e.g., [6]), and control tasks have also been included (e.g., [9]). Second, potential inequalities in the color metric are unable to account for infants’ response to color. As for studies of adult color categories, there is an issue about which color metric is appropriate to equate same- and different-category colors: using a physical metric (e.g., wavelength), a perceptual metric such as Munsell or CIE ($L^*u^*v^*$), or a cone-opponent color space (e.g., MacLeod-Boynton color space). Importantly, infant color categories appear to transcend this metric issue. Infants’ pattern of response is identical across studies which vary in the color metric used. Additionally, at least under some conditions, infants’ novelty response to color appears completely invariant to the color metric, as distances in any of these metrics do not predict the degree of infants’ chromatic novelty preference or change detection. Infants treat same- but not different-category chromatic differences as equivalent on habituation or visual oddball tasks, irrespective of the size of the same-category chromatic difference. Only the categorical relationship between stimuli can account for this pattern of response.

Theoretical Questions and Future Directions

The converging evidence for infant color categories strongly suggests that there is a nonlinguistic route to parsing the continuum of color into

discrete groups. However, there is also convincing evidence that the influence of color categories on adults' cognition and perception is dependent on language and that different languages also vary in how they divide up the color spectrum [see Roberson]. This raises two sets of challenging questions for future research to consider and address. First, what is the relationship between prelinguistic and linguistic color categories? Why do some languages not have color terms for categories that infants have? For example, many of the world's languages do not have separate terms for blue and green, yet infants appear to form separate categories for these. The answer to this question may be in the analysis of the World Color Survey [see TBA]. Multiple analyses of the World Color Survey strongly suggest that, despite variation in color lexicons, there is also striking commonality in how the world's languages divide up the color spectrum. For example, the centers of color categories across languages tend to cluster around particular points in the color space, with distinct clusters for "blue" and "green." It has therefore been suggested by several groups that there are perceptual constraints on how categories form in color lexicons (see [1]). Therefore, although not all languages have separate terms for blue and green, there appears to be a common pressure in the world's color lexicons for separate categories of blue and green to form. One possibility is that this commonality arises because infant color categories partially constrain the world's color lexicons, while local variation arises due to the additional influence of culture and communication [1].

Alternatively, infant color categories and those in language could operate independently from each other but could have similar organizing principles. This raises the question of what happens when prelinguistic and linguistic color categories converge when color terms are learnt. What are the mechanisms of the switch from a nonlinguistic to a language dominant division of the color spectrum? How do prelinguistic color categories affect color term acquisition? The presence of infant color categories strongly suggests that any difficulty that children have in learning color words is unlikely to be due to a lack of category structure

but rather provides support for the argument that there are general attentional, conceptual, and linguistic constraints that make learning color words difficult relative to nouns, but not other adjectives (see [15]). However, it is theoretically possible that color term learning may be more difficult for lexicons which map less well onto the prelinguistic category structure that is there in infancy.

Second, what are the organizing principles of infant color categories? Where do infants get their category structure from? Several theories have been proposed to identify the organizing principles of color categories. For example, it has been proposed that color categories derive from (i) the clustering of colors taken from surfaces in the natural chromatic environment; (ii) reflection properties, with some colors (those that correspond to the unique hues) reflecting incoming light in a much simpler way than others; or (iii) an irregularly shaped perceptual color space combined with basic principles of categorization (see [1]). None of these organizing principles rely on language, and all are contenders for explaining how infants are able to categorize color even before color terms have been learnt. Importantly, color categories may not be inbuilt into our visual system but rather could be a cognitive response to inequalities in our visual system's processing of color. Some hues could be more salient than others, for example, due to having greater maximum levels of saturation or different reflection properties. This could lead to some colors being more informative than others, and inequalities such as these could provide structure for cognition when categorizing. The actual categorical response to color in infants could therefore come from an interaction of perceptual inequalities in color and a general cognitive strategy to categorize. It is well known that categorization is a core part of infant cognition and is essential in helping infants to navigate an uncertain world (e.g., [2]). If color categories do result in part from general cognitive strategies, differences between humans and animals in the extent of the categorical response to color could arise due to differences in cognition, even if color vision is similar (e.g., baboons; see Weber and Changizi).

These questions on how infant color categories relate to those in language, and on the organizing principles and underlying mechanisms of infant color categories, provide clear direction for future research. Investigations which empirically address such questions will resonate beyond color and will have implications for fundamental questions on how categories are formed, how categories in language and thought interact, and how categories are expressed in the brain.

Acknowledgements The writing of this entry was supported by a European Research Council Grant to AF (Project CATEGORIES, ref: 283605).

Cross-References

- [Ancient Color Categories](#)
- [Color Categorical Perception](#)
- [Comparative Color Categories](#)
- [World Color Survey](#)

References

1. Regier, T., Kay, P.: Language thought and color: Whorf was half right. *Trends Cogn. Sci.* **13**, 439–446 (2009)
2. Mareschal, D., Quinn, P.C.: Categorization in infancy. *Trends Cogn. Sci.* **5**, 443–450 (2001)
3. Bornstein, M., Kessen, W., Weiskopf, S.: Color vision and hue categorisation in young human infants. *J. Exp. Psychol. Hum. Percept. Perform.* **2**, 115–129 (1976)
4. Catherwood, D., Crassini, B., Freiberg, K.: The nature of infant memory for hue. *Br. J. Dev. Psychol.* **5**, 385–394 (1987)
5. Catherwood, D., Crassini, B., Freiberg, K.: The course of infant memory for hue. *Aust. J. Psychol.* **42**, 277–285 (1990)
6. Franklin, A., Davies, I.R.L.: New evidence for infant colour categories. *Br. J. Dev. Psychol.* **22**, 349–377 (2004)
7. Franklin, A., Pilling, M., Davies, I.R.L.: The nature of infant color categorisation: evidence from eye-movements on a target detection task. *J. Exp. Psychol.* **91**, 227–248 (2005)
8. Franklin, A., Drivonikou, G.V., Bevis, L., Davies, I.R.L., Kay, P., Regier, T.: Categorical Perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3221–3225 (2008)
9. Clifford, A., Franklin, A., Davies, I.R.L., Holmes, A.: Electrophysiological markers of color categories in the infant brain. *Brain Cogn.* **71**, 165–172 (2009)
10. Franklin, A., Drivonikou, G.V., Clifford, A., Kay, P., Regier, T., Davies, I.R.L.: Lateralization of categorical perception of color changes with color term acquisition. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 18221–18225 (2008)
11. Witzel, C., Gegenfurtner, K.: Is there a lateralized category effect for color? *J. Vis.* **11**, 1–25 (2011)
12. Gerhardstein, P., Renner, P., Rovee-Collier, C.: The roles of perceptual and categorical similarity in colour pop-out in infants. *Br. J. Dev. Psychol.* **17**, 403–420 (1999)
13. Davies, I.R.L., Franklin, A.: Categorical similarity may affect colour pop-out in infants after-all. *Br. J. Dev. Psychol.* **20**, 185–203 (2002)
14. Franklin, A.: Pre-linguistic categorical perception of color cannot be explained by color preference: response to Roberson and Hanley. *Trends Cogn. Sci.* **13**, 501–502 (2009)
15. Franklin, A.: Constraints on children's color term acquisition. *J. Exp. Psychol.* **94**, 322–327 (2006)

Influences Between Light and Color

- [Interactions of Color and Light](#)

Instrument: Colorimeter

Shao-Tang Hung

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Definition

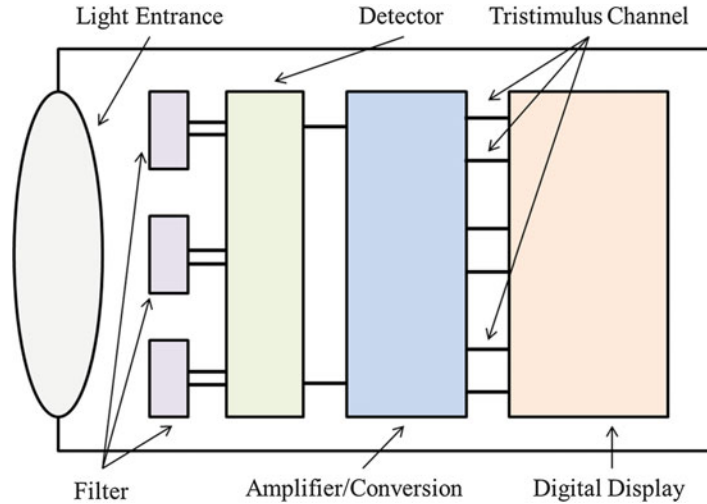
Synonyms

[Tristimulus colorimeter](#)

Definition

There are two kinds of modern colorimeters: one that includes and one that excludes a light source for measuring the surface or a self-luminous stimulus respectively. Both types of colorimeter measure International Commission on Illumination (CIE) tristimulus values under particular illuminant and observer conditions, e.g., D50/2°.

Instrument: Colorimeter,
Fig. 1 Schematic structure
 of a colorimeter [3]



The instruments are applied to measure a test stimulus displayed on a monitor, printer or a light stimulus. CIE International Lighting Vocabulary [1] defines the colorimeter as an instrument for measuring colorimetric quantities, such as the tristimulus values of a color stimulus.

Overview

The colorimeter measures XYZ values in a luminance unit. This overview is divided into four sections: optical structure, measurement geometry, spectral responsivity and measuring uncertainty, and applications [1] (► [Photodetectors](#)).

Optical Structure

The structure of the colorimeter can be divided into two parts: the measuring head and the digital display. Figure 1 shows the schematic structure of a colorimeter. The measuring head contains tristimulus filters, detectors, and spatial corrections of the detector response. There are filters in front of the detectors to filter the input light, which is from the entrance window. The filters mimic the CIE color-matching functions. The parallel photodetectors sense the filtered light in the electronic signal ► [Photodetectors](#). The multiplex amplifier can integrally form the signal to feed via output channels. Finally, the XYZ tristimulus is reported on a digital display [2].

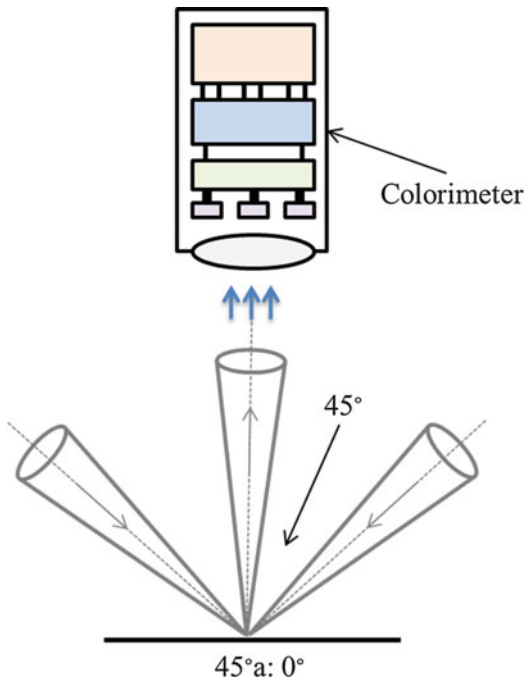
Measuring Geometry

There are three factors affecting the measurement results: the limiting aperture, the acceptance area, and the optical axis of the colorimeter head. The limiting aperture affects the quantity of the incoming light. The acceptance area evaluates the incident radiation to improve accuracy. The optical axis is the middle point of the measured surface or light source, which is related to the light entrance of the measuring head [3]. Figure 2 reveals $0^\circ/45^\circ$ measuring geometry; more details are shown under (► [Photometer](#)).

Spectral Responsivity and Measuring Correction

The measured spectrum is reported across a visible spectrum of 380–780 nm. Because the color-matching functions (CMF) $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ approximate the three human spectral responsivities, the real relative spectral responsivities (RSR) $s_x, r(\lambda)$, $s_y, r(\lambda)$, $s_z, r(\lambda)$ show some deviation. They do not match the color-matching functions exactly.

The mismatch correction of the spectrum must be considered; many of the corrected factors are used to evaluate the mismatch and to correct these deviations. For example, there are three types of factors corrected by comparing actual and target functions: the spectral mismatch, ultraviolet response ($f_{UV, i}$), and infrared response ($f_{IR, i}$). The corrected indexes are also used for the correction of the spectral transmittance factor ($\bar{\tau}(\lambda)$).



Instrument: Colorimeter, Fig. 2 0°a/45° measuring geometry

Applications

The colorimeter can measure XYZ tristimulus values in the illuminant industry, in paper reflectance, and in digital imaging. All of the measuring data, which calculate the color tristimulus to compare the test with reference values, following the CIE CMF, are employed, the user demand to determine the visual angle, and choose the corresponding CMF data (2°/10°) [2]. Moreover, the reference illuminant is also considered when the colorimeter calculates the XYZ tristimulus [4].

Most important for the colorimeter is that it is used to assist in calibrating color devices. For a printer, the colorimeter measures the different color patches that are printed on the paper. The measurement results are compared with the reference printer data, and the color difference is calculated to ensure colorimetric consistency.

Applications in the lighting industries can measure XYZ tristimulus in units of luminance (cd/m^2). The luminance can be represented clearly to correct the illuminant. Furthermore, the handheld colorimeter can be conveniently used

outdoors to measure the correlated color temperature (T_{cp}) and luminance.

Cross-References

► [Photodetector](#)

References

1. International Commission on Illumination: ILV: International Lighting Vocabulary. CIE S017/E:2011, Vienna (2011)
2. International Commission on Illumination: Methods for Characterising Tristimulus Colorimeters for Measuring The Colour of Light. CIE 179:2007, Vienna (2011)
3. Ohta, N., Robertson, A.: Colorimetry: Fundamentals and Applications. Wiley, Hoboken (2006)
4. X-Rite: A Guide to Understanding Color Communication. Grand Rapids (2007)

Instrument: Photometer

Shao-Tang Hung

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

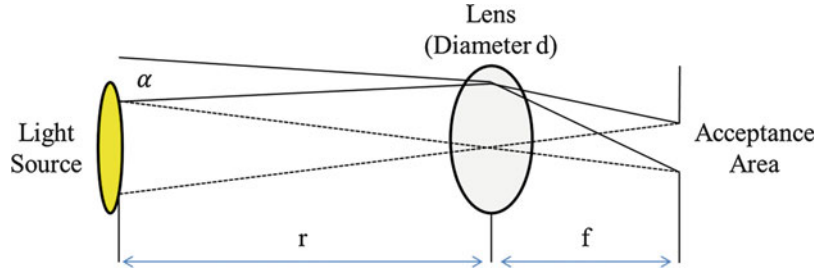
Synonyms

[Luminous intensity meter](#)

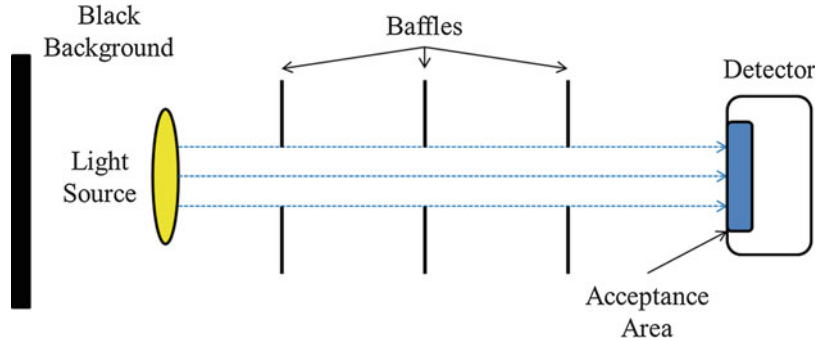
Definition

The photometer is an instrument which can be used to measure photometric quantities such as luminance, illuminance, etc. The parameters are derived from means of law of additives and CIE spectral luminous efficiency functions. The instruments usually transport the measured data either digitally or analogly to display on the screen. CIE had defined the international vocabulary of the photometer: instrument for measuring photometric quantities [1].

Instrument: Photometer,
Fig. 1 Luminous intensity
 measurement [2]



Instrument: Photometer,
Fig. 2 The perfect
 conditions for luminous-
 intensity meter
 measurement [2]



Overview

The photometer is a universal vocabulary which has plenty of types. According to the applications, the types can be divided to luminance meter, illuminance meter, etc. The luminous-intensity meter is a conventional photometer, which can measure directly luminous intensity (candela, cd). This article will focus on the luminous-intensity meter and is divided into three sections to introduce luminous-intensity meter structure and measuring theories, measuring fields and range, and applications and comparison [1–3].

Luminous-Intensity Meter Structure and Measuring Theories

The luminous-intensity meter consists mainly of detector unit. Following the measuring theories, the spectral weighting or the spectral dispersion device is arranged in front of the detector to separate the light. The lenses and apertures assist light transportation for luminous intensity measurement, see Fig. 1. The use of additional lenses can change the field of the focus to make it more flexible [4]. Luminous intensity can be ensured by measuring theories.

The data is measured by a detector which is included in the measuring detector that is transformed in terms of the digital signal, pass-through amplifier, and digital processes. If users must calculate other illuminant parameters such as luminance, illuminance, etc., the data can be outputted to the computer for further analysis.

Acceptance Area

Following the CIE definition, the acceptance area is the area of the luminous-intensity meter detector which is receiving and directionally evaluating the incident light. The acceptance area affects directly luminous intensity. In manufacturing processes of luminous-intensity meter, the measuring field must be defined accurately.

The measurement results are affected certainly by the environmental condition such as the background. The entire scene which includes light is also evaluated before the light input to the luminous-intensity meter detector. In the metrologically perfect conditions, the explanation is presented in Fig. 2. The light source is installed in front of the black background. The stray light is sheltered by baffles; the detector can receive the

pur light that is not interfered with by the other nonstationary light [2].

Applications

The SI unit of luminous intensity is the candela (cd). The luminous-intensity meter can be used to measure light source in a particular direction per unit solid angle. The luminous-intensity meter is an adaptive tool to calibrate the optical instruments. The sensitive detector is also applied to measure flicker or nonstationary optical wave. Furthermore, the other types of photometers such as luminance meter and illuminance meter can measure environmental variety. For example, luminance and illuminance are measured continuously for realizing LED light and overall light respectively. The correlation also can be found.

References

1. International Commission on Illumination: ILV: International Lighting Vocabulary, CIE S017/E:2011, Vienna (2011)
2. International Commission on Illumination: The Measurement of Absolute Luminous Intensity Distributions, CIE 70–1987, Vienna (2011)
3. International Commission on Illumination: Methods of Characterizing Illuminance Meters and Luminance Meters, CIE 69–1987, Vienna (2011)
4. International Commission on Illumination: Measurement of Luminous Flux, CIE 84–1989, Vienna (2011)

Instrument: Spectrophotometer

Shao-Tang Hung

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

Spectroreflectometer (for reflectance measurement)

Definition

The spectrophotometer is an instrument for measuring the reflectance or transmittance quantities

of an object, including specular and diffuse reflectance and regular and diffuse transmittance. On request, the instrument can measure the above quantities at different viewing/illumination geometry. The International Commission on Illumination (CIE) defines the spectrophotometer in its International Lighting Vocabulary [1] as an instrument for measuring the ratio of two values of a radiometric quantity at the same wavelength.

Overview

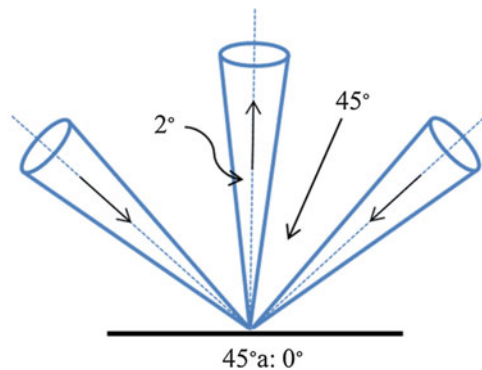
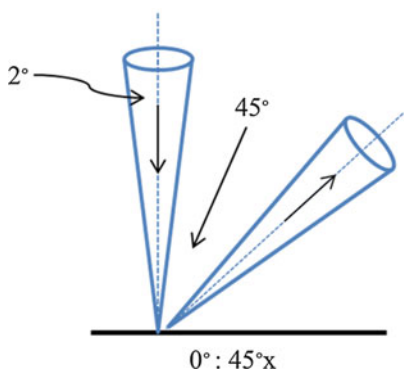
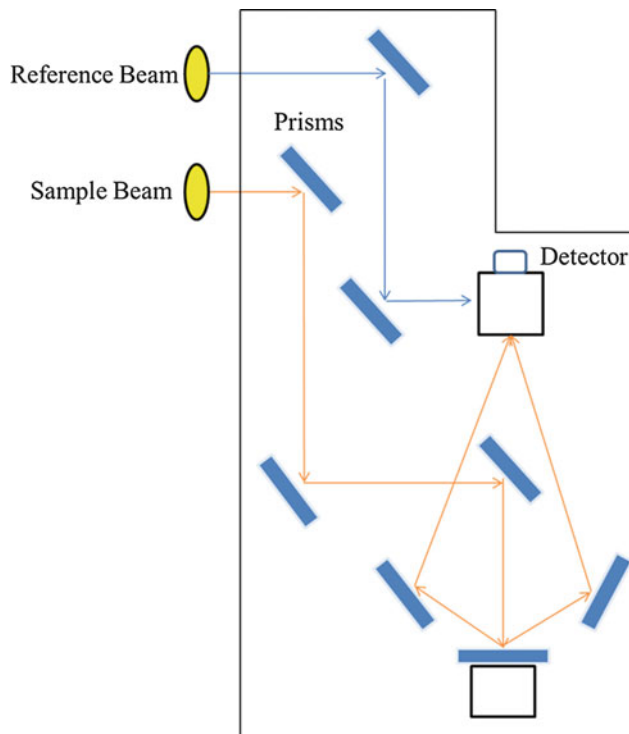
The spectrophotometer is used to measure spectral reflectance or transmittance across a visible spectrum. The structure consists mainly of a light source, a system of optical transmission, and a detector. Furthermore, the viewing geometries also affect the measurement results. The instrument is composed of the following three parts: structure, viewing/illumination geometries, and application [1, 2].

Structure

The spectrophotometer consists of multiple structures to carry out light transmission and sample measurement. First of all, the measurement condition divides the wavelength into three sections: UV, visible, and IR sections. The deuterium lamp is used to measure the UV section. The visible and IR sections are measured by the halogen lamp. Second, the beam is divided into the reference beam and the sample beam. The reference beam and the sample beam are used to monitor stability and measure the sample, respectively. Finally, the two beams are transmitted by the multiple prisms and focused on the detector. The optical transmission is revealed in Fig. 1.

Viewing/Illumination Geometries

The geometries for measuring reflectance depend on radiation field and angle; several standardized geometries are defined by the CIE. For color measurement, some of the standard geometries are shown in Fig. 2. The spectrophotometer commonly uses bidirectional geometry $0^\circ:45^\circ$ to calibrate the color in the National Measurement Laboratory R.O.C.

Instrument:
Spectrophotometer,
Fig. 1 The optical transmission of a spectrophotometer [3]

Instrument: Spectrophotometer, Fig. 2 Standard geometries for spectral reflectance [1]

The Field of View

The tristimulus values are calculated from spectral reflectance, the factors are affected by the field of view ($2^\circ/10^\circ$). According to the CIE definition, the field of view can be decided by a range; if more than 4° , the field of view is 10° , and if less than 4° , the field of view is 2° . The field of view is one of the important elements for evaluating color. According to the field of view, the color-matching

function (CMF) is divided into two types: the CIE 1931 2° Standard Observer and the CIE 1964 10° Standard Observer. The two types of CMF are used to fit the different measurement conditions.

Applications

For color measurement, the spectrophotometer measures a variety of surface colors, such as textiles, coatings, and plastics. The results are reported as

spectral reflectance or transmittance across the visible spectrum. The spectrophotometer is the appropriate device for measuring color data, and provides the correct color information for the manufacturing of textile products, printer products, and paint and material products to solve color difficulties.

References

1. Schanda, J. (ed.): Colorimetry: Understanding the CIE System. Wiley, Hoboken (2007)
2. International Commission on Illumination.: ILV: International Lighting Vocabulary. CIE S017/E: 2011, Vienna (2011)
3. Industrial Technology Research Institute Center for Measurement Standards.: Spectrophotometer MSVP. Hsinchu (2009)

Integrating Sphere

Cheng-Hsien Chen

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

Ulbricht sphere

Definition

An integrating sphere is a hollow spherical cavity with a diffuse-reflecting material inside the sphere that has a spatially uniform surface, and a flat profile of spectral power distribution. The structure of integrating spheres generally consists of a baffle, different apertures, a light source, and a detector. The ideal integrating sphere has the following characteristics:

- (i) Small detector and external port.
- (ii) Miniature devices (light source, baffle, and mount) inside the sphere, which cannot affect optical radiation after its first reflection.

- (iii) The surface of the sphere is a homogeneous Lambertian reflector, with reflectance independent of the wavelength.

In color applications, an integrating sphere is used to produce a uniform light source from a luminaire or for measuring the transmittance or reflectance of a surface in a particular viewing geometry [1, 2].

Overview

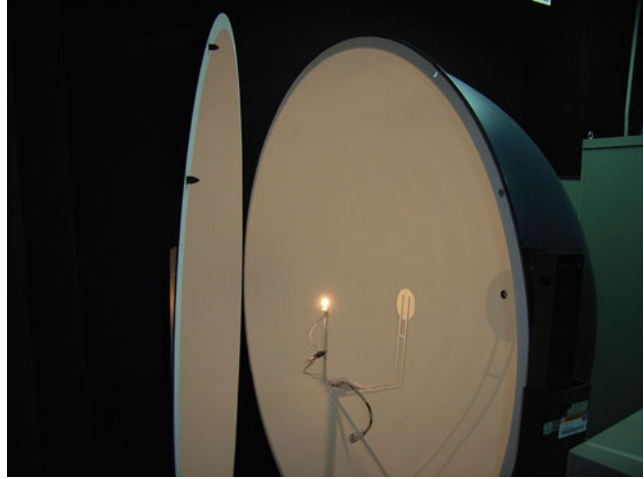
As mentioned earlier, one application is for the measurement of the color properties of a luminaire. In general, the substitution method is utilized to calculate the color properties of a test and a standard lamp, or transfer standard. The standard lamp provides the spectral flux, which can calculate the colorimetry and photometry quantities. Two types of detector are used on the sphere: a broadband $V(\lambda)$ -corrected photometer, and a narrow-band spectroradiometer [3]. The spectral mismatch error of the calibrated photometer should be corrected, according to the technical report of the International Commission of Illumination (CIE) [4]. The spectral response of the photometer is easy to apply to the measurement of the luminous flux of a test lamp. The spectroradiometer may be applied for spectral power distribution and colorimetry of test lamp measurement (Fig. 1).

When measuring surface properties, an integrating sphere is used as a uniform lighting source or a diffuser of the incident light. For the surface property measurement system, the integrating sphere is set to diffuse the reference beam and geometry structure to analyze the characteristics of the surface. According to various standard geometries, the integrating sphere provides various conditions to realize different optical properties (photometry, colorimetry, haze, gloss, clarity, shading, etc.) of the surface colorimetry and photometry measurement [5, 6].

The integrating sphere method requires a standard lamp, and the substitution method is often utilized to calculate the total luminous flux of the test lamp. There are some important components

Integrating Sphere,

Fig. 1 An image of an integrating sphere. The baffle inside the integrating sphere is designed to directly avoid the light incident on the detector port. The shadow from the baffle must cover the detector



inside the integrating sphere to consider. One such device is the baffle. The size of the baffle, used to prevent incident light from reaching the detector, depends upon both the size of the detector, and the distance between the lamp and the detector. A second device to consider is an auxiliary lamp. In practice, the shape of the test lamp is different from that of the standard lamp. The auxiliary lamp can provide the absorption correction function between the standard lamp and the test lamp.

The equation of luminous flux Φ_t is

$$\Phi_t = \frac{I_t}{I_s} \cdot \frac{A_s}{A_t} \cdot \Phi_s \quad (1)$$

where I_t is the signal of the test lamp, I_s is the signal of the standard lamp, A_s is the signal of the auxiliary lamp with the standard lamp, and A_t is the signal of the auxiliary lamp with the test lamp.

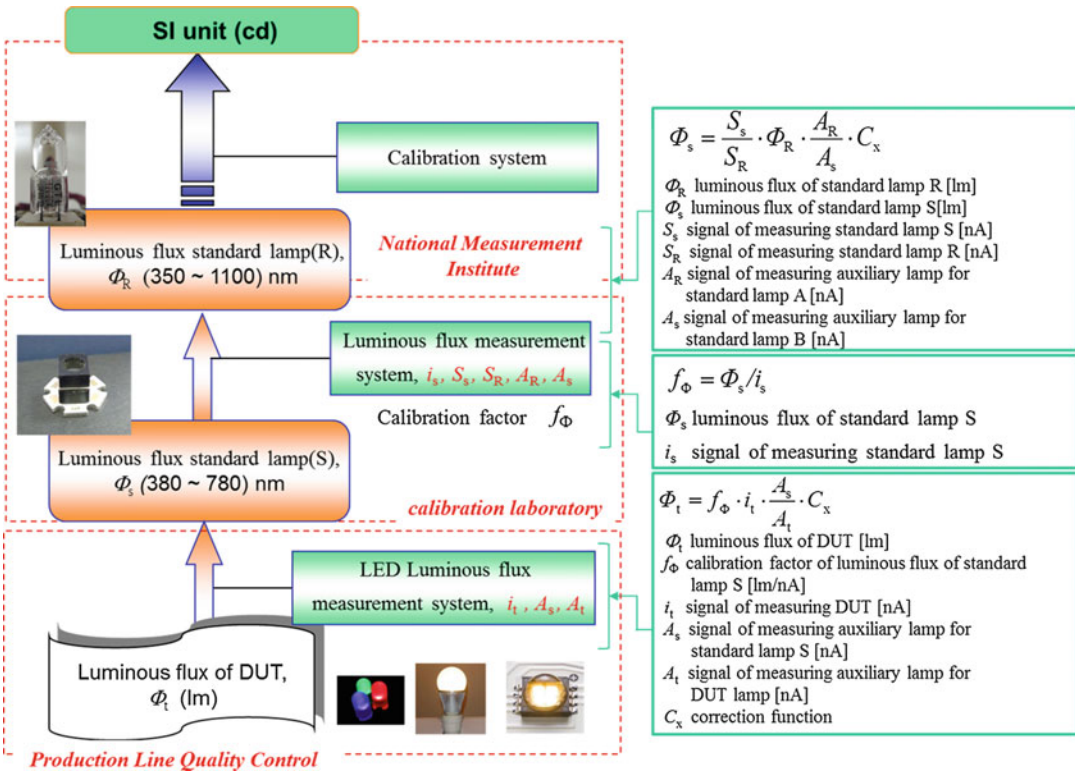
The photodetector shows some mismatches between the spectral luminous efficiency function and the spectral response of the detector and lamp source relation. The nonuniformity of the integrating sphere causes the directional light source to contribute measurement errors inside the sphere. The baffle size and different geometry inside the sphere provide varying spatial response distribution. Therefore, the equation of luminous flux Φ_t is given by

$$\Phi_t = \frac{I_t}{I_s} \cdot \frac{A_s}{A_t} \cdot ccf \cdot scf \cdot \Phi_s, \quad (2)$$

where ccf is the spectral mismatch correction factor, and scf is the spatial response distribution function (SRDF) of the integrating sphere [7, 8].

A good measurement system of an accreditation laboratory should have a suitable traceability chain. The metrological traceability chain, according to the International Standards Organization (ISO) guide 99 is the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. Therefore, the product line quality control of manufacturing needs a transfer standard from the calibration laboratory. The calibration laboratory uses the reference standard trace to the National Measurement Institute (NMI), and then the trace to the SI unit shown in Fig. 2.

The stable and repeatable transfer standards, which have similar characteristics to the test lamp, are needed for the measurement system to transfer photometric and colorimetric units from the NMI or an accreditation laboratory. In practice, the transfer standard is made by a light-emitting diode (LED) lamp whose optical characteristics are similar to those of the test lamp and are measured using the same measurement instrument. For the calibration laboratory, the luminous



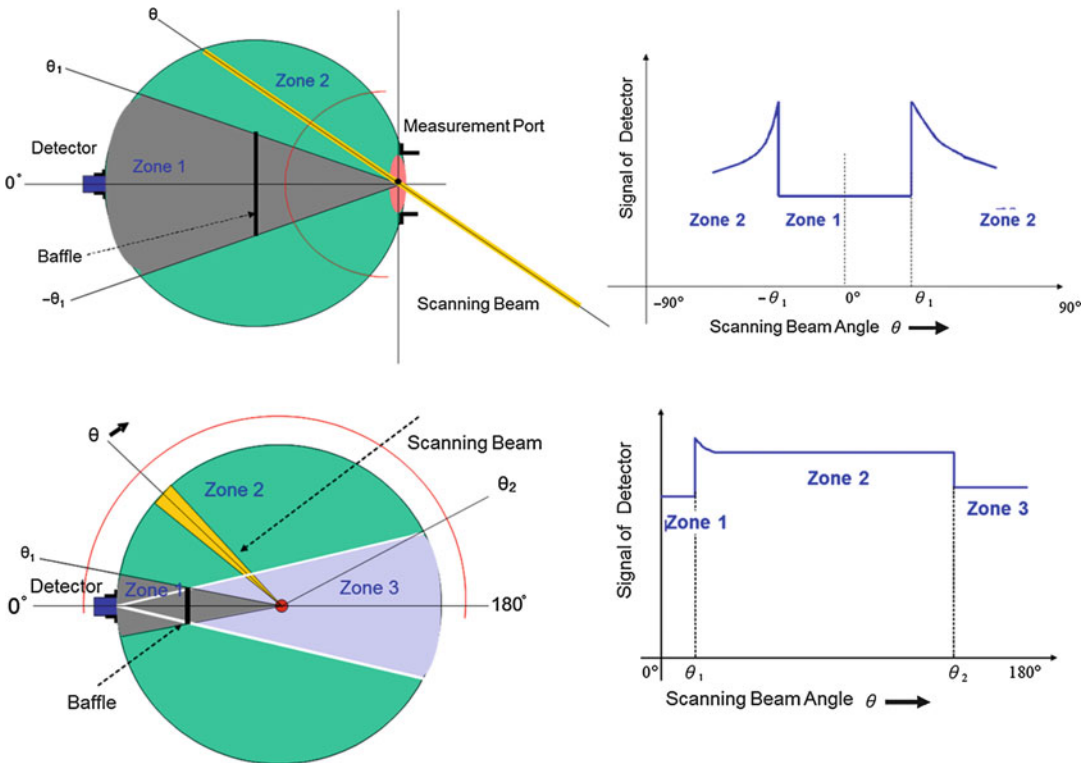
Integrating Sphere, Fig. 2 The metrological traceability chain of a luminous flux standard lamp calibration system

flux measurement system transfers the standard from the reference standard to the LED transfer standard. The NMI transfers the standard, which is made by a stable, repeatable, and broadband spectrum lamp, such as a tungsten or halogen lamp, to the calibration laboratory.

For the luminous flux measurement of lighting, an integrating sphere is sometimes used together with calibration with a standard lamp that has similar characteristics and by using a comparison method to transfer luminous flux quantity to the measurement instrument. Measurement of LED lighting involves two kinds of measurement geometries. First, the test sample is mounted on the integrating sphere. Second, the test lamp is located in the center of the sphere. Why do the test lamp and standard lamp have similar characteristics? Because the measurement result is achieved by comparing the input and output signals. If we consider the spatial nonuniformity of integrating, by rotating a scanning beam, we will

find the nonperfect output of the integrating sphere. If we use the narrow beam light source to scan inside the sphere, we measure the different signal from the detector in the different scanning region shown in Fig. 3.

The 2π geometry would have two zones, and zone 2 would have a drift signal as the large scanning beam angle. This means that when using a wide diverging angle lamp such as a halogen lamp or a diffused lamp such as a standard lamp, and the narrow diverging angle of the test lamp such as some 5-mm LEDs, there would be more measurement uncertainty for calibration, because the large size of the integrating sphere and perfect Lambertian surface were good for measuring the luminous flux of the LED. For the 4π geometry, zone 1 is the baffle region with the scanning beam near the edge of the baffle, and we would measure a higher signal from the detector. Some light is directly incident to the detector because of scattering. If we use a large sphere



Integrating Sphere, Fig. 3 Spatial nonuniformity of the integrating sphere

and a baffle and mount of smaller geometry, a large area of the flat region would be zone 2. Zone 3 is the shadow region made from the baffle and detector by the scanning beam.

Most important is the light distribution between the standard lamp and the test lamp. If the light distribution is different, the spatial nonuniformity of the integrating sphere should be considered. The nonuniformity of the integrating sphere causes the directional light source to make a measurement error inside the sphere. The baffle size and different geometry inside the sphere provide a different spatial response distribution.

Ideally, the light distribution of the device under test (DUT) lamp and the standard lamp should be similar. If not, the spatial distribution mismatch correction is applied complexly. If measurement of a directional light source is required, for which the light distribution is different from the standard lamp, the spatial

response distribution function (SRDF) should be considered. The SRDF is relative to the baffle size, the DUT lamp, and the geometry inside the sphere. In practice, the point source in the center of the sphere should be used and the light distribution of the DUT measured to calculate the SRDF. In addition, the same method of evaluating the SRDF to measure the directional light source in the center of the integrating sphere can be used. Overall, by knowing the limitation of the integrating sphere, the directional light source can be precisely measured, and the measurement uncertainty can be reduced [9].

Cross-References

- ▶ Spectral Power Distribution
- ▶ Spectroradiometer
- ▶ Standard Lamp

References

1. International Commission on Illumination.: ILV: International Lighting Vocabulary. CIE S017/E: 2011, Vienna (2011)
2. Gigahertz-Optik website. <http://www.light-measurement.com/ideal-integrating-sphere/> (2014)
3. Illuminating Engineering Society of North America.: Electrical and Photometric Measurements of Solid-State Lighting Products. IES LM-79-08, New York (2008)
4. International Commission on Illumination.: Measurement of LED. CIE 127:2007, Vienna (2007)
5. International Commission on Illumination.: Measurement of Luminous Flux. CIE 84:1989, Vienna (1989)
6. International Commission on Illumination.: The Basis of Physical Photometry. CIE 18.2:1983, Vienna (1983)
7. Ohno, Y.: Integrating sphere simulation: application to total flux scale realization. *Appl. Optics* **33**(13), 2637–2647 (1994)
8. Ohno, Y., Daubach, R.: Integrating sphere simulation on spatial nonuniformity errors in luminous flux measurement. *J. IES* **30**(1), 105–115 (2001)
9. Chen, C.H., Pong, B.J., Jiaan, Y.D., Lin, H.L.: Lamp orientation dependence of an integrating sphere response for directional light source in luminous flux measurement. *NCSLI Measure J. Meas. Sci.* **8**(2), 42–47 (2013)

Interactions of Color and Light

Ulrich Bachmann¹ and Ralf Michel²

¹Institute of Colour and Light, Zurich, Switzerland

²Institute Integrative Design | Master Studio Design, University of Applied Sciences FHNW Basel, Basel, Switzerland

Synonyms

Color and light effects; Color and light interdependence; Influences between light and color

Definition

Color cannot be observed without light, and without material artifacts light is not visible – thus one can simplify and say: no color exists without light and no light could be present without color. Color is a visual sensation that depends, among other

factors, on stimulation of the visual system by radiation within the visible range – light – either emitted by a source or transmitted or reflected by *pigmented* objects. Indefectibly, light has color; there is no colorless light (light can be white, but white is also a color sensation). The chromatic composition of incident light affects the perceived color of pigmented surfaces. Color and light are like both sides of the same coin; they cannot be separated. For the creative disciplines involving design (design, *interior design*, *scenography*, architecture, etc.), the use of color and light and the knowledge about how one interacts and influences each other is a fundamental issue.

Interactions of Color and Light in the Context of Creative Training and Practices

Initial Position

In creative training and practices, the determining aspects of the interactions of color and light are generally not represented in a systematized manner. The present authors have responded to this shortcoming in collaboration with Florian Bachmann, Marcus Pericin [1], and others by examining the relevance of the interaction of color and light and by describing the aspects of it which are most important to creative training and practices. In addition, they use different approaches to exemplify the makings of a new teaching model.

Color cannot be observed without light, and without material artifacts light is not visible. Thus, one can simplify and say: no color without light and no light without color. For the creative disciplines, the use of color is one of the fundamental dimensions. And yet, despite the reciprocal interdependence of color and light, there has as yet been no comprehensive and systematizing concept for theory and creative practice.

Research on the Interactions of Color and Light in the Creative Disciplines

This entry is based on three research projects on the interactions of color and light which build

upon each other's findings: ColorLightLab, LED-ColorLab, and Colour and Light – Materials for a Theory of Colour and Light [2].

The research starts from the premise that design systems using color and light which are employed in design practice (and by analogy in *interior design*, *lighting design*, etc. as well) do not relate to each other. This despite the fact that material color cannot be perceived without light and that light cannot be seen in the absence of material artifacts. This shortcoming is reflected in the practice and methods of creative teaching. There is no theory which combines color and light in a systematic manner. The aim of the research was to pursue a practically oriented investigation of the principles involved in the interaction of color and light in terms of the requirements imposed by design and, in addition, to express research findings in practically relevant materials towards a color-light theory.

– ColorLightLab

This project investigated the principles of the interactions of light and material color. The research was carried out using the methods of artistic research and research through design [3–5] and culminated in a presentation of the phenomena under investigation and the publication *Colours Between Light and Darkness* [6].

– LED-ColorLab

The focus was on an investigation of the interaction of dynamic light and colored surfaces in a spatial context. Here the interactions of dynamic light in the color spectrum were systematically observed, as were those in the spectrum of white light. The findings were presented in a laboratory display and evaluated in surveys with experts (architects, *interior designers*, designers, *lighting designers*, etc.) and with consumers. Approaches from the realm of research through design were combined with survey methods.

– Colour and Light – Materials for a Theory of Colour and Light

Here the focus was on conveying the practical relevance of the previous findings.

Materials for a theory of color and light (installations, digital and physical tools) were developed in an experimental context and evaluated by experts as well as school and university students (methods used were research through design as well as qualitative and quantitative surveys). These materials were summarized in a multimedia publication [7] which comprises: description of the installations, contextualization of practice, theory and research, compendium of a DVD with videos about the installations, a specifically developed software, and a relaunched website setting attitudes to the interaction of color and light in the context of historical systems of color.

The experiments and empirical investigations were set up on the basis of the following research questions:

- How is the *perception of color* on surfaces and in spaces affected by lighting?
- What effect does the physical composition of surfaces, materials, and spaces have on the perception of color under different lighting?
- What interactions occur between dynamic light and colored surfaces?
- How can one arrive at a comprehensible description of the interaction of dynamic light and colored surfaces?
- What are the requirements to be met by materials and instruments used in a theory of color and light?

These questions derive from experiences gained in design practice and theory.

Summary of Investigations

First, models and prototypes were built, and several test measurements were carried out to compare the differentiation of various surfaces and various light sources; then a measuring system was constructed, where calibrated light falls on two precisely defined and identical surfaces in such a way that the remitting light can be measured by means of sensors. Thus a kind of color-sampling booth (*LED-ColorCase*) was produced, enabling surface colors under different light

sources to be visually compared and methodologically described. The LED-ColorCase makes it possible to record and evaluate *hue* shifts visually and in terms of spectral analysis. What is striking is the powerful influence of the light sources used on the difference of the resultant color effects. It was also possible to verify the possibility of calculating and predicting the resultant color impressions.

The experiments and laboratory investigations in the LED-ColorLab project showed that hue shifts develop a far greater differential under different light sources than was previously supposed and that they are clearly perceptible. In general terms one can say that the effect of and alteration caused by *hue*, *saturation*, and *brightness* are underestimated and generally not sufficiently taken into account in color and light design. In an ideal situation, it should be possible to give such alterations primacy during planning and to move towards predicting them.

Summary of Surveys and Empirical Aspect

The empirical aspect of the LED-ColorLab research project resulted in clear findings in terms of practical applicability. Surveys carried out during the LED – Staging Color and Light [8] exhibition (a public laboratory situation and part of the LED-ColorLab research project) showed that one can recognize the importance of the interactions of dynamic light (across the whole color spectrum but also dynamically in the white light spectrum from cool white to warm white): experts in the fields of architecture, *interior design*, design, and *lighting design* rate this potential highly in terms of their own creative practice. At the same time, the experts who were questioned demonstrated a lack of methodological and technical knowledge. End users questioned in the LED-ColorLab installations use recent lighting technology either as mood lighting or task lighting (this relates to LED). They rate highly the potential for integrated use of color and light in their personal environment and their place of work. Regarding the effects of the color shifts described above, acceptance is almost at 90 % overall.

Similar results are seen in analysis of the responses to the research project Colour and Light – Materials for a Theory of Colour and Light, which took as its focus the development and evaluation. In the course of in-depth interviews and on the basis of surveys in the display of materials, 71 % of the experts questioned acknowledged not only the quality of the materials produced but also the underlying approach towards a theory of color and light for creative practice.

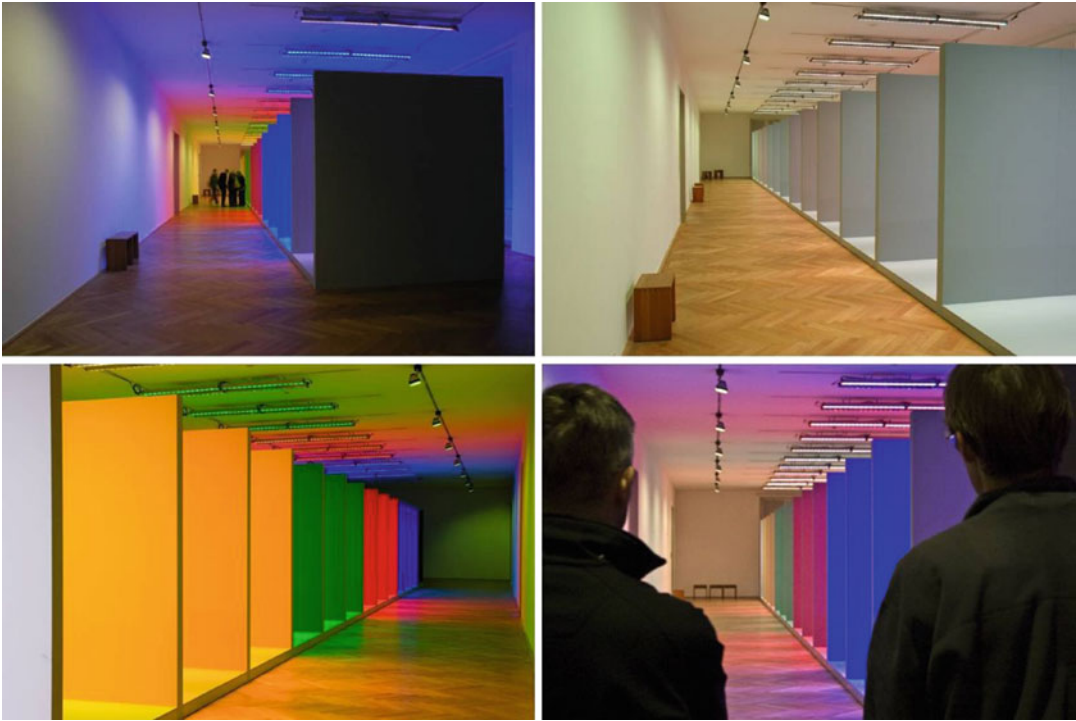
Implementing Knowledge as Mediation Tools and Materials

The principles of interactions of dynamic light and colored surface were laid out in the multimedia publication *Colour and Light: Materials for a Theory of Colour and Light* [7]. Here the aspect of direct sensual perception is given especial prominence. The authors are convinced that a theory of light and color has a particularly strong influence on creative practice if it exploits direct sensual perception of the phenomena of the interactions of color and light. In this the authors are following, among other influences, the concept of the German philosopher Gernot Böhme who writes, using the argument of an aesthetic (a general theory of perception) about the effects of atmospheres: "...the primary subject of sensual perception is not the things one perceives but what one feels: atmospheres. . ." [9]. As a consequence, conclusions drawn from the creative experiments as well as experience gained from the installations and tools were used to develop elements of a didactic concept in which sensual perception of the phenomena of the interactions of color and light are linked to the direct experience of learners.

Selected Instruments of a Theory of Color and Light

Installations

The authors identify as installations such arrangements as permit different perceptual phenomena relating to color and light to be placed in a spatial context so that they can be experienced by the



Interactions of Color and Light, Fig. 1 Color-light keyboard installation; project: staging LED-Color. It is a 25-m-long spatial installation, which consists of 15 large-scale wall panels painted on one side in the same gray and on the other with different, highly pigmented monochrome colors: white, gray, green, red, ultramarine, and black. The surfaces are illuminated with changing colored LED

lighting according to an animated score of color and light. Because the installation can be entered on foot, the dynamic interactions of color and light can be directly experienced in a spatial context. In this way the installation points towards the great scenic and dramaturgical potential of such compositions of color and light for design and the performing arts

senses. They evolve from observations such as can also be made in everyday life. In contrast, the phenomena thus presented are detached from the context of their surroundings and then staged and varied spatially, such that these aspects, or some of them, can be perceived in isolation from it. In this process *hue*, *saturation*, and *brightness* of the colors observed are affected to varying degrees (Fig. 1).

Tools

The term tools refers to materials which facilitate play-based and experimental access to various perceptual phenomena around light and color. These include physical as well as digital tools. They encourage independent or directed experimentation and thus facilitate the acquisition of

experiential knowledge in individualized learning processes (Fig. 2).

Color-Light Toy

The Color-Light Toy is a virtual 3D environment where complex interactions of surface colors and light and various contrast effects can be simulated interactively (Fig. 3).

Conclusion

The authors' research demonstrates the lack of integrative principles and practices in the field of designing using color and light, especially in the context of architecture, *interior architecture*, design, and *scenography*. The research



Interactions of Color and Light, Fig. 2 Color-light playground tool; project: Colour and Light – Materials for a Theory of Colour and Light. This tool stimulates experimentation with dynamically controlled LEDs and color-coated angled sheet metal which can be assembled

in a variety of combinations of color and space. Hue, saturation, and brightness can be altered manually, enabling active experience of the interactions between surface colors and colored light



Interactions of Color and Light, Fig. 3 The color-light toy

demonstrates the high degree of willingness to apply the phenomena of interaction of color and light professionally and as end users. The approach of implementing the abovementioned phenomena into a mediation concept which can be experienced and experimented on, and which is focused on aspects of direct and sensual perception, is being further developed by researchers into color and light at various institutions [10].

Cross-References

- ▶ [Architectural lighting](#)
- ▶ [Color Combination](#)
- ▶ [Complementary Colors](#)
- ▶ [Color Contrast](#)
- ▶ [Environmental Color Design](#)
- ▶ [Light-Emitting Diode, LED](#)
- ▶ [Primary Colors](#)
- ▶ [Pigment, Inorganic](#)

References

1. Colourlight-Center, Zurich University of the Arts. www.colourlight-center.ch (2013)
2. Research Projects at the Zurich University of the Arts: ColourLightLab, supported by SNF/DORE, Swiss National Science Foundation (2005–2006); LED-ColourLab supported by KTI, Swiss Committee for Technology and Innovation (2008–2009); Colour and light – materials for a theory of colour and light, supported by SNF/DORE, Swiss National Science Foundation (2010–2011). www.colourandlight.ch and www.colourlight-center.ch
3. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage Publications, International Educational and Professional Publisher, Thousand Oaks/London/New Delhi (2008)
4. Jonas, W.: Forschung Durch Design. In: Erstes Design Forschungssymposium. Swiss Design Network (2004)
5. Jonas, W.: Design Research and its Meaning to the methodological Development. In: Design Research Now. Michel, R. (ed.). Birkhäuser Verlag AG, Basel/Boston/Berlin (2007)
6. Bachmann, U.: Colours Between Light and Darkness. Niggli Publisher, Sulgen (2006)
7. Bachmann, U.: Colour and Light – Materials for a Theory of Colour and Light. Niggli Publisher, Sulgen (2011)
8. Exposition, LED – Licht und Farbe inszenieren, Gewerbemuseum Winterthur (2008–2009)
9. Böhme, G.: Atmosphäre, p. 15. Suhrkamp, Frankfurt (1995)
10. Institute of Colour and Light, Zurich, www.colourandlight.ch; Institute Integrative Design, Academy of Art and Design Basel, www.masterstudiodesign.ch; Colourlight-Center, Zurich University of the Arts, www.colourlight-center.ch

Interior Lighting

Martine Knoop
Berlin University of Technology, Berlin,
Germany

Synonyms

Indoor lighting; Lighting for indoor spaces; Lighting for interior spaces

Definition

Lighting concepts for interior spaces can be found everywhere. Be it the concept to allow daylight to

penetrate in a building without causing glare or overheating, be it the concept to create appropriate lighting conditions in offices, or be it the specifics of museum lighting to highlight art without damaging it by radiation or heat. Interior lighting is about choosing the appropriate lighting components in a lighting design to fit the requirements of a certain application (examples in Fig. 1).

Overview

Interior lighting can be applied to:

- Support safety from a visual point of view
- Provide adequate lighting conditions for activities in terms of quantity and quality
- Enhance the desired room appearance/lit environment
- Create an atmosphere suitable for the space that is lit
- Enhance health and well-being
- Support an architectural design

The first five aspects are related to the individuals (individual well-being) using the space; the last aspect is related to the space itself. Lighting for safety and lighting for performing activities need to be considered as minimum lighting conditions. This so-called functional lighting is typically referred to in standards, whereas aesthetic lighting, clearly related to atmosphere creation and support of architectural designs, is less specified. Depending on the application, these aspects will vary in their emphasis. For example, atmosphere creation is of higher importance in restaurants and shops; enhancing well-being with lighting is typically taken into consideration in lighting designs for, e.g., schools, hospitals, or office buildings.

Safety needs in normal operation, for example to safeguard not to stumble over obstacles in a corridor or to safely operate industrial machines, are usually fulfilled when lighting conditions to perform activities in the space are realized. Safety in temporary, untypical situations, mainly in case of electrical power loss, needs to be addressed as well. In this case, the lighting needs to guarantee



Interior Lighting, Fig. 1 Examples of interior lighting (All pictures: Philips Lighting)

that the building can be vacated safely without electrical energy in a given time frame. Requirements for the so-called emergency lighting can be found in, for example, EN 1838:1999 Lighting applications – Emergency lighting and CIE S 020/E:2007 Emergency Lighting [1, 2].

Lighting to realize adequate conditions for activities focuses on the lighting needs for visual performance and visual comfort. Aspects to be considered are, for example, protection from glare caused by either daylight or artificial lighting in office rooms, minimum lighting levels to perform visual tasks in almost all applications, color rendering requirements to enable differentiation between colors, which is specifically important, for instance, for color inspection, and in a large number of areas in health-care premises.

Lighting can be applied to enhance room appearance. Several studies indicate that people prefer relatively bright vertical surfaces [3–5]. Research [6] has indicated that the appearance of a space is related to the lighting characteristics of a 40° band of view for an observer (20° above and 20° below the line of sight). The

luminance of this band has a large effect on the perception of dim-bright and uniform-nonuniform. A room is perceived bright from an average luminance of $30\text{--}40\text{ cd/m}^2$ [6, 7]. Additionally, the average scene luminance and ceiling luminance are also relevant parameters in the satisfaction rating of a lighting system [8].

Lighting to create an atmosphere sets the right scene for the purpose of the interior space. In retail and hospitality applications, lighting is an influential parameter to create an inviting space, to stimulate communication or to attract shoppers. Research has been conducted in the 1970s [9] and more recently [6, 10, 11]. It indicates, for example, that to realize an intimate atmosphere, diffuse warm lighting should be applied; a businesslike, neutral atmosphere is created with diffuse cold lighting. Visual clarity is achieved with high workplane and wall luminances. A place works confined, narrowing when direct overhead lighting is applied at low lighting, without illuminating the walls. Nonuniform, moderate light levels and nonuniform wall lighting induce pleasantness.

Light is important to people's health and well-being. It can affect people's physiological and psychological state. Although already researched in the 1980s and 1990s, it received a broader interest after the discovery of a new photoreceptor (ipRGC) in the retina of the eye [12]. When light reaches these cells, signals are sent to the central biological clock (suprachiasmatic nuclei – SCN), which regulates circadian and circannual rhythms of a large variety of bodily processes. This influences, for example, sleepiness and alertness. The lighting community is still exploring how light can be used to support the healing process in hospitals, to improve learning conditions for pupils in schools, or to improve shift workers' sleep quality during the day and alertness during the night. Current research indicates that higher illuminances and a wider range of color temperatures than typically specified in standards can stimulate people and enhance their well-being (► [Non-visual lighting effects and their impact on health and well-being](#)).

Lighting can also be used to support the architectural design, by using coves and large lit surfaces to define the space, as well as by highlighting architectural details, such as columns, arches, and galleries (► [Architectural Lighting](#)).

Lighting Quality

Although no precise definition for lighting quality exists [13], lighting quality is typically related to the degree of which the lighting conditions meet needs of users of the space. A large field investigation in office buildings [8] has indicated that the effect of lighting that is perceived of higher quality can be considerable. The study shows that, through environmental satisfaction, lighting quality can affect job satisfaction, organizational commitment, and turnover, as well as health problems and self-assessed performance.

In order to realize good quality lighting, the lighting solution needs to fulfill the requirements set by/coming from the various target groups, for example, the user, the owner, and/or the designer. The quality of the lighting solution is determined

by the fulfillment of the set lighting purpose(s) (see before) as well as the economic, environmental, and design aspects of the solution, such as style, integration into architecture, appropriateness of appearance, form, and size (from [5, 14]). Quality lighting can be achieved by the right choice of lighting components (light sources, luminaires, and lighting control) in an appropriate lighting design.

This part on interior lighting will focus on quality aspects of lighting conditions for activities, as mentioned above. Typically, requirements for this lighting can be found in standards and recommendations (e.g., [13, 15–18]), which ensure that the user in the application is able to comfortably perform basic visual work. The following aspects are usually seen as minimum considerations:

Task Visibility

In order to have appropriate task visibility in indoor spaces, a minimum required (average) illuminance level is given in standards and recommendations. Just as an example, in the European standard EN 12464-1 [16], a minimum average illuminance of 500 lx is given for writing, typing, reading, and data processing in office buildings; color inspection in chemical, plastics, and rubber industry requires 1,000 lx. A minimum uniformity of illumination (E_{\min}/E_{av}) is demanded to prevent frequent reaccommodation of the eyes. In the European Standard the required uniformity lays between 0.4 (for an easy task or in the surrounding area of a task) and 0.7 (difficult task).

In order to prevent errors, fatigue, or accidents, one should ensure that reflections in specular surfaces (veiling reflections) are not present. These reflections, which partially or totally obscure the details to be seen by reducing the contrast, will diminish task visibility. Maximum average luminance values are given to avoid reflected glare on computer screens [16]. When it comes to reflections in glossy material (e.g., magazine) or glass (e.g., vitrine or pictures) (see Fig. 2), design rules that define the appropriate position of the light sources (outside the so-called offending zone) can be applied to prevent veiling reflections (see, e.g., [19]).



Interior Lighting, Fig. 2 Examples of veiling reflections

Visual Comfort

When too high luminance differences are offered within the field of view, glare can be caused. Discomfort glare is the type of glare that does not directly reduce the ability to see an object, but it produces a sensation of discomfort and might affect performance in the longer run. Glare which impairs vision is called disability glare. In indoor lighting it is assumed that disability glare does not occur, when both shielding angles for bright light sources and discomfort glare restrictions are applied. The degree of discomfort glare can be assessed with the Unified Glare Rating or Visual Comfort Probability. Standards and recommendations provide maximum (UGR) and minimum (VCP) values to ensure comfortable lighting conditions in a specific application. For example, for offices the required UGR is typically 19, the recommended VCP value is 70. For situations with VDT tasks, a VCP value of 80 is recommended (► [Glare](#)).

Whereas the task visibility and visual comfort aspects are related to the lighting solution in a specific application, standards and recommendations refer to minimum requirements for characteristics of light sources as well. Both flicker and color rendering are qualities to be assessed, as they might hamper visual performance when requirements are not met.

Visible, but also invisible, flicker can trigger headaches, fatigue, and distraction. It has been shown to reduce performance. If it results in stroboscopic effects and objects appear to move discretely rather than continuously, it can lead to critical situations, e.g., in industry when

interacting with rotating or reciprocating machinery. The Critical Flicker Frequency and the Flicker Index are used in the past to assess flicker, but research indicates that not all relevant parameters are covered in both measures [20–22]. New metrics are looked into, to be used to predict both flicker and stroboscopic effects for all light sources, including LED (e.g., [23]).

For visual performance, colors in the environment shall be rendered naturally and correctly. The general color rendering index (R_a) of a light source indicates how well this light source reproduces colors of an object in comparison with a reference (ideal or natural) light source. The maximum value of R_a is 100, and the required value is depending on the application. For example, according to the European Standard EN 12464-1 [16], indoor circulation areas require a minimum R_a of 40, offices a value above 80, and areas for fabric control in textile manufacturing or examination and treatment rooms a minimum R_a of 90. The color rendering index of a light source can be obtained from the manufacturer, the package, or the lamp itself (► [CIE Color-Rendering Index](#)).

A number of quality measures go beyond task performance. Generally, these are listed in standards or recommendations as well, with proposed minimum levels to achieve a higher lighting quality, to benefit human health and well-being (e.g., in [16, 18]). Examples are:

- Color appearance, which is referring to the color of the light, quantified by its correlated color temperature (T_{CP}). For artificial lighting, it ranges typically from 2,700 K (warm white

light) to 6,500 K (cool white light), with a maximum of 17,000 K. The choice for a certain color appearance depends on the application, the atmosphere to be created, the colors of furniture and room partitions, the illuminance level, climate, and culture (► [Kruithof, preferred colour temperature](#)).

- **Modelling of faces for visual communication and recognition of faces.** In order to achieve this, the lighting in a room should not be too directional nor too diffuse. Design guidelines propose to use as little as 20 % of the total illuminance from a diffuse or reflected light source and avoid strongly directional lighting (e.g., [17, 18, 24]).

It needs to be noted that the abovementioned quality criteria typically focus on minimum requirements to assure visual performance and visual comfort. As indicated before, room appearance is of great importance for the user's well-being. The standards, mainly focusing on minimum requirements for task illumination, will not sufficiently address this. Although minimum required vertical illuminance is mentioned in the IES Handbook [14] and the revised EN 124641:2011 [16], the needed lighting levels for quality lighting are higher. As LED solutions can be tightly controlled optically, thus offering the possibility to realize appropriate task illuminances without spill light to the walls, room appearance seems to be a relevant topic for future lighting solutions.

Interior Lighting Design

Interior lighting is about choosing the lighting components in a lighting design to fit the requirements of a certain application.

As indicated before, a lighting design consists of functional lighting and/or aesthetic lighting. Typical application areas with a focus on functional lighting are industry (e.g., assembly work, storage) and office (e.g., open plan office, cell office, meeting room, corridors). Typical application areas with additional aesthetic lighting are retail (e.g., department stores, boutiques), hotels

(e.g., entrance areas, reception areas), and restaurants.

Resulting, interior lighting is often built up from layers of light. Depending of the application, one or more layers are applied, typically distinguishing between

- **General lighting:** lighting providing (horizontal) illumination over the total area of the room with a certain degree of uniformity. General lighting is often offered in a regular grid of point, line, or area light sources (see Fig. 3) (► [Coefficient of utilization, lumen method](#), ► [Diffuse Lighting](#)).
- **Task lighting:** additional lighting to general lighting in the area in which the task will be performed. Task lighting is used to fulfill the requirements, without having to illuminate the entire room to that specific level, in order to save energy (see Fig. 4) (Task Lighting).
- **Accent lighting:** directional lighting to emphasize a particular object or to draw attention to a part of the field of view (see Fig. 5).
- **Decorative lighting:** the lighting itself provides a point of interest or attraction in an interior (see Fig. 6).
- **Architectural lighting:** lighting to form or underline the architecture of the space (see Fig. 7) (► [Architectural Lighting](#)).

All layers have specific requirements, which the lighting components have to fulfill.

Lighting Components for Interior Lighting

Lighting for indoor places can be provided by daylight, artificial lighting, or a combination of both. In this chapter “interior lighting,” only artificial lighting is discussed. As interior lighting is about choosing the lighting components in a lighting design to fit the requirements of a certain application, the lighting components will be briefly discussed below.

Interior lighting is used in a wide spread of applications with a wide range of requirements. Resulting, a large variety of light sources and



Interior Lighting, Fig. 3 Examples for general lighting: point (Two pictures on the top), line, and area light sources (First picture: Philips Lighting)

luminaire types are applied in interior lighting. The most relevant characteristics that determine the applicability of a light source in a certain application are luminous flux, efficacy, luminance, color rendering, correlated color temperature, lifetime, run-up time, reignition time and form factor of the light source. In the design process the application requirements on these aspects need to be specified, in order to select the appropriate light source.

To give an impression, without being conclusive, one can roughly say that halogen lamps are generally applied in aesthetic lighting, in accent lighting, and in decorative lighting (► [Incandescent Lamp](#)). Fluorescent lamps can be found in applications with a focus on functional lighting, in both general lighting and task

lighting, such as industry, health care, educational buildings, and office. Next to that fluorescent lamps are used in architectural lighting solutions, such as coves and luminous ceilings (► [Tubular and compact fluorescent lamp](#)). High-intensity discharge lamps (e.g., ► [Metal Halide Lamp](#)) can be found in general lighting solutions as well as accent lighting solutions, in, for example, industry and retail applications. LEDs have been successfully introduced in interior lighting as well and can be found in both functional and aesthetic lighting solutions. The diversity of LEDs allows for usage of this light source in general, task, accent, decorative, as well as architectural lighting (► [Light emitting diode, LED](#)).

With respect to the luminaires (► [Luminaires](#)), a similar specification of requirement needs to be



Interior Lighting, Fig. 4 Examples for task lighting (All pictures: Philips Lighting)

made. Based on the chosen required lighting concept (functional and aesthetic, from general lighting to architectural lighting), the following characteristics of the luminaire can support luminaire selection: the light distribution (luminous intensity distribution – ► [Light Distribution](#)), efficiency (reflected in light output ratio), IP class, form factor, and type (recessed, surface mounted, suspended, wall mounted, free floor standing, or desk luminaire). The variety is large, serving the broad bandwidth of requirements in interior lighting.

In the overall lighting design, lighting controls need to be considered. Reasons to use lighting controls in interior lighting are to:

- Make use of potential energy savings, such as time control, occupancy control, and controls that respond to the daylight availability in the room (daylight supply control)
- Increase user comfort and well-being, such as personal control, glare control, and controls that offer lighting settings for specific tasks (task demand control)
- Support building appearance, ambience, and company representation, for example, using scene setting

Lighting controls can enhance the quality of a lighting installation. Once more, the specification will depend on the application. The use of



Interior Lighting, Fig. 5 Example for accent lighting (Picture: Philips Lighting)



Interior Lighting, Fig. 6 Examples for decorative lighting (Second picture, Philips Lighting; photo credits, Cinimod Studio – third picture, Philips Lighting; photographer, Leon Verlaek)

occupancy control is, for example, suitable in areas where people generally stay a relatively short period or do not feel responsible for switching off the lighting, such as bathrooms, stair cases, and

warehouse aisle. Daylight supply control is applied in situations with sufficient daylight and an appropriate choice, for example, in open plan offices and hospital wards (► [Lighting Controls](#)).



Interior Lighting, Fig. 7 Examples for architectural lighting (Second and third picture: Philips Lighting)

The combination of components – light source, luminaire, and controls – will determine the overall quality of the lighting.

Concluding

Interior lighting serves a number of objectives. Each application requires a specific approach, typically build up from a few layers of light, fulfilling demands of the user of the space and its architecture. Good quality lighting respects these demands, as well as the economic, environmental, and design aspects. A combination of suitable light sources, luminaires, and lighting controls in an appropriate design will offer good quality lighting in interior spaces.

Cross-References

- ▶ [Architectural Lighting](#)

- ▶ [CIE Color-Rendering Index](#)
- ▶ [Coefficient of Utilization, Lumen Method](#)
- ▶ [Diffuse Lighting](#)
- ▶ [Glare](#)
- ▶ [Incandescent Lamp](#)
- ▶ [Kruithof Curve](#)
- ▶ [Light Distribution](#)
- ▶ [Light-Emitting Diode, LED](#)
- ▶ [Lighting Controls](#)
- ▶ [Luminaires](#)
- ▶ [Metal Halide Lamp](#)
- ▶ [Non-visual Lighting Effects and their Impact on Health and Well-being](#)
- ▶ [Tubular and Compact Fluorescent Lamp](#)

References

1. CEN: EN 1838:1999 Lighting Applications – Emergency Lighting. Published by the European national standards institutes CEN (1999)
2. CIE: CIE S 020/E:2007 Emergency Lighting. CIE, Vienna (2007)

3. Newsham, G.R., Marchand, R.G., Veitch, J.A.: Preferred surface luminances in offices, by evolution. *J. Illum. Eng. Soc.* **33**(1), 14–29 (2004)
4. Van Ooyen, M.H.F., Van De Weijert, J.A.C., Begemann, S.H.A.: Preferred luminances in offices. *J. Illum. Eng. Soc.* **16**(2), 152–156 (1987)
5. Veitch, J.A.: Lighting for high-quality workplaces (NRCC-47631 – <http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=rtdoc&an=20377130&article=0&fd=pdf>. Accessed 10 August 2012), also published in Clements-Croome, D. (ed.). *Creating the Productive Workplace*, pp. 206–222, 2nd edn. Taylor & Francis, London (2006)
6. Loe, D.L., Mansfield, K.P., Rowlands, E.: Appearance of lit environment and its relevance in lighting design: experimental study. *Light. Res. Technol.* **26**(3), 119–133 (1994)
7. Newsham, G.R., Arsenault, C., Veitch, J.A., Tosco, A.M., Duval, C.: Task lighting effects on office worker satisfaction and performance, and energy efficiency. *Leukos* **F1**(4), 7–26 (2005)
8. Veitch, J.A., Newsham, G.R., Mancini, S., Arsenault, C.D.: Lighting and office renovation effects on employee and organizational well-being NRC report IRC – RR-306 <http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=rtdoc&an=20374532&article=0&fd=pdf> (2010). Accessed 10 Aug 2012
9. Flynn, J., Spencer, T.J., Martyniuk, O., Hendrick, C.: Interim study of procedures for investigating the effect of light on impression and behavior. *J. Illum. Eng. Soc.* **3**, 94–96 (1973)
10. Vogels, I.: Probing experience atmosphere metrics: a tool to quantify perceived atmosphere. *Philips Res. Book Ser.* **8**(1), 25–41 (2008)
11. Custers, P.J.M., De Kort, Y.A.W., IJsselstein, W.A., De Kruijf, M.E.: Lighting in retail environments: atmosphere perception in the real world. *Light. Res. Technol.* **42**(3), 331–343 (2010)
12. Berson, D.M., Dunn, F.A., Takao, M.: Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**(5557), 1070–1073 (2002)
13. Veitch, J.A. (eds.): CIE: Proceedings of the first CIE Symposium on Lighting Quality (1998) CIE Central Bureau, Vienna
14. DiLaura, D., Houser, K., Mistrick, R., Steffy, G. (eds.): IESNA: Lighting Handbook: Reference & Application, 10th edn. IESNA, New York (2011)
15. CIE: CIE S 008/ISO 8995-1: Lighting of work places – part 1: indoor (2002)
16. CEN: EN 12464-1:2011 Light and lighting – lighting of work places – part 1: indoor work places. Published by the European National Standards Institutes CEN (2011)
17. IESNA: ANSI/IESNA RP-1-04 American National Standard Practice for Office Lighting. IESNA, New York (2004)
18. IESNA: IES DG-18-08 A Guide to Designing Quality for People and Buildings. IESNA, New York (2008)
19. Ganslandt, R., Hofmann, H.: Handbook of Lighting Design. ERCO Edition, Verlag Vieweg ISBN 3-528-08895-8. http://www.erco.com/download/data/30_media/20_handbook/en_erco_lichtplanung.pdf (1992). Accessed 10 Aug 2012
20. Vogels, I., Sekulovski, D., Perz, M.: Visible artefacts of LEDs conference proceedings of the 27th session of the CIE, Sun City, 10–15 Jul 2011 **1**(1):42–51 (2011)
21. Bullough, J.D., Sweater Hickcox, K., Klein, T.R., Narendran, N.: Effects of flicker characteristics from solid-state lighting on detection, acceptability and comfort. *Light. Res. Technol.* **43**(3), 337–348 (2011)
22. Bullough, J.D., Sweater Hickcox, K., Klein, T.R., Lok, A., Narendran, N.: Detection and Acceptability of Stroboscopic Effects from Flicker Lighting Research and Technology first published on Oct 11 2011 as doi: 10.1177/1477153511414838 (2012)
23. ASSIST Alliance for Solid-State Illumination Systems and Technologies: ASSIST Recommends ... Flicker Parameters for Reducing Stroboscopic Effects from Solid-State Lighting Systems 11(1) Lighting Research Center, Troy <http://www.lrc.rpi.edu/programs/solidstate/assist/pdf/AR-Flicker.pdf> (2012). Accessed 10 Aug 2012
24. Boyce, P.R., Miller, N.J., Wilwol, K.: Lighting for good facial modelling <http://www.lrc.rpi.edu/resources/pdf/68-1999.pdf> (1999). Accessed 10 Aug 2012

Interpolation of Spectral Data

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

Synonyms

Subsampling

Definition

Spectral data can be interpolated using Sprague interpolation when the original data are evenly spaced. This involves fitting a fifth-order polynomial between two points P_i and P_{i+1} thus:

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 \quad (1)$$

which can be used to interpolate between the points P_i and P_{i+1} and where x is the proportional distance between P_i and P_{i+1} of the point to be evaluated. However, the polynomial is constrained so that the gradients and curvature are consistent with the points P_{i-2} , P_{i-1} and P_{i+2} , P_{i+3} (the two points either side of the points that are being interpolated between). This constraint results in the coefficients of equation being easily calculated from the points P_{i-2} – P_{i+3} [1] thus:

$$\begin{aligned}
 a_0 &= P_i \\
 a_1 &= (2P_{i-2} - 16P_{i-1} + 16P_{i+1} - 2P_{i+2})/24 \\
 a_2 &= (-P_{i-2} - 16P_{i-1} - 30P_i + 16P_{i+1} - P_{i+2})/24 \\
 a_3 &= (-9P_{i-2} + 39P_{i-1} + 70P_i + 66P_{i+1} - 33P_{i+2} \\
 &\quad + 7P_{i+3})/24 \\
 a_4 &= (13P_{i-2} - 64P_{i-1} + 126P_i - 124P_{i+1} \\
 &\quad + 61P_{i+2} - 12P_{i+3})/24 \\
 a_5 &= (-5P_{i-2} + 25P_{i-1} - 50P_i + 50P_{i+1} - 25P_{i+2} \\
 &\quad + 5P_{i+3})/24
 \end{aligned} \tag{2}$$

However, for N points P_i where $i = 1:N$ the method cannot be directly used to interpolate between P_1 and P_2 nor between P_{N-1} and P_N because the additional points on either side do not exist. The CIE recommends a method for extrapolating four additional points, two at shorter wavelength intervals than available (P_{-1} and P_0) and two at longer wavelength intervals than available (P_{N+1} and P_{N+2}) thus:

$$\begin{aligned}
 P_0 &= (884P_1 - 1960P_2 + 3033P_3 - 2648P_4 \\
 &\quad + 1080P_5 - 180P_6)/209 \\
 P_1 &= (508P_1 - 540P_2 + 488P_3 - 367P_4 \\
 &\quad + 144P_5 - 24P_6)/209 \\
 P_{N+1} &= (-24P_{N-5} + 144P_{N-4} - 367P_{N-3} \\
 &\quad + 488P_{N-2} - 540P_{N-1} + 508P_N)/209 \\
 P_{N+2} &= (-180P_{N-5} + 1080P_{N-4} - 2648P_{N-3} \\
 &\quad + 3033P_{N-2} - 1960P_{N-1} + 884P_N)/209
 \end{aligned} \tag{3}$$

Background

CIE tristimulus values can be calculated from spectral reflectance factors, spectral illuminant relative power distributions, and the CIE color-matching functions (CMFs). However, it is sometimes the case that the CMFs, for example, are available at intervals of 5 nm, but the spectral reflectance factors are available at intervals of 10 nm. Interpolation of spectral data may be required and interpolation methods therefore need to be considered.

A straight line can be drawn to fit exactly through any two points and a parabola through any three points. More generally, an n th-degree polynomial can pass through any $n + 1$ points. Imagine we wish to fit some spectral data $P(\lambda)$ at four wavelengths, λ_1 – λ_4 . We can write:

$$\begin{aligned}
 P(\lambda_1) &= a_1\lambda_1^3 + a_2\lambda_1^2 + a_3\lambda_1 + a_4 \\
 P(\lambda_2) &= a_1\lambda_2^3 + a_2\lambda_2^2 + a_3\lambda_2 + a_4 \\
 P(\lambda_3) &= a_1\lambda_3^3 + a_2\lambda_3^2 + a_3\lambda_3 + a_4 \\
 P(\lambda_4) &= a_1\lambda_4^3 + a_2\lambda_4^2 + a_3\lambda_4 + a_4
 \end{aligned} \tag{4}$$

which represents four simultaneous equations and four unknowns. In terms of linear algebra, Eq. 4 can be represented by Eq. 5:

$$\mathbf{A}\mathbf{p} = \mathbf{q} \tag{5}$$

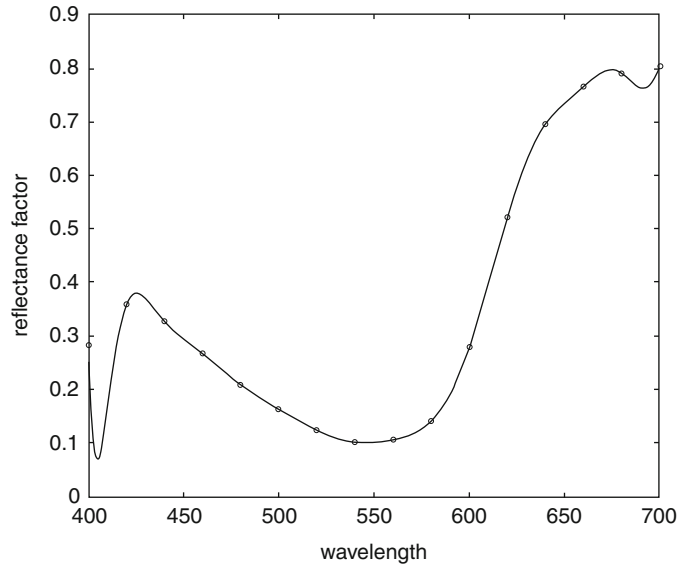
where \mathbf{q} is a 4×1 column matrix of reflectance values, \mathbf{p} is a 4×1 column matrix containing the coefficients a_1 – a_4 , and \mathbf{A} is a special 4×4 matrix known as the Vandermonde matrix. In the example shown as Eq. 4, the Vandermonde matrix would be constructed with the entries thus:

$$\begin{pmatrix} \lambda_1^3 & \lambda_1^2 & \lambda_1 & 1 \\ \lambda_2^3 & \lambda_2^2 & \lambda_2 & 1 \\ \lambda_3^3 & \lambda_3^2 & \lambda_3 & 1 \\ \lambda_4^3 & \lambda_4^2 & \lambda_4 & 1 \end{pmatrix}$$

The polynomial in Eq. 4 is referred to as the Lagrange polynomial. It is trivial to solve Eq. 5 for the coefficients in \mathbf{p} using MATLAB thus

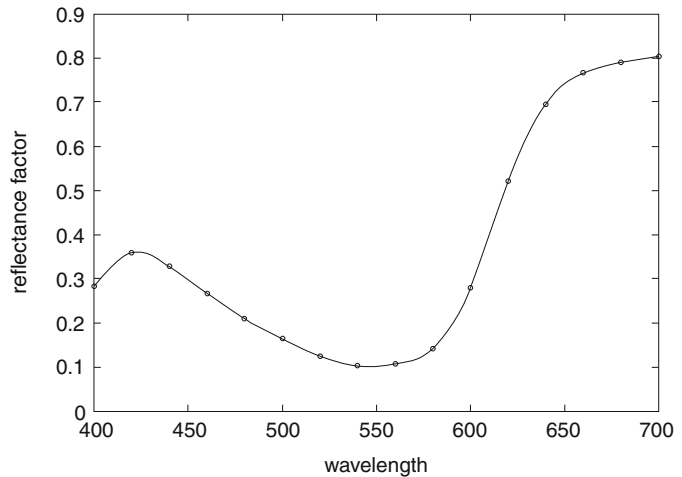
Interpolation of Spectral Data, Fig. 1

Example of interpolating 20-nm spectral data (symbols) using a Lagrangian polynomial (solid line)



Interpolation of Spectral Data, Fig. 2

Example of interpolating 20-nm spectral data (symbols) using the Sprague method (solid line)



$\mathbf{p} = \mathbf{A} \backslash \mathbf{q}$ or, alternatively, using MATLAB's *polyfit* function. However, the use of a Lagrangian polynomial in this way is not the best way to interpolate spectral data; the resulting polynomial may lead to wide excursions between the points [1].

Figure 1 illustrates a Lagrangian polynomial that was fitted to spectral data at intervals of 20 nm and interpolated at intervals of 1 nm. The interpolation is likely to be inadequate at the short and long ends of the spectrum.

One way to address this problem is to use a family of polynomials (each of which fits a

relatively small number of points) rather than trying to fit the whole spectrum with a single polynomial. Repeated application of a low-order polynomial to a restricted range of the data is an example of a "piecewise method." CIE Publication Number 15.2 recommends that interpolation of spectra be carried out using a third-degree polynomial from neighboring data within twice the measurement interval [2]. This means that a Lagrange interpolation formula should be used for four data points (two either side of the point to be interpolated). A later CIE publication [3]

recommends one of four interpolation methods: (1) the third-order polynomial interpolation (Lagrange) from the four neighboring data points around the point to be interpolated, (2) a cubic spline interpolation formula, (3) a fifth-order polynomial interpolation from the six neighboring data points around the point to be interpolated, or (4) a Sprague interpolation.

Current Status

Sprague interpolation was recommended following a study by a CIE working group [1], and the method for doing this is defined by Eqs. 1 and 2. Figure 2 illustrates some spectral data at intervals of 20-nm Sprague interpolation at intervals of 1 nm.

The method can be used equally for reflectance factors or for illuminant spectral power distributions.

Cross-References

► [CIE Tristimulus Values](#)

References

1. CIE Publication No. 167: Recommended practice for tabulating spectral data for use in colour, Central Bureau of the Commission Internationale d'Eclairage (2005)
2. CIE Publication No. 15.2: Colorimetry, 2nd edn, Central Bureau of the Commission Internationale d'Eclairage (1986)
3. CIE Publication No. 15: Colorimetry, 3rd edn, Central Bureau of the Commission Internationale d'Eclairage (2004)

Interreflection

► [Global Illumination](#)

Intrinsically Photosensitive Retinal Ganglion Cells

► [Melanopsin Retinal Ganglion Cells](#)

Iodine Lamps

► [Halogen Lamp](#)

ipRGCs

► [Melanopsin Retinal Ganglion Cells](#)

Iridescence (Goniochromism)

Richard J. D. Tilley
Queen's Buildings, Cardiff University,
Cardiff, UK

Synonyms

[Goniochromatism](#); [Irisation](#); [Labradorescence](#); [Opalescence](#); [Pearlescence](#); [Polychromatism](#); [Schiller](#)

Definition

Iridescent is the adjective used to describe objects or surfaces that show bright, metallic-looking, shining colors that change noticeably as the position of the observer changes. Iridescence is the noun that describes the colors that arise from an iridescent object.

Description and Examples

The root cause of iridescence lies in the (coherent) scattering of white incident light by transparent or semitransparent surface and subsurface structures, as was stated by Robert Hooke in his description of iridescent peacock feathers (*Micrographia* 1665). He also gave a technique for distinguishing iridescent colors from those that are due to pigmentation – when the color is wetted, it will disappear if its origin is refraction and reflection.

For color to be observed, the substructures have to vary on a size scale equivalent to about half the wavelength of visible light. Moreover, iridescent reflections are often strongly polarized.

Among inanimate examples, colors of a soap film or films of oil on water can be thought of as archetypal [1]. A number of minerals such as labradorite and opal show iridescence and are used as gemstones. Materials that oxidize to produce thin transparent films overlaying a dark background also show strong iridescence, typified by silicon carbide (SiC) and bornite (Cu_5FeS_4) also called peacock ore. Dichroic filters, which are made by coating glass disks with multilayers of oxides, show iridescent colors. Dichroic glass, made in the same way, is used by artists. Iridescent gift bags and tags are commonplace, and iridescence is utilized in car paints, cosmetics, and other decorative and artistic objects. The iridescent reflected colors from CD and DVD storage disks and holograms used as security labels are familiar (Fig. 1a–c) [2–4].

Living organisms display an enormous variety of iridescent colors [5–10]. Many creatures are able to see near-ultraviolet radiation, and iridescence in the ultraviolet is of importance in the wild even though it is undetected by humans. Thin-film iridescence similar to that in soap films can be seen on close examination of the transparent wings of flies, bees and wasps, beetles, and many other flying insects. More obviously iridescent coloring is characteristic of many groups of birds (feathers), beetles (exoskeleton), and butterflies (wing scales), although iridescent colors are not restricted to these groups. In these cases the visual effect is often enhanced by backing the iridescent structures with a melanin-related black pigment so that stray reflections do not diminish the vividness of the perceived colors. Pearls and a number of shells show iridescence, notably that of the New Zealand paua, *Haliotis iris* (Fig. 1d–f).

In biological areas there is great interest in explaining the evolution and adaptive advantage of iridescent coloration. In photonics there is considerable effort expended in mimicking natural iridescence by artificially copying biological examples.

Mechanisms

The interaction of light with an iridescent surface alters the phases of the various components of the scattered radiation, allowing constructive and destructive interference to occur. Because phase shift is sensitive to the optical path through the membranes intercepted, the colors observed change dramatically with variation in viewing angle. The mechanism of color production is often described in terms of reflection or diffraction, and although such a clear cut division is somewhat artificial, this provides a starting point for discussion.

Reflection

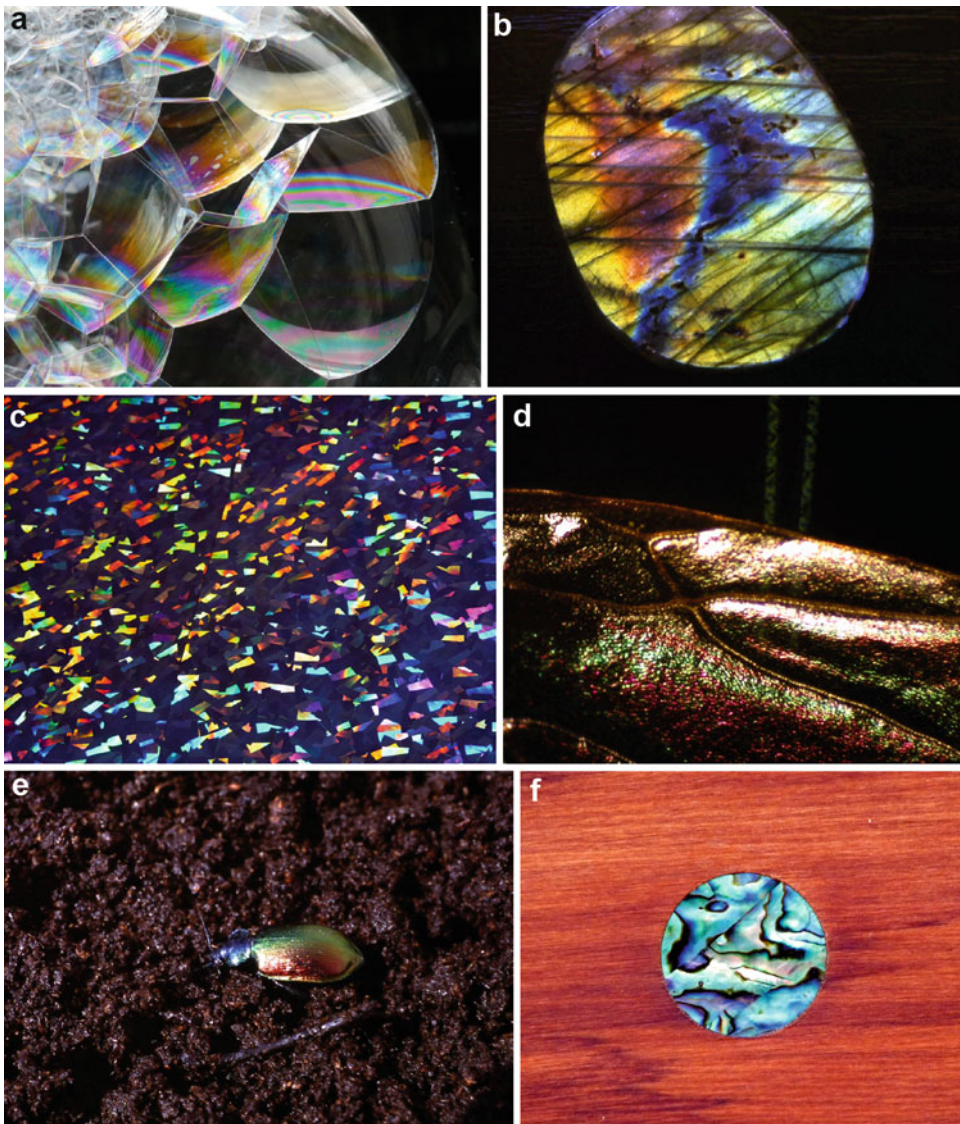
The simplest model for iridescence is the reflection of white light from a single transparent thin film, the cause of color in soap films and transparent insect wings (Fig. 1a–d). The phase shift that leads to interference and hence color production is between, in the simplest case, that part of the incident beam that is reflected off the top surface and that part that is transmitted through the film and reflected off the bottom surface (Fig. 2). At *normal incidence* it is found that a monochromatic incident beam of wavelength λ reflected back from a thin film of refractive index n and thickness d in air or on a substrate of lower refractive index than the film will show constructive interference and so appear *bright* when

$$2nd = \left(m + \frac{1}{2}\right)\lambda \quad (m = 0, 1, 2, 3 \dots)$$

When the light is incident at a small angle to the normal, it is possible to use the *approximate* relationship:

$$2nd \cos \theta = \left(m + \frac{1}{2}\right)\lambda$$

If the film is on a substrate of higher refractive index, the reflected light will show destructive interference and appear dark under the same conditions. Similarly, light reflected from a film in air or on a substrate of lower refractive index than the



Iridescence (Goniochromism), Fig. 1 Iridescence: (a) soap film; (b) labradorite; (c) wrapping paper; (d) ant wing; (e) scarab beetle; (f) paua shell decoration (Photographs R J D Tilley)

film will show destructive interference and so appear *dark* when

$$2nd = m\lambda \quad (m = 1, 2, 3 \dots)$$

If the film is on a substrate of higher refractive index, the reflected light will show constructive interference and appear bright under the same conditions. When the light is incident at a small

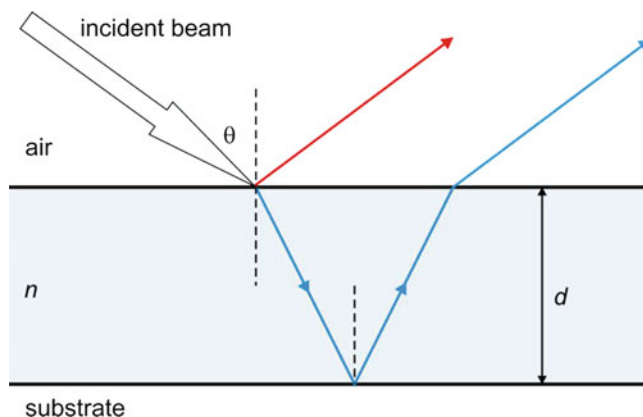
angle to the normal, it is possible to use the *approximate* relationship:

$$2nd \cos \theta = m\lambda$$

Note that the equations for viewing at an angle do not take polarization into account and must be employed with caution.

Iridescence**(Goniochromism),**

Fig. 2 Production of thin-film iridescent colors due to phase difference between rays reflected from top (*red*) and bottom (*blue*) surfaces



To determine the appearance of a film illuminated with white light, the reflectivity has to be summed across all of the wavelengths in the visible. The result is that some colors will be suppressed because that particular wavelength will correspond to destructive interference, while those that correspond to constructive interference will be augmented so that the film will appear colored. The colors seen depend upon film thickness and cycle through several series or *orders*. The $\cos \theta$ angle dependence given above indicates that, as the viewing angle moves away from perpendicular, the color observed moves towards shorter wavelengths: a *blue shift*. Because the color observed depends upon the film thickness and angle of observation, in the case of unstable films such as soap bubbles, the colors swirl and change as the film thins and flexes.

Many decorative films and paints are made by using thin-film iridescence. On bags and gift tags, this is often a thin polymer film, sometimes containing a dye, overlaying a reflective substrate (Fig. 1c). Metallic car paints often contain flakes of a mineral such as mica overlaid with a thin film of a transparent oxide such as TiO_2 .

A similar effect, but generally with greater brightness, is found if the reflecting object is composed of multiple layers (often called multilayer stacks or distributed Bragg reflectors) consisting of alternating lamellae of higher and lower refractive index. The mineral labradorite (Fig. 1b) contains alternating lamellae of feldspars of slightly different constitution, thus giving lamellae with different refractive indices. The iridescent colors

of many scarab beetles (Fig. 1e) are due to a multilayer structure consisting of chitin and air layers, while in paua shells (Fig. 1f) the color arises in layers of nacre, a composite of aragonite (CaCO_3) plates and proteins. The reflectivity of multilayers depends upon the thickness of the layers, their refractive indices, and the angle of view. Relatively simple formulas have been derived for the reflectivity of stacks where two layer types are involved, each of which is $\lambda/4$ thick (called a quarter wave stack), viewed at normal incidence. The peak reflectivity of such a stack with refractive indices n_1 and n_2 , of thicknesses d_1 and d_2 , occurs at a wavelength given by $2(n_1d_1 + n_2d_2)$. The optical behavior of more complicated arrangements, such as those found in dichroic interference filters and dielectric mirrors, can be calculated using computer software.

There are a number of differences between the colors exhibited by multilayers compared to those from single thin films. In particular, the blueshift observed as the observation angle moves away from normal to the stack is much less pronounced compared to single films.

Diffraction

Colors arise when white light is diffracted by objects with a dimension similar to that of visible wavelengths. For example, a single small hole, disk, speck, or droplet gives rise to diffracted beams in the form of nested cones with maxima and minima given by

$$\sin \theta = m\lambda/d$$

Iridescence**(Goniochromism),**

Fig. 3 Irisation shown by high clouds (Photograph R J D Tilley)



where d is the diameter of the object and m takes values of 0, the straight-through beam; 1.220, first minimum (dark ring); 1.635, first bright ring (called the first-order diffracted beam); 2.333, second minimum (second dark ring); 2.679, second bright ring (second-order diffracted beam); and so on. For orders other than 0, the angle of scattering is wavelength dependent so that each diffracted cone will consist of a spectrum, with violet diffracted least and red the most.

A random distribution of scattering centers of differing sizes will not show iridescence as the orders overlap and cancel. However, if the collection is made up of similar-sized objects, iridescence can result. This occurs on rare occasions when high clouds are made up of uniformly sized droplets and viewed at an angle of approximately $10\text{--}45^\circ$ from the direction of the sun, a phenomenon called *irisation* (Fig. 3). The fact that strong iridescent colors are only rarely seen is likely to be due to the necessity of having uniform drop sizes with diameters of the order of 0.003 mm. Similarly, iridescence arising from small evenly sized spots and droplets that form on mirrors can sometimes be seen.

A surface which contains a series of equally spaced parallel grooves or slits with a spacing of about the wavelength of visible light forms a *one-dimensional diffraction grating*. Light incident upon such a grating of spacing d (either in transmission or reflection) will be scattered strongly in directions given by the *grating equation*:

$$d(\sin \theta_i + \sin \theta_m) = m\lambda$$

Where θ_i and θ_m are the angles of incidence and diffraction of the order m , and θ_m is positive when the diffracted and incident beam lie on the same side of the grating normal and negative when they lie on opposite sides of the normal. Far from the grating the (Fraunhofer) diffraction pattern will consist of a row of bright spots called *principal maxima*. The spacing of these spots is proportional to $1/d$ and lies along a line perpendicular to the grating grooves. They are separated by much smaller secondary maxima which will not concern us here as these do not give rise to strong iridescent effects. The zero-order, $m = 0$, diffracted beam is equivalent to an undeviated beam (transmission) or specular reflection (reflection). For the other orders, the angle of scattering, θ_m , is wavelength dependent so that white light will be spread out into a spectrum, with violet diffracted least and red the most (Fig. 4). The spectra produced by such gratings are intense, giving rise to marked iridescence (Fig. 1c).

Scattering centers can be distributed across a surface to form a *two-dimensional* grating. Such a grating can consist of an array of apertures arranged in an ordered pattern in an opaque screen, an ordered set of pits or protuberances arranged on a plane, and so on, most easily visualized for a rectangular array (Fig. 5a). Each row on the pattern acts as a one-dimensional grating

Iridescence**(Goniochromism),**

Fig. 4 Iridescent color formation by a one-dimensional reflection diffraction grating

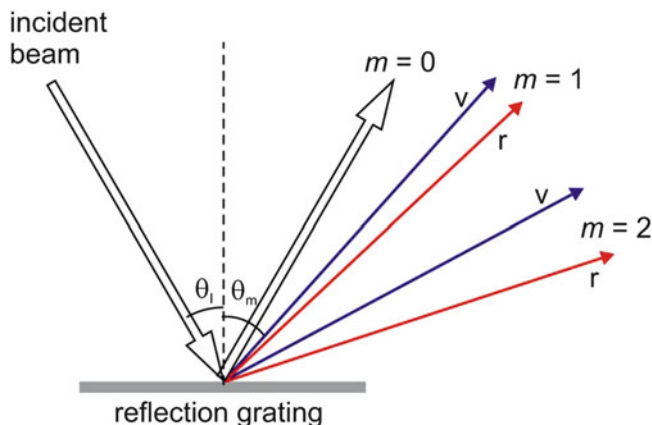
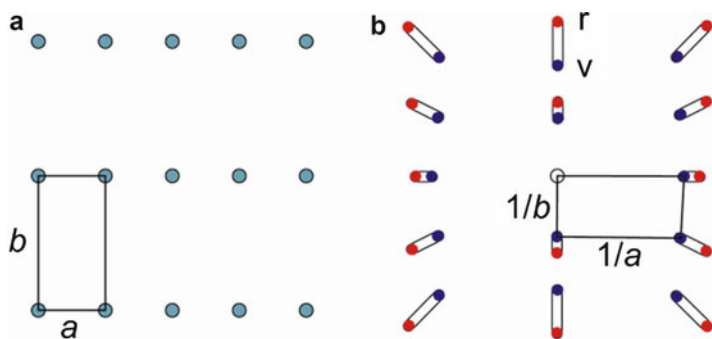
**Iridescence****(Goniochromism),**

Fig. 5 Iridescence from two-dimensional diffraction grating: (a) grating can be envisaged as intersecting one-dimensional gratings; (b) diffraction pattern, with violet diffracted least and red most



and the diffraction pattern may be understood intuitively as arising from intersecting one-dimensional gratings. Each row will give rise to principal maxima (described by the grating equation) running normal to the row selected, separated by weak intermediary subsidiary spots that do not give rise to strong iridescence. A complete pattern will only be seen if the conditions of the grating equation are fulfilled for every row simultaneously. For a rectangular grating illuminated by a beam normal to the grating, the conditions for diffraction are

$$a \sin \theta_m = m\lambda \quad (\text{grating spacing } a)$$

$$b \sin \theta_p = p\lambda \quad (\text{grating spacing } b)$$

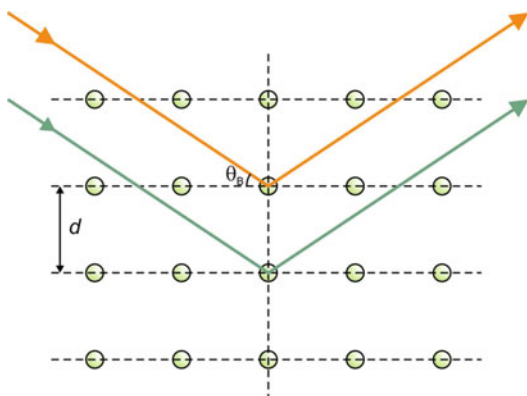
(When this condition is met, all the other rows will be perpendicular to the incident beam.) Each

diffracted order will now be specified by two indices m and p so that

$$d_{mp} \sin \theta_{mp} = mp\lambda$$

where d_{mp} is the spacing of the grating elements along the row giving rise to the diffraction pattern row containing the order mp . As before, the zero-order beam corresponds to undeviated transmission or specular reflection and is not colored. The diffraction orders, each forming a spectrum with violet diffracted least and red the most, will now be arranged on a grid with symmetry that matches that of the grating (Fig. 5b).

Many iridescent gift tags and bags show two-dimensional diffraction effects. The surface of these consists of a thin plastic film, often overlaying a reflecting layer. The polymer



Iridescence (Goniochromism), Fig. 6 Diffraction by three-dimensional array of scattering centers

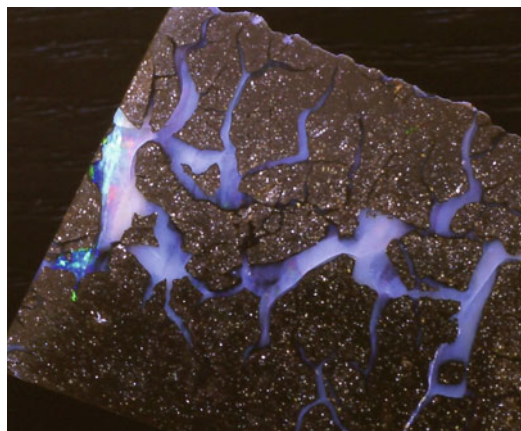
surface is embossed with several sets of grooves. Two sets at right angles give rise to square or rectangular diffraction effects while three sets at 60° give rise to hexagons and star diffraction patterns.

Natural and artificial three-dimensional ordered crystal-like arrays of scattering centers, also called *photonic crystals*, are also known. In this case diffraction is constrained by the spacing of the scattering centers in three dimensions. The condition for strong diffraction to occur is given by Bragg's law:

$$m\lambda = 2d \sin \theta_B$$

Where m is the order of diffraction, d is the distance between the planes of scattering centers giving rise to the diffracted beam, and θ_B is the angle between the incident beam and the scattering planes (Fig. 6). As in previous examples, the position of all diffracted beams except for the zero order is a function of wavelength, and each diffracted order will be pulled out into a spectrum with violet diffracted least and red the most.

The most familiar natural object that displays such iridescence is the gemstone precious opal (Fig. 7). This consists of arrays of ordered silica spheres embedded in a matrix of amorphous silica. As the opal is tilted, the color observed moves to shorter wavelengths giving a blueshift.



Iridescence (Goniochromism), Fig. 7 Iridescent vein of opal in ironstone (Photograph R J D Tilley)

Artificial opals, also called colloidal opals, are made by sedimenting evenly sized polymer spheres. Under controlled conditions, these will pack into close-packed arrays as in natural opal and show typical opaline iridescence. If the spaces between the spheres are infiltrated with an inorganic material, typically an oxide such as TiO_2 or ZrO_2 , and the polymer subsequently dissolved, an array of close-packed shells forms. These show similar iridescence to natural opals.

If a structure forms in which the layers are composed of birefringent molecules, polarization of the scattered beam may occur. When the orientation of the molecules is such that there is an approximately constant rotation angle from one layer to the other, the diffracted light is circularly polarized. The color observed is given by Bragg's law, in which the pitch of the spiral is equivalent to d and the observed color changes with viewing angle as in the case of opals. This is found in a number of beetles and is exploited in liquid crystal thermometers.

Holograms

Iridescent colors that change hue and content dramatically with both viewing angle and viewing distance are produced by the holograms commonly seen as security tags on credit cards, banknotes, and many novelty gift items such as



Iridescence (Goniochromism), Fig. 8 Suppressed iridescence: (a) rose chafer, *Cetonia aurata* (Photograph Ralph Hobbs, 28 April 2010, Lewes, E. Sussex, with permission); (b) silver spangles on the underside of the

Queen of Spain fritillary, *Issoria lathonia* (Photograph R J D Tilley); (c) Adonis blue, *Lysandra bellargus* (Photograph R J D Tilley)

bookmarks, when viewed in white light. These are called rainbow holograms, Benton holograms, or white light transfer holograms.

Holograms are essentially interference patterns formed by the superposition of coherent light from a reference beam and the same light after it has been scattered by an object. The “reconstructed” image is observed when the eye receives light from the reference beam alone that has interacted with the hologram. Holograms are able to record both phase and amplitude information, so that the image contains considerable information content not normally inherent in an orthodox interference pattern. When conventional holograms are viewed in white light, each different wavelength produces its own distorted image of the object. These overlap and a satisfactory

image is not observed. Rainbow holograms are produced by a sophisticated technique that allows undistorted colored images of the object to be viewed sequentially as the hologram is tilted. Each image, of a single color, appears as if viewed through a narrow slit. As the angle of observation changes, the virtual slit giving rise to the image changes, giving rise to different colors and different images.

Rainbow holograms are embossed on a thin plastic sheet by conventional embossing techniques. A master hologram is initially created and this is copied to produce numerous metal “stamper shims” that are used in the embossing process. The interference pattern is recorded as complex series of ridges and whorls in the transparent film. To create an image, white light passes

through the film containing the interference pattern and is reflected back from the backing to pass through the hologram again before reaching the eye. Passage through the holographic structure imposes a phase change on the components of the light which is interpreted as described.

Suppressed Iridescence

Many creatures, especially insects and birds, show shining metallic reflections with all of the characteristics of iridescence except that there is no significant change of hue with viewing angle, as if the iridescent color-changing ability has been suppressed. These non-iridescent or suppressed iridescent colors arise in similar ways to iridescence, from coherent scattering of incident white light. The change in the optical effect is due to a change in the nanostructural organization of the scattering materials. For example, ordered high-reflectivity stacks give rise to noticeable iridescence. As the stacks become less well ordered, due to refractive index variation or an uneven distribution of layers, the single angle-dependent color reflected is replaced by a shining single color that is fairly constant over a wide range of angles (Fig. 8a). If the layers in the stack are very disordered, omnidirectional metallic silver reflection is produced (Fig. 8b).

Similar complexity modifies the diffraction characteristics of diffraction grating iridescence. As an ordered grating becomes replaced by structures of a quasicrystalline nature, made of interconnected sheets, pillars, tubes, or globes of chitin or β -keratin, the characteristic angular-dependent iridescence may be replaced by a single shining color that does not change significantly with viewing angle (Fig. 8c). In many of these natural nanostructures, Fourier transform analysis shows that short-range order is important in color production.

In some cases iridescence appears to be suppressed because the scattering medium is composed of slightly misaligned microdomains. At any particular angle of incidence, only some of

the microdomains reflect coherent light which is brightly colored. As the viewpoint is changed, some of these are extinguished while others shine out. The uniform non-iridescent shining color is the macroscopic effect of the microdomains winking on and off as the angle changes. This augments the suppressed iridescence in many blue butterflies of the *Lycaenidae* family (Fig. 8c).

References

1. Isenberg, C.: *The Science of Soap Films and Bubbles*. Dover, New York (1992)
2. Tilley, R.J.D.: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
3. Nassau, K.: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
4. Lynch, D.K., Livingston, W.: *Colour and Light in Nature*. Cambridge University Press, Cambridge (1995)
5. Meadows, M.G., et al.: Iridescence: views from many angles. *J. R. Soc. Interfaces* **6**(Suppl 2), S107–S265 (2009)
6. Vukusic, P.: Structural colour. In: Schwartz, J.A., Contescu, C.I., Putyera, K. (eds.) *Dekker Encyclopedia of Nanoscience and Nanotechnology*, vol. 5, pp. 3713–3722. Marcel Dekker, New York (2004)
7. Lee, D.: Chapter 10, Iridescent plants. In: Lee, D. (ed.) *Nature's Palette*. University of Chicago Press, Chicago (2007)
8. Parker, A.R.: 515 million years of structural colour. *J. Opt. A Pure Appl. Opt.* **2**, R15–R28 (2000)
9. Vukusic, J.P., Sambles, R.: Photonic structures in butterflies. *Nature* **424**, 852–855 (2003)
10. Kinoshita, S., Yoshioka, S., Miyazaki, J.: Physics of structural colours. *Rep. Prog. Phys.* **71**, 076401 (2008). 30 pp

Irisation

- [Iridescence \(Goniochromism\)](#)

Isochromatic Color Confusions

- [Pseudoisochromatic Plates](#)

Itten, Johannes

David Briggs¹ and Stephen Westland²

¹Public Programs, National Art School, Sydney, NSW, Australia

²Colour Science and Technology, University of Leeds, Leeds, UK



Johannes Itten was a Swiss expressionist painter, designer, and teacher and one of the main pedagogical forces behind the Bauhaus in its earliest phase. His 1961 book *The Art of Color* presented color theory in a simplified form that largely excluded scientific developments from the mid-nineteenth century onward. His approach has permeated much subsequent teaching of color in the arts.

Itten was born in Südern-Linden (Switzerland) on 11 November 1888. He trained and practiced as a teacher in Bern before studying under the abstract painters Eugène Gilliard in Geneva (1912) and Adolf Hoelzel in Stuttgart (1913–1916). He then ran his own art school in Vienna until the director of the Bauhaus in Weimar, Walter Gropius, appointed him as one

of its first teachers in 1919. Itten played a key role in the development of the “preliminary course” that would teach students the basics of material characteristics, composition, and color. However, conflict involving Itten’s ambitions, his promotion of the eastern-inspired Mazdaznan sect, and his opposition to involvement in commercially oriented design, by which Gropius hoped to validate the state-funded Bauhaus in a hostile political and economic climate, prompted Itten to leave in 1923. He subsequently taught in the Mazdaznan community in Zurich (1923–1926) before establishing an art and architecture school in Berlin (1926–1934) and directing the Advanced Vocational School for Textile Art in Krefeld (1932–1938). Itten then settled in Zurich, serving as director of the Museum and School of Applied Arts (1938–1953), the Silk Industry Vocational School (1943–1960), and the Rietberg Museum (1949–1956).

In his retirement Itten published his main book on color theory, *The Art of Color*, in 1961 [1–3] and an account of his Bauhaus preliminary course, *Design and Form*, in 1963 [4]. Some of his ideas on color had appeared previously in the rare hand-printed *Tagebuch* of 1930 [5] and *Die Farbe*, an exhibition catalog from 1944 [6].

Color Star, Color Circle, and Color Sphere

At the Bauhaus, Itten taught color theory using a “color star” of radiating tint and shade scales that he printed as a lithograph in 1921. Its 12-hue scale was derived via Hoelzel from one Bezold had proposed as being perceptually equal, but was modified to align what Itten regarded as the warm-cool boundary (between yellow and yellow green) vertically. Unlike this 1921 scale, which placed yellow, “purple” (magenta), and cyan blue in a symmetrical triad, Itten’s post-Bauhaus color diagrams were all structured around a symmetrical triad of perceptually pure red, yellow, and blue primaries, producing an unequal hue scale with larger perceptual steps in the yellow-green-blue

sector. Three secondary hues (orange, green, and violet) and six intermediates (red-orange, yellow-orange, etc.) complete the circle. This change reflects Itten's adherence to the view, widely held in science until the mid-nineteenth century, that all object colors are mixtures of red, yellow, and blue. For a three-dimensional model, Itten ignored the quantitative systems produced by Munsell and Ostwald and used a simple sphere externally resembling the one published by Runge in 1810. This sphere places the strongest colors of all hues on the equator, with the result that the vertical dimension does not represent lightness consistently.

Color Contrasts

Hoelzel incorporated a broad range of sources into a system of seven or eight "contrasts" of color that were central to his teaching. Itten simplified Hoelzel's classification and language into a list of seven contrasts that is one of the most widely cited elements of his system: contrast of hue, of light and dark, of cold and warm, of complements, of saturation, of extension, and simultaneous contrast.

Subjective and "Objective" Color Harmony

Itten encouraged exploration of the color preferences of the individual, but warned that this "subjective harmony" must often be subordinated to "objective" (though scientifically unexamined) laws. "Objective" harmony required balance of the three traditional primaries, which could be obtained (following Hoelzel) from complementary pairs, equilateral and isosceles triads, and rectangular, square, and trapezoidal tetrads in his 12-hue circle, but also by tilting these shapes in any direction within his color sphere. Balance also required that the three primaries be present in a set ratio that Hoelzel had derived ultimately from Schopenhauer, but which Itten misattributed to

Goethe. Unbalanced, "discordant" combinations could however be used for expressive effect.

Color Expression

Itten also regarded color expression as involving objective rules and presented a system of "dictionary" meanings of colors and color combinations, in which complementary colors were expected to have opposite meanings, secondary colors were expected to combine the meanings of the primaries they "contain," and meanings could be modified by contrast effects with surrounding colors.

References

1. Itten, J.: *The Art of Color: The Subjective Experience and Objective Rationale of Color*. Reinhold (1961). (English translation of *Kunst der Farbe. Subjektives Erleben und objektives Erkennen als Wege der Kunst*)
2. Itten, J.; Birren, F. (ed.): *The Elements of Color*. Van Nostrand Reinhold Company. New York (1971)
3. Itten, J.: *The Art of Color: The Subjective Experience and Objective Rationale of Color*. Wiley, New York (1974)
4. Itten, J.: *Mein Vorkurs am Bauhaus. Gestaltungs- und Formlehre*. Otto Maier Verlag, Ravensburg (1963). (Translated as *Design and Form: The Basic Course at the Bauhaus*, Thames and Hudson. London (1964))
5. Itten, J.: *Tagebuch: Beiträge zu einem Kontrapunkt der bolenden Kunst*. Berlin: Verlag der Itten-Schule (1930); reissued Zurich (1962)
6. Itten, J.: *Die Farbe* (exhibition catalogue), Kunstgewerbemuseum. Gewerbeschule der Stadt, Zurich (1944)

Further Reading

- Briggs, D.J.C.: <http://www.huevaluechroma.com/>. Accessed 17 Feb 2014
- Forgacs, E.: *The Bauhaus idea and Bauhaus politics*. Central European University Press, Budapest (1995)
- Parris, N.G.: *Adolf Hoelzel's structural and color theory and its relationship to the development of the basic course at the Bauhaus*. Ph. D. thesis, University of Pennsylvania (1979)
- Poling, C.V.: *Color theories of the Bauhaus artists*. Ph.D thesis, Columbia University (1973)
- Wagner, C. *Johannes Itten. Biografie in Bildern*. In: Uthemann, E.W. (Hrsg.) *Johannes Itten: alles in einem – alles im Sein*. Ostfildern-Ruit 2003, pp. 81–96. (2003)

ITULAB

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

Ives, Frederic Eugene

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Ives was born on February 17, 1856, in Litchfield, CT, to Hubert L. Ives and his wife Ellen. His father passed away when Frederic was still a child. He left school when he was 12 and began an apprenticeship as printer at the *Litchfield Enquirer*. Photography and engraving became his hobby. At age 18, based on his reputation and without having a formal education, he was invited to run the photography laboratory at Cornell University in Ithaca, NY, where he stayed for 4 years, a key period in regard to his inventions.

While at Cornell he invented a practically valuable method of the halftone printing process for black-and-white photographs [1]. In 1879 he got married and moved to Philadelphia where he established an association with a major manufacturer of woodcut engravings who was interested in photographic reproduction methods and Ives' halftone process. Until 1880, illustrations in books, magazines, and newspapers were largely based on woodcuts as well as lithography; by the turn of the century, the industry had mostly switched over to Ives' halftone process.

Ives was issued a total of 70 patents in his lifetime, the earliest related to his halftone printing process and many related to trichromatic halftone printing. He revolutionized color printing in books, magazines, and newspapers in a manner similar to that for black-and-white printing. Up to that time the process in commercial use of color printing was lithography, with the image constructed from up to a dozen handmade limestone or metal engravings used to print with differently colored inks. He regularly lectured on the subject of his inventions at Philadelphia's Franklin Institute. Circa 1890 he moved to New Jersey, outside New York. Frederic Eugene Ives died on May 27, 1937, in Philadelphia, Pennsylvania. His portrait was attached to a postage stamp celebrating his invention of the halftone process (Fig. 1) [2].

F. E. Ives' son Herbert obtained a PhD degree from Johns Hopkins University in 1908. He was also active in the field of color science, publishing in 1915 an important article on the transformation of color-matching functions between different colorimetric systems [3].

Major Accomplishments/Contributions

Ives was granted patents covering the halftone process in 1881 [4]. In 1890 he obtained a patent for his "trichromatic" color photography process [5] and in 1892 for a related camera [6]. There were two steps to Ives' halftone color printing process: (1) Color separations were obtained by

making successive exposures through colored filters onto film, and halftone plates were made from these filtered exposures. (2) The three separations were printed onto paper, on top of each other in exact registration, with yellow, magenta, and

blue-green inks [7]. Unlike the system proposed for color photography by J. C. Maxwell in 1855 [8] using red, green, and blue filters, Ives learned from experiments that the most accurate results were obtained when using yellow, magenta, and cyan (blue green) as primaries. This process and further developments resulted in a revolution of the technical methods of printing colored images.

In 1894 he obtained a patent for the Kromskop (Fig. 2), a stereoscopic version of his trichromatic camera [9]. It had limited success, and some years later, Ives transferred the related patents to the Eastman Kodak Company where they were used to develop the Kodachrome process. Two more significant inventions were a portable “colorimeter” and the “photometer,” both patented in 1908. The former was used to measure and define color stimuli in terms of trichromatic designations, the latter to obtain measures for hue and intensity of lights. In 1909 and 1911 Ives was granted patents for the Tripak paper color film, used in the Tripak color camera [10].



Ives, Frederic Eugene, Fig. 1 US postage stamp celebrating Ives' invention of the halftone printing process, 1996 (Copyright The U.S. Government)

Ives, Frederic Eugene, Fig. 2 Stereoscopic color image as viewed in the Kromskop, 1897. On the left are the B&W stereoscopic images for the three primary colors as used in the Kromskop. Top right is a digital recreation of the two stereoscopic versions of the image. On bottom right is an enlarged version of the image based on a photographic print made in the 1950s



References

1. Stulik, D.C., Kaplan, A.: Halftone. Getty Institute, Los Angeles (2013)
2. <https://multimediaman.wordpress.com/> search entry: f e ives. Accessed 5 May 2015
3. Ives, H.E.: The transformation of color mixture equations from one system to another. *J. Franklin Inst.* **115**, 673–701 (1915)
4. US Patents 237,664 and 245,501, 1881
5. US Patent 432,530, 1890
6. US Patent 475,084, 1892
7. Ives, F.E.: Photography in the colors of nature. *J. Franklin Inst.* **81**, 1–21 (1891)
8. Maxwell, J.C.: Experiments on colour, as perceived by the eye, with remarks on colour-blindness. *Trans. R. Soc. Edinb.* **21**(Part II), 275–298 (1855)
9. US Patent 531,040, 1894
10. US Patent 927,244, 1909

Judd, Deane Brewster

Michael H. Brill¹ and Rolf G. Kuehni²

¹Datacolor, Lawrenceville, NJ, USA

²Charlotte, NC, USA



Deane Brewster Judd (1900–1972) was an American physicist who contributed to the fields of colorimetry, color discrimination, color order, and color vision. Born in South Hadley Falls, Massachusetts, he attended Ohio State University

and Cornell University and received there a Ph.D. in physics in 1926. He was a Munsell research associate in colorimetry at the National Bureau of Standards (NBS) in Washington in 1926.

In 1927 Judd joined the NBS, where he remained until his retirement in 1969 and subsequently continued as a guest worker. He was the USA's representative in colorimetry at eight meetings of the International Commission on Illumination (CIE) from 1931 to 1967 and thereby a key force in the development of the CIE standard system of colorimetry, e.g., 1931 and 1964 standard observers, standard illuminants B and C, daylight illuminants like D65, and definition of colorimetric purity [1]. Largely responsible for coining the term “psychophysics,” he wrestled throughout his career with the relationship between color stimuli and color perception.

Colorimetric System

Judd introduced the concept of keeping luminosity and chromaticness separate in the CIE system. He was active in the colorimetric definition of color temperature and introduced the CIE colorimetric system to US industries. Together with D. L. MacAdam and G. Wyszecki, in 1964, he used principal component analysis to show that natural daylights are largely composed of three components from which daylights at any

correlated color temperature can be defined (CIE method of calculating D illuminants) [2].

Color Difference

In a series of papers in the 1930s, Judd represented then-available color-scaling data first into a chromaticity diagram based on color-matching functions introduced by the Optical Society of America in 1922, then represented the resulting diagram in a Maxwell-type primary triangle, and finally embedding unit difference ellipses into the CIE chromaticity diagram. This work became the basis for the CIE (u,v) color diagram in 1960, slightly modified in the 1976 CIELUV color space. In 1939 he was instrumental in developing the NBS color difference formula. When in 1947, at the suggestion of the US National Research Council, the Optical Society of America (OSA) undertook to develop a perceptually uniform color space. Judd became its chairman and remained in that position until 1968, when D. L. MacAdam assumed the chairmanship, with results published in 1974. One of the key findings was “that strictly uniform color scales of all kinds are not homologous with Euclidean space” to which Judd proposed a solution implemented in the OSA Uniform Color Space [1].

Systematic Color Names

The perceived need for systematic naming of colors resulted in 1939 in ISCC-NBS Method of Designating Colors, based on the Munsell color system, with a revised edition published in 1955 [3, 4].

Color Constancy

Judd presented on color constancy first in 1940. He then presented with Helson and Warren, in 1952, results of meticulous work on the subject [5, 6].

Miscellaneous

At NBS Judd investigated impaired color vision, whiteness measurement of paper, opacity, color stimulus measurement, and a flattery index for artificial light sources. In 1951 he proposed a modification of the CIE 1924 luminous efficiency function $V(\lambda)$ below 460 nm that became known as Judd-modified $V(\lambda)$, not implemented in the CIE system but used in some vision research works [7]. In addition to his positions at the CIE, Judd was president of the Optical Society of America from 1953 to 1955 and of the Inter-Society Color Council from 1940 to 1944. He was president of the Board of Trustees of the Munsell Color Foundation from 1942 to 1972.

In his homage, the International Colour Association (AIC) instituted its Judd Award, a prize that since 1975 is bestowed every 2 years to persons who have made relevant contributions in color research.

Judd was the author of one book [8] and more than 200 articles, 57 of which were reprinted as a volume by the NBS [1].

References

1. Judd, D.B.: In: MacAdam, D.L. (ed.) Contributions to Color Science. NBS Special Publication 545. U.S. Department of Commerce, Washington, DC (1979)
2. Judd, D.B., MacAdam, D.L., Wyszecki, G.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1040 (1964)
3. Judd, D.B., Kelly, K.L.: Method of designating colors. *J. Res. NBS* **23**, 355–366 (1939)
4. Kelly, K. L., Judd, D. B.: Color, universal Language and Dictionary of Names. U.S. Department of Commerce, Washington, DC (1955/1976)
5. Judd, D.B.: Hue, saturation, and lightness of surface colors under chromatic illumination. *J. Opt. Soc. Am.* **30**, 2–32 (1940)
6. Helson, H., Judd, D.B., Warren, M.H.: Object-color changes from daylight to incandescent filament illumination. *Illum. Eng.* **47**, 221–233 (1952)
7. Report of U.S. Secretariat Committee on Colorimetry and Artificial Daylight, CIE Proceedings, vol. 1, part 7, Stockholm, 1951
8. Judd, D. B.: Color in Business, Science and Industry. Wiley, New York (1959), 2nd and 3rd edn. with Gunter Wyszecki, 1963 and 1975

K

Katz, David

Rolf G. Kuehni
Charlotte, NC, USA

Biography

Katz was born on October 1, 1884 as the seventh of eight children of his parents in Kassel, Germany, where he attended the local Realgymnasium. As a child and student, he became a talented amateur painter, interested in color and other sensory abilities. In 1902 he began studying at the nearby University of Göttingen. There he became interested in psychology taught by G. E. Müller. In 1906 he obtained a PhD degree in psychology, physics, and philosophy, and in 1907 he assumed the position of assistant to Müller. In 1911 he was named a professor, based on his work on color phenomenology [1]. From 1914 to 1918 he was a volunteer soldier in the German Army. In 1919 he received an invitation from the University of Rostock to assume a then new professorship in psychology. In the same year he married the psychologist Rosa Heine with whom he had two sons. Over the years they both were active in a wide field of psychology, including that of children and animals and had a number of well-known students. In 1929 Katz was for a time a guest professor in Maine in the USA. Shortly after Hitler's move to power in 1933, Katz and his wife were officially retired from their positions. In the same year Katz, with the help of a colleague, moved to England,

followed a year later by his wife and sons. In 1937 Katz received an invitation to fill the first professorship in psychology in Sweden at the University of Stockholm. Here he was broadly active in many fields of psychology. In 1952, already retired, he received an honorary professorship from the University of Hamburg. On February 2, 1953, he died in Stockholm. Katz produced over 200 research papers and books, most in the general field of psychology [2].

Major Accomplishments/Contributions

Beginning with his dissertation, Katz was specifically interested in the phenomenology of vision, including shape, structure, space, color, and movement, to the exclusion of related physics and psychophysics. He defined three appearance modes of colors: film colors, surface colors, and volume colors. Film colors are those experienced when looking into the black tube of a spectroscope and seeing spectral colors, having a slightly spongy appearance without having a specific location in space. Surface colors are those commonly referred to as object colors. Volume colors he defined as those of colored glasses or gels. An important concept in his work was the subjective gray we experience either with closed eyes in the absence of strong light or with open eyes in a completely dark room. Another one was the reduction screen, a black sheet of paper with a small hole through which a small section of a colored object could be viewed and the resulting film color



Image source: University of Rostock, Germany

appearance compared to the surface color without the screen. Katz and his students performed many experiments related to color appearance involving illumination changes, surround and shadow effects, monocular versus binocular vision, exposition time, color constancy and lack thereof, light and dark adaptation, color and depth contrast effects, and the effects of chromatic illumination, to name a few. He introduced novel concepts, translated as “pronouncedness” and “insistence,” to describe nuanced phenomenological color effects. A key subject was the problem of definition of the “true” color of an object. To measure phenomenological magnitudes, he used innovative disk mixture methodology.

In his book [1] Katz discussed his agreements and disagreements with other researchers and their findings, such as Hering and of his time K. Bühler, A. Gelb, E. Jaensch, J. von Kries, and E. G. Müller. An important psychological concept of that period was Gestalt psychology based on the idea that, as its primary formulator M. Wertheimer expressed it, “the brain is holistic” by considering in assessing a given situation the total amount of information received. Together with self-organizing tendencies, this results in “the whole is greater than its parts.” Katz published a textbook on the subject in 1944, issued in several editions and multiple languages [3].

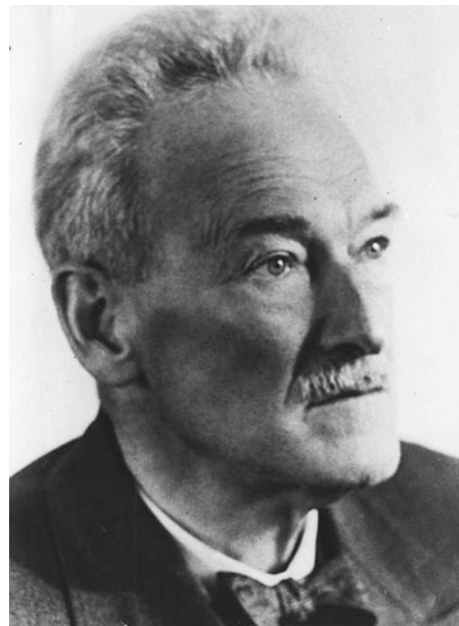
Katz’s book of 1911 was issued in a revised and updated version as *Der Aufbau der Farbwelt* (The constellation of the world of colors) in 1930 [4]. After he moved to London a shortened, translated version was published in 1935 in England as *The world of colour* [5].

References

1. Katz, D.: Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung (Appearances of colors and how they are influenced by individual experience). Barth, Leipzig (1911)
2. www.ipprdk.uni-rostock.de, search: david katz
3. Katz, D.: Gestaltpsychologie. Schwabe, Basel (1944)
4. Katz, D.: Der Aufbau der Farbwelt (The constellation of the world of colors). Barth, Leipzig (1930)
5. Katz, D.: The world of colour (R. B. MacLeod, trans.). Kegan Paul, Trench, Trubner, London (1935)

Kohlrausch, Arnt

Rolf G. Kuehni
Charlotte, NC, USA



Portrait © Jürgen Killmann and Universität von Tübingen, 1969

Biography

Arnt Kohlrausch was born on October 30, 1884 in Hannover, Germany, a member of an extended family of scientists from the early nineteenth century to the present. His father was professor of electrical technology at the Technical College of Hannover. Arnt Kohlrausch studied medicine at the universities of Marburg, München, and Rostock, graduating in 1911. In 1918 he became lecturer in physiology at the University of Berlin. In 1926 he moved to the University of Greifswald, today in the German province of Mecklenburg. In 1928 he became chair of physiology at the University of Tübingen, where he remained until his retirement in 1951. Kohlrausch was active in several fields of physiology. Perhaps his best-known work is *Körperliche und psychische Lebenserscheinungen* (Bodily and psychological phenomena of life) [2]. He also did research in visual perception and was one of the founders of the journal *Die Farbe* (Color). A peculiarity of low-level brightness perception, the *Kohlrausch-Knick* (bend) is named after him. It refers to a bend in the curve of absolute luminance threshold versus time during dark adaptation, the bend being due to the crossover of absolute thresholds of rods and cones as a function of time. The perceptual effect on an observer is that a very dim scene suddenly appears brighter as the rod sensitivity rises enough to dominate the visual signal. In English it is known as the “rod-cone break.” Kohlrausch died on July 13, 1969 [1].

Major Accomplishments/Contributions

In 1923 Kohlrausch published an article *Über den Helligkeitsvergleich verschiedener Farben, Theoretisches und Praktisches zur heterochromen Photometrie* (Comparing brightness of different colors, theoretical and practical aspects of heterochromatic photometry) [3] in which he investigated in detail the effect of hue on perceived brightness, mentioned earlier by Helmholtz [4]. This was followed in 1935 by the article *Zur Photometrie farbiger Lichter* (Photometry of colored lights) [5]. As reported by D. B. Judd [6], in 1939, J. Urbanek and E. Ferencz introduced the term Helmholtz-Kohlrausch effect for the phenomenon, a term still in use today [7]. Figure 1 shows some examples. The six circles have luminance values identical to that of the achromatic surround. To most observers the circles appear, to various degrees, lighter than the gray of the surround.

Helmholtz used the term *glühend* (glowing) to describe the appearance of the chromatic colors compared to the achromatic surround. The strength of the effect varies by hue. Quantitative investigations of the Helmholtz-Kohlrausch effect (HKE) based on the CIE colorimetric system were performed by Wyszecki and Sanders in 1964 [8], the basis of the inclusion of the HKE in the lightness formula of the Optical Society of America Uniform Color Scales [9].

Kohlrausch, Arnt,

Fig. 1 Examples of the Helmholtz-Kohlrausch effect. All circles have the same colorimetric lightness values as the gray surround



References

1. Kohlrausch, A.: Körperliche und psychische Lebenserscheinungen. Kohlhammer, Stuttgart (1934)
2. https://de.wikipedia.org/wiki/Arnt_Kohlrausch. Accessed 7 May 2015
3. Kohlrausch, A.: Über den Heligkeitsvergleich verschiedener Farben. Theoretisches und Praktisches zur heterochromen Photometrie. Pflugers Arch. **200**, 210–215 (1923)
4. von Helmholtz, H.: Handbuch der physiologischen Optik, 2nd edn. Voss, Hamburg (1896)
5. Kohlrausch, A.: Zur Photometrie farbiger Lichter. Das Licht **5**, 259–275 (1935)
6. Judd, D.B.: A new look at the measurement of light and color. Illum. Eng. **53**, 61–71 (1958)
7. Urbanek, J., Ferencz, E.: Comité d'Etudes sur la Photométrie Visuelle – Rapport du Secretariat, 10e Session, p. 44. Commission Internationale de l'Éclairage, Scheveningen (1939)
8. Wyszecki, G., Sanders, C.L.: Correlate for lightness in terms of CIE tristimulus values. J. Opt. Soc. Am. **47**, 840–842 (1957)
9. MacAdam, D.L.: Uniform color scales. J. Opt. Soc. Am. **64**, 1691–1702 (1974)

Kries, Johannes Adolph von

Mark D. Fairchild

College of Science, Rochester Institute of Technology, Rochester, NY, USA



J. v. Kries

Johannes von Kries was a German physiological psychologist, or what might now be called a neuroscientist, and student of Helmholtz. In color science, he is known as the father of chromatic adaptation models for his work on the coefficient theory of adaptation. He also made significant contributions to a variety of areas including probability theory and the structure of the retina and human visual system [1].

von Kries published profusely (in German, with few works translated to English) on topics such as blood flow in arteries, duplicity theory, chromatic adaptation, zone theory, probability theory, physiology, psychology, and history. For example, observations of the Purkinje shift led von Kries to postulate the existence of two separate visual systems, rods and cones, also referred to as duplicity theory. It is well known that Helmholtz described the trichromatic theory of color vision and participated in a famous academic dispute with Hering, the main proponent of opponent-colors theory. It is less widely recognized that von Kries was among the first proponents of the so-called zone theory of color vision that allowed for trichromatic receptors (the cones) and opponent processing of the visual information. It is little wonder that von Kries is sometimes referred to as the “greatest German disciple” of Helmholtz. Considering that Helmholtz’s students included the likes of Max Planck, Wilhelm Wien, Arthur König, A.A. Michelson, and Wilhelm Wundt, that is quite a statement indeed.

In the second half of the nineteenth century, von Kries was applying probability theory to the evaluation of the effectiveness of new drugs. He realized that the computation of probability distributions depends on the classification of symptoms and pathologies into diseases, often very subjective data. Given that the important uncertainty was determined by the experimenters’ definition of events, he developed the logical foundations of a probability theory where the subjectivity of mental representations impacts the assignment of numerical values to probabilities. Interestingly, with some distortion and misunderstanding, these ideas passed on to Keynes and formed the core of his economic theory. In his book on the

topic, von Kries developed a highly original interpretation of probability, illustrating it to be both logical and objective.

Returning to color science, von Kries' most recognized and long-lived contribution came from his works on the theory of chromatic adaptation (1902, 1905) [2, 3] in which he proposed the coefficient rule that lives on to this day as the von Kries coefficient law or simply the von Kries model of chromatic adaptation. Interestingly von Kries did not write out the mathematical formula with which he is credited; he simply stated the theory in words. In MacAdam's translation of von Kries' (1902) words:

This can be conceived in the sense that the individual components present in the organ of vision are completely independent of one another and each is fatigued or adapted exclusively according to its own function.

Perhaps it shouldn't be too surprising to realize that von Kries (1902) himself foresaw that this model was too simple to explain all chromatic adaptation phenomena. The next line after his description of what is now referred to as the von Kries model reads:

But if the real physiological equipment is considered, on which the processes are based, it is permissible to doubt whether things are so simple.

Indeed things are not so simple, yet von Kries' theory forms the basis of all effective modern models of chromatic adaptation, including that embedded in CIECAM02.

References

1. von Kries, J.: Beitrag zur Physiologie der Gesichtsempfindungen (Physiology of visual sensations), (1878) (Translation: MacAdam, D. L.: Sources of Color Science. MIT Press, Cambridge, (1970))
2. von Kries, J.: Chromatic adaptation, Festschrift der Albrecht-Ludwig-Universität, (Fribourg) (1902) (Translation: MacAdam, D.L.: Sources of Color Science. MIT Press, Cambridge, (1970))
3. von Kries, J.: Die Gesichtsempfindungen, In: Handbuch der Physiologie des Menschen, vol. 3, pp. 109–282. Vieweg, Braunschweig, (1905)

Kruithof Curve

Wout van Bommel
Nuenen, The Netherlands

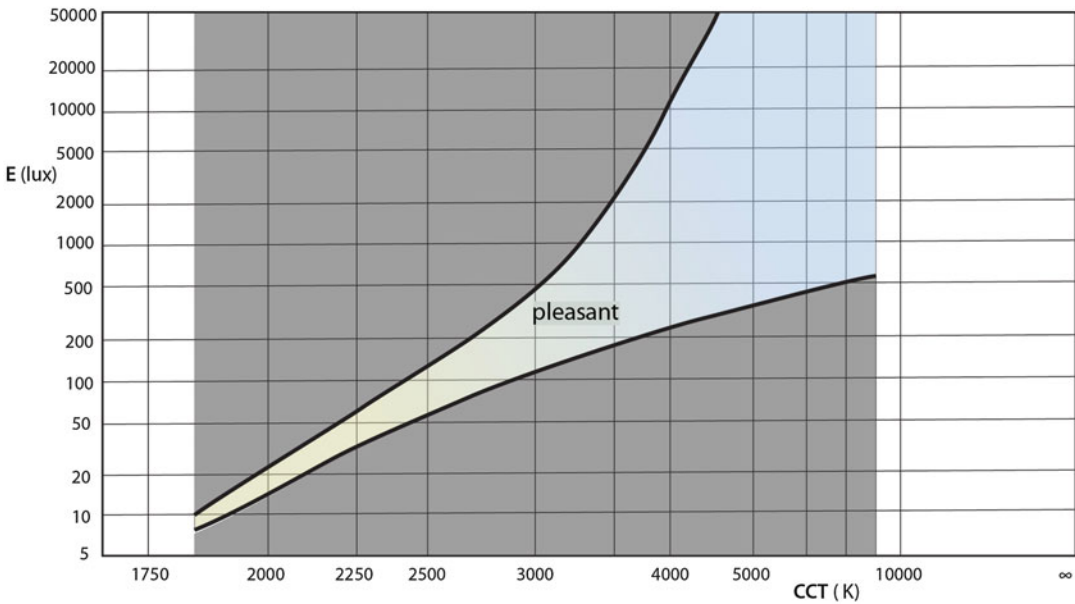
Definition

“Kruithof's graph,” as published in 1941 by Kruithof, that shows, for general lighting in interiors, what combinations of illuminance level and correlated color temperature result in a “pleasing” or “not pleasing” assessment by persons in that interior.

Preferred Correlated Color Temperature (CCT) for General Lighting in Interiors

Around the time of the introduction of tubular fluorescent lamps in the early 1940s, Kruithof was the first who carried out a study on preferred correlated color temperature (CCT) for interior lighting. He investigated the relationship between lighting level in an interior and the CCT of the light sources used, on the assessment of people in terms of pleasantness [1]. His study received much recognition, although many subsequent researchers got inconsistent results. It must be realized that not only the lighting level but aspects as climatic conditions and cultural background influence the preferred illuminance as well.

Lighting not only has a visual effect but also a nonvisual biological effect that influences health and well-being. The CCT of lighting plays an important role in the effectiveness of lighting in this respect. This relatively new knowledge has led to dynamic interior lighting installations where both the lighting level and the CCT gradually change in the course of the day [2]. This, together with the availability of LED light sources with a wide range of CCTs, has given rise to renewed interest in the research of Kruithof and other researchers, on preferred CCT in interior lighting.



Kruithof Curve, Fig. 1 Kruithof's graph with combinations of illuminance levels, E , and correlated color temperatures, CCT, that are judged as "pleasant." Logarithmic

E scale and inverted reciprocal CCT scale, as in Kruithof's original black and white graph [1]

CCT and Lighting Level

The purpose of the 1941 research of Kruithof, then one of the early developers of tubular fluorescent lamps, in fact was the determination of the fluorescent powder composition for the production of commercial fluorescent lamps. Note that at that moment of time, manufacturers planned to bring fluorescent lamps on the market with only one type of fluorescent powder. Kruithof's pleasantness assessment tests were carried out with observers in a room where the lighting level and the CCT of the light sources used could be changed independently. He presented the results in a graph as in Fig. 1. Here the area in between the two curves represents illuminance level and CCT combinations that result in lighting that is judged as "pleasant." The area above the upper curve represents the judgment "too cold" and the area below the lower curve "too dim."

The conclusion of Kruithof is straightforward: at low illuminances, low CCT light sources should be used and, at high illuminances, high CCT light sources. As the big market for

fluorescent tubes in that era was in offices and industries where relatively high illuminances were needed, the first fluorescent tubes on the market indeed had relatively high CCTs (higher CCT lamps could also be produced more efficiently).

Kruithof's results have often been used to draw general conclusions on preferred CCTs in interior lighting design. In this context, however, it is important to note some details of his research:

- Kruithof's graph is based on a pilot study in which only two observers participated [3].
- The research has been carried out in a laboratory setting.
- Kruithof used fluorescent lamps with different fluorescent powder compositions for the midrange of CCTs; for the lower CCT range, incandescent lamps on a dimmer; and for the higher CCT range, daylight or daylight luminescence lamps. This means that the color-rendering qualities of his light sources were

widely different and may have had an important influence on his results.

- The observers had a short habituation time for each different situation. Recent research into the acceptance of extreme high CCT values of up to 17,000 K shows that a habituation time of some 1–2 weeks influences the acceptance of these CCTs [4].
- The research has been carried out in the Netherlands (Mid-Europe). It is known that the climatic conditions and the cultural background of the observers play a role in the preference of CCT values (see next sections).

Since the publication of Kruithof, many other researchers at many different institutions have carried out investigations on the same theme. Many of them gave less straightforward conclusions [3, 5]. With the introduction of LED light sources with a wide range of CCTs for the lighting of interior spaces, the subject has received renewed interest. Again the research results are, to some degree, conflicting. Often the research results validate the statement that the use of high CCT lamps at low illuminances results in unpleasant assessments. However, low CCT lamps do sometimes result in pleasant situations when used at higher illuminances, and some studies show a general preference for the use of lower CCT lamps [6].

CCT and Climatic Conditions

The predominantly prevailing climate also has an influence on the preference for color impression in interiors. In a warm climate, the preference is more toward light sources with higher CCTs, while in cooler climates, preference is more for lower CCTs. These preferences are confirmed by sales analysis of the various color types of fluorescent tubes per region. With the availability of dynamic lighting installations with which CCT can be varied, sometimes the CCT during the winter and summer season is varied according to the above rule, especially in shop lighting.

CCT and Cultural Background

The cultural background plays, probably because of long-time habituation, a role in the acceptance

and preference of both CCT and light levels. A typical illustration is Japan, one of the few countries in the world where fluorescent tubes have been used in domestic lighting, right from the beginning of their introduction in the 1940s. These lamps were originally only available in CCT versions of around 4,000–5,000 K. Today, in Japan, a preference for higher CCT lamps is seen than in other countries with similar climatic conditions where the standard in domestic lighting was incandescent lamps with relatively low CCTs.

CCT and Nonvisual Biological Effects

The eye has light-sensitive cells (intrinsic photo-sensitive retinal ganglion cells) that connect with the biological clock (suprachiasmatic nucleus) in the brain, which in turn connects with the pineal gland that controls for a part the rhythm of the hormone metabolism. Through these pathways, light has a nonvisual biological effect that influences health and well-being. Figure 2 shows the action spectra for nonvisual biological effects together with the spectral sensitivity for photopic vision, V_λ . It shows that light with a large blue component (cool-white light of a high CCT) has a larger nonvisual biological effect than light with a large red component (warm-white light of low CCT).

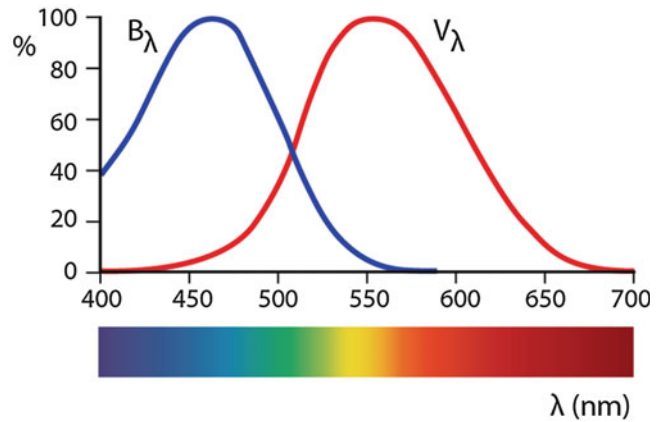
This relatively new knowledge has already led to two new developments in lighting, especially in office, industrial, school, and hospital lighting:

- Dynamic lighting installations that automatically vary the lighting level and CCT in the course of the day to get, at the right moments, biologically activating and relaxing effects. Figure 3 shows an example of how in such a system the lighting level and CCT may vary.
- The introduction and use of extreme high CCT lamps with good color rendering ($R_a > 80$) for use in interior lighting installations (e.g., in dynamic installations as described above).

These developments will probably also have a meaning for the preferred CCT in interiors.

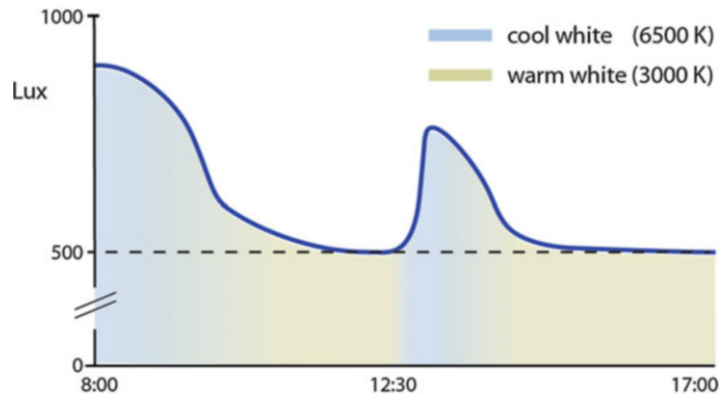
Kruithof Curve,

Fig. 2 Action spectrum for nonvisual biological effects, B_λ , based on melatonin suppression [7], and the relative spectral sensitivity of the eye for photopic vision, V_λ



Kruithof Curve,

Fig. 3 Example of the variation of lighting level and CCT in the course of a working day as obtained from an automated dynamic lighting installation [2]



Cross-References

- [Non-visual Lighting Effects and their Impact on Health and Well-being](#)
- [Color Preference](#)

References

1. Kruithof, A.A.: Tubular luminescence lamps for general illumination. Philips Tech. Rev. **6**, 65–96 (1941)
2. Van Bommel, W.J.M., van den Beld, G.J.: Lighting for work, a review of visual and biological effects. Light Res. Technol. **36**, 255–269 (2004)
3. De Boer, J.B., Fischer, D.: Interior Lighting, 2nd edn. Kluwer Technische Boeken, Deventer (1981)
4. Mills, P.R., Tomkins, S.C., Schlangen, L.J.M.: The effect of high correlated colour temperature office lighting on employee wellbeing and work performance. J. Circadian. Rhythm. **5**, 2 (2007)
5. Boyce, P.R., Cuttle, C.: Effect of correlated color temperature on the perception of interiors and color discrimination performance. Light Res. Technol. **22**, 19–36 (1990)
6. Vienot, F., Durand, M., Mahler, E.: Kruithof's rule revisited using LED illumination. J. Mod. Opt. **56**, 1433–1446 (2009)
7. CIE Publication: Ocular Lighting Effects on Human Physiology and Behaviour, vol. 158. International Commission on Illumination, CIE, Vienna (2004)

Kubelka, Paul

Michael H. Brill¹ and Michal Vik²

¹Datacolor, Lawrenceville, NJ, USA

²Faculty of Textile Engineering, Laboratory Color and Appearance Measurement, Technical University of Liberec, Liberec, Czech Republic



Paul Kubelka (1900–1956) was a Czechoslovakian chemical engineer whose many accomplishments include a theory of light absorption and scattering by a layer of paint. That theory, originally published in 1931 with Franz Munk [1], underlies much software that performs colorant-recipe prediction (colorant formulation). Whereas the 1931 theory assumed that light flows in one dimension (two fluxes, upward and downward within the layer), in 1948 Kubelka derived the same equations (up to a factor of 2) assuming spherical scatter within the paint layer [2]. Later he generalized the theory to inhomogeneous layers [3].

The main contribution of these articles was a closed-form function relating the reflectance R of

a layer to two constants characteristic of small particles within the layer: the absorption coefficient K and the scattering coefficient S (both assessed in a unit thickness of the layer). If the layer is opaque, then the reflectance is a function of K/S ; otherwise R depends on K and S separately, as well as on the reflectance of the material behind the layer. The Kubelka-Munk analysis also includes equations for total transmittance of a translucent layer.

Although Kubelka and Munk independently developed their analysis for paint layers, the underlying theory originated with an astronomical motivation, starting with Arthur Schuster's 1905 paper [4] relating to transmission of light through clouds.

To render Kubelka-Munk analysis useful for colorant formulation, one needed the additivity principle described by Duncan in 1940 [5]. This principle says that the total absorption coefficient K of a layer is the concentration-weighted sum of the K values of the components ($K = c_1 K_1 + c_2 K_2 + \dots$) and similarly for S ($S = c_1 S_1 + c_2 S_2 + \dots$).

Once the Kubelka-Munk theory, the additivity principle, and computer technology had emerged, colorant formulation was on its way. By 1958 Davidson and Hemmendinger introduced the analog Colorant Mixture Computer (COMIC), and this was quickly followed by dedicated digital devices, which in turn yielded to software packages that ran on general-purpose digital computers.

The above discussion places Paul Kubelka's color-science contribution in historical context. His life story offers an interesting context as well (see his 1947 autobiography <http://www.graphics.cornell.edu/~westin/pubs/kubelka-autobio.html>, accessed 8 Jan 2014).

Kubelka was born in 1900 to Austrian parents in Czechoslovakia. He was educated in the German language and attended elementary school in Kladno and secondary school in Brno and Prague. In 1918 he served half a year in the Austrian Army and began studies at the Technical

University in Prague. In 1922 he passed his final examination as a chemical engineer. He then served seven months in the Czechoslovakian Army, where he soon commanded the military analytical laboratory. He also worked with L. Storch, professor of Physical Chemistry at the Technical University.

After leaving military service, Kubelka collaborated with Werner Mecklenburg, the well-known colloid chemist, at Verein für Chemische und Metallurgische Produktion, in Aussig (SPOLCHEMIE, in Ústí n. L.), Czechoslovakia. During this period, he worked on activated charcoal, resulting in patents and in the gas mask charcoal "G 1000." One theoretical investigation of this period earned Kubelka a Doctorate of Engineering in 1926.

During this time Kubelka married Margarethe Schönhöfer and had two children.

In 1928 he led both the Inorganic and Analytic laboratories in Ústí n. L. His laboratory investigated (among other things) high temperature reactions, the preparation of pigments, and activated charcoal.

In 1931 Kubelka resolved to enter an academic career. Because of his publication on absorption and capillary condensation, he was nominated docent of the University of Prague. There he investigated the absorption of vapors by silica gel, which led to an exact method of measurement of surface tension of crystals. His work interested Fritz Haber, and he expected to be nominated as professor at a German University. [It was during this year that Kubelka published the famous paper with Munk, but he does not mention this in his autobiography.]

In 1933 the situation was changed by the Nazi revolution in Germany. Kubelka refused to go to Germany and had few choices in the rest of Central Europe, so he returned to technological contributions – soon founding the company Kubelka Schuloff & Co., which eventually changed to Dr. P. Kubelka & Co. The company established a research laboratory and later a small factory. Products such as the fungicide Cuprenox

were successful. References [6] and [7] are two representative patents from that era. But World War II disturbed this success and prevented the realization of other inventions. Furthermore, Kubelka was forced to change to German citizenship and then was ostracized by the Nazis.

During the postwar events of 1945, Kubelka and his wife were interned by mistake. His wife died in the camp, and he was released finally by the efforts of Czech and Jewish friends in September 1945. His German citizenship was nullified, he was promised to regain Czechoslovak citizenship, and he accepted a position as a research chemist at the Film Company at Cesky Brod near Prague, branch factory of the Aussig Combine (SPOLCHEMIE). There he worked out a new photomechanical emulsion and reorganized the testing system. Upon deciding to go to America, he was told that the Czech authorities intended to prevent him as a specialist from leaving the country. The only place to emigrate legally was to Germany. Through the date of writing of his autobiography, he then lived in Bavaria with his children, working on the optical theory of light-scattering materials and thermodynamics of absorption and capillary condensation. In March 1947 he married Dr. Brigitte Gade.

After emigrating to Brazil in 1950, Paul Kubelka conducted further research to generalize his optical-transfer theory to inhomogeneous layers [3]. He passed away in 1956 in Rio de Janeiro, where he was Head of the Research Laboratory at the Brazilian Ministry of Agriculture.

References

1. Kubelka, P., Munk, F.: Ein Beitrag zur Optik der Farbanstriche. *Z. Techn. Physik*, **12**, 593–601 (1931). See also English translation by S. Westin (An article on optics of paint layers, <http://www.graphics.cornell.edu/~westin/pubs/kubelka.pdf>)
2. Kubelka, P.: New contributions to the optics of intensely light-scattering materials. Part I. *J. Opt. Soc. Am.* **38**, 448–457 (1948) [Also see errata, *ibid* p. 1067.]

3. Kubelka, P.: New contributions to the optics of intensely light-scattering materials. Part II. Non-homogeneous layers. *J. Opt. Soc. Am.* **44**, 330–334 (1954)
4. Schuster, A.: Radiation through a foggy atmosphere. *Astrophys. J.* **21**, 1–22 (1905). [see <http://articles.adsabs.harvard.edu/full/1905ApJ....21....1S/0000001.000.html>]
5. Duncan, D.R.: The colour of pigment mixtures. *Proc. Phys. Soc. London* **52**, 380–390 (1940). [also see his article by the same title in *J. Oil Colour Chem. Assoc.* **32**, 296–321 (1949)]
6. Kubelka, P., Srebeek, J.: Rutile pigments and process of making the same. US Patent 2,062,133 (1936)
7. Kubelka, P., Srebeek, J.: Rutile pigments. US Patent 2,062,134 (1936)

L

Lab

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

Labradorescence

- [Iridescence \(Goniochromism\)](#)

Lambert, Johann Heinrich

Rolf G. Kuehni
Charlotte, NC, USA

Biography

Lambert was born on August 26, 1728 in the city of Mulhouse, then an enclave of Switzerland (now part of France). He was largely self-educated, going to school only until age 12. By age 17 he assumed the job of secretary to a newspaper publisher in nearby Basel, Switzerland. He also began to work as a private tutor. At age 20 he became tutor to three boys in the family of Count Peter von Salis in Chur, Switzerland, a position he held for 10 years. There he had access to the count's

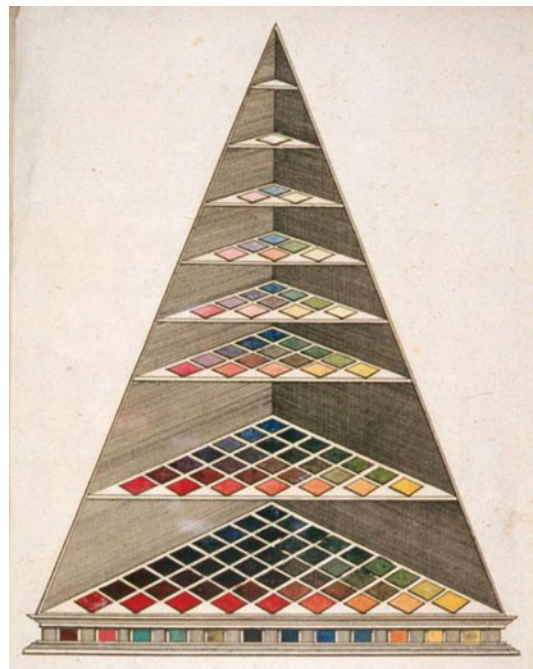
large library and was able to travel widely in Europe with his charges. In 1755 he began to publish scientific articles on a number of subjects. In 1756 he traveled with his pupils to Göttingen in Germany where he met Tobias Mayer and was elected a member of the *Königliche Gesellschaft der Wissenschaften* (Royal Society for the Sciences). In 1759 he published his work on light measurement, *Photometria* [1], introducing his mathematical formula for the law of absorption of light (Lambert's law), described nonmathematically a few years earlier by Pierre Bouguer. In 1764 he followed an invitation by the Swiss mathematician Leonhard Euler to come to Berlin where, after some initial difficulties, Frederic II appointed him to a position in the *Königlich-Preussische Akademie der Wissenschaften* (Royal Prussian Academy of Sciences). Lambert established an important position as mathematician, physicist, astronomer, and philosopher. He also had considerable interest in the art of painting. Among many other achievements, Lambert was the first to mathematically prove the irrationality of the number π . He died on September 25, 1777 in Berlin, Germany [2].

Major Accomplishments/Contributions

In 1758 Tobias Mayer presented his public lecture on a three-dimensional color order system, of



which a report was published in *Göttingische Anzeigen für gelehrte Sachen* (Göttingen reports on learned matters), read by Lambert. In 1768 Lambert published an article *Mémoire sur la partie photométrique de l'art du peintre* (Dissertation on the photometric component of the art of the painter) [3] in which he discussed the effect of light on the appearance of colored materials. Soon after and as a result of Mayer's premature death, he began work on an implementation of Mayer's conceptual double pyramidal system, with assistance of the Prussian court painter Benjamin Calau (1724–1785). The result was published in 1772 as *Beschreibung einer mit Calauischem Wachse ausgeführten Farbenpyramide* (Description of a color pyramid painted with Calau's wax) [4] containing a hand-colored abbreviated version of the conceptual color pyramid (Fig. 1) [3]. Lambert and Calau had to solve a number of practical issues, for example, they determined the relative strength of the pigments they used. When mixing the three primary paint samples, they obtained near-black



Lambert, Johann Heinrich, Fig. 1 Lambert's illustration of his triangular color pyramid, 1772

colors. As a result, Lambert saw no need for the lower half of Mayer's double pyramid. The Lambert-Calau pyramid is the first three-dimensional illustrated representation of a systematically developed color solid.

References

1. Lambert, J.H.: Photometria, sive de Mensura et Gradibus Luminis, Colorum et Umbrae. Klett, Augsburg (1760)
2. Kraus, A.: Lambert, Johann Heinrich. In: Neue Deutsche Biographie, Berlin: Duncker & Humblot GmbH, vol. 13, pp. 437–439. (1982)
3. Lambert, J.H.: Mémoire sur la partie photométrique de l'art du peintre. Mémoires de l'Académie des Sciences de Berlin **XXIV**, 313–334 (1768)
4. Lambert, J. H.: Beschreibung einer mit Calausischem Wachse ausgeführten Farbenpyramide. Haude und Spener, Berlin (1772). English translation available at www.iscc.org

Land Retinex Theory

► Retinex Theory

Lanthanide Ions

► Lanthanoid Ion Color

Lanthanides

► Lanthanoid Ion Color

Lanthanoid Ion Color

Richard J. D. Tilley
Queen's Buildings, Cardiff University,
Cardiff, UK

Synonyms

Lanthanide ions; Lanthanides; Rare earths; Rare-earth elements; Rare-earth ions

Definition

The lanthanoids (often designated Ln) are the 15 elements with atomic numbers 57 (lanthanum) to 71 (lutetium).

Colors, Electron Configurations, and Energy Levels

Color

Most of the lanthanoid ions exhibit rather pale characteristic colors when introduced into transparent solids or in water solutions, the most important being the Ln^{3+} state (Table 1). These

Lanthanoid Ion Color, Table 1 Characteristic color of lanthanoid ions

Element	Ion ^a	4f occupancy ^b	Color
Lanthanum	La^{3+}	$4f^0$	Colorless
Cerium	Ce^{4+}	$4f^0$	Colorless
	Ce^{3+}	$4f^1$	Pale yellow
Praseodymium	Pr^{4+}	$4f^1$	Colorless
	Pr^{3+}	$4f^2$	Green
Neodymium	Nd^{3+}	$4f^3$	Lilac/violet
Promethium	Pm^{3+}	$4f^4$	Pink
Samarium	Sm^{3+}	$4f^5$	Pale yellow
	Sm^{2+}	$4f^6$	Red/green
Europium	Eu^{3+}	$4f^6$	Pink
	Eu^{2+}	$4f^7$	Red brown
Gadolinium	Gd^{3+}	$4f^7$	Colorless
Terbium	Tb^{4+}	$4f^7$	Colorless
	Tb^{3+}	$4f^8$	Pale pink
Dysprosium	Dy^{3+}	$4f^9$	Pale yellow
	Dy^{2+}	$4f^{10}$	Brown
Holmium	Ho^{3+}	$4f^{10}$	Yellow
Erbium	Er^{3+}	$4f^{11}$	Pink
Thulium	Tm^{3+}	$4f^{12}$	Green
	Tm^{2+}	$4f^{13}$	Green
Ytterbium	Yb^{3+}	$4f^{13}$	Colorless
	Yb^{4+}	$4f^{14}$	Colorless
Lutetium	Lu^{3+}	$4f^{14}$	Colorless

^a Ln^{3+} is the principal ionic state encountered

^bThere is uncertainty about the exact f-orbital occupation in many compounds

colors arise from electronic transitions between the ionic ground state and energy levels derived from 4f electron configurations lying between 1.77 and 3.10 eV above it, giving absorption maxima in the visible wavelength range (700–400 nm). Of more practical importance is color produced when ions excited to higher energy levels fall back to these 4f-derived levels and thence to the ground state, giving rise to characteristic visible emission spectra, which are used in many applications including fluorescent printing inks used as security markers on banknotes.

The 4f electrons in the lanthanoids are well shielded beneath an outer electron configuration, ($5s^2 5p^6 6s^2$), and so are little influenced by the surrounding solid matrix, and although crystal-field effects (see Transition-Metal Ion Colours) contribute to the fine structure of the electronic spectra of the lanthanoid ions, these do not have a gross effect upon the color. This implies that the most important optical properties attributed to the 4f electrons on any particular lanthanoid ion do not (usually) depend significantly upon the host structure, so that lanthanoid elements find use in phosphors, lasers, and other light-emitting devices, where a host lattice can be chosen with respect to processing conditions without significantly changing the desirable color properties of the ion.

Lanthanoid Free-Ion Energy Levels

The energy levels of a free lanthanoid ion are usually labeled with atomic *term symbols* derived by *Russell-Saunders (LS) coupling* (although other coupling schemes are also used in this respect). A *term* is a set of states which are very similar in energy, and the appropriate term symbol is written as ^{2S+1}L where L is a many-electron quantum number describing the total orbital angular momentum of all of the electrons surrounding the atomic nucleus and S is a many-electron quantum number representing the total electron spin. The superscript ($2S+1$) is called the *multiplicity* of the term and is given a name: 1, singlet; 2, doublet; 3, triplet; 4, quartet; and so on. The total angular momentum quantum number L is given a letter symbol: $L = 0$, S; $L = 1$, P; $L = 2$, D; $L = 3$, F; and

thereafter alphabetically, omitting J. The energies of the terms must be determined by quantum mechanical calculations, except for that of the ground state, which is given by Hund's second rule: the ground state is the term with the highest multiplicity and, if more than one term of the same multiplicity is present, by that with the highest L value.

The term symbol does not account for the true complexity of the energy levels of the lanthanoid ions. This arises from the interaction between the spin, S , and the orbital momentum, L , called *spin-orbit coupling*. For this the quantum number, J , is needed. It is given by:

$$J = (L + S), (L + S - 1) \dots |L - S|$$

where $|L - S|$ is the modulus (absolute value) of the quantity $L - S$. Each value of J represents a different energy level. The new quantum number is incorporated as a subscript to the term, now written $^{2S+1}L_J$ and called a *level*. It is found that a singlet term always gives rise to 1 level, a doublet 2, a triplet 3, and so on (Table 2). The energies of these levels can be sorted in terms of energy by Hund's third rule: the level with the lowest energy

Lanthanoid Ion Color, Table 2 Ground state terms and levels of principal lanthanoid ions

Ion	Term	Levels ^a
La ³⁺ f ⁰	¹ S	¹ S ₀
Ce ³⁺ f ¹	² F	² F _{5/2} ² F _{7/2}
Pr ³⁺ f ²	³ H	³ H ₄ ³ H ₅ ³ H ₆
Nd ³⁺ f ³	⁴ I	⁴ I _{9/2} ⁴ I _{11/2} ⁴ I _{13/2} ⁴ I _{15/2}
Pm ³⁺ f ⁴	⁵ I	⁵ I ₄ ⁵ I ₅ ⁵ I ₆ ⁵ I ₇ ⁵ I ₈
Sm ³⁺ f ⁵	⁶ H	⁶ H _{5/2} ⁶ H _{7/2} ⁶ H _{9/2} ⁶ H _{11/2} ⁶ H _{13/2} ⁶ H _{15/2}
Eu ³⁺ f ⁶	⁷ F	⁷ F ₀ ⁷ F ₁ ⁷ F ₂ ⁷ F ₃ ⁷ F ₄ ⁷ F ₅ ⁷ F ₆
Eu ²⁺ f ⁷	⁸ S	⁸ S _{7/2}
Gd ³⁺ f ⁷	⁸ S	⁸ S _{7/2}
Tb ³⁺ f ⁸	⁷ F	⁷ F ₆ ⁷ F ₅ ⁷ F ₄ ⁷ F ₃ ⁷ F ₂ ⁷ F ₁ ⁷ F ₀
Dy ³⁺ f ⁹	⁶ H	⁶ H _{15/2} ⁶ H _{13/2} ⁶ H _{11/2} ⁶ H _{9/2} ⁶ H _{7/2} ⁶ H _{5/2}
Ho ³⁺ f ¹⁰	⁵ I	⁵ I ₈ ⁵ I ₇ ⁵ I ₆ ⁵ I ₅ ⁵ I ₄
Er ³⁺ f ¹¹	⁴ I	⁴ I _{15/2} ⁴ I _{13/2} ⁴ I _{11/2} ⁴ I _{9/2}
Tm ³⁺ f ¹²	³ H	³ H ₆ ³ H ₅ ³ H ₄
Yb ³⁺ f ¹³	² F	² F _{7/2} ² F _{5/2}
Lu ³⁺ f ¹⁴	¹ S	¹ S ₀

^aIn ascending energy order, ground state level **bold**

is that with the lowest J value if the valence shell is up to half full and that with the highest J value if more than half full. The separation between the components of the spin-orbit energy levels is of the order of 0.1–0.25 eV. (For comparison, crystal-field splitting of these energy levels, which is due to the interaction of the f orbitals with the surrounding atoms in a solid or liquid, is about 0.01 eV.)

In the presence of a magnetic field, the spin-orbit levels are split further due to the Zeeman effect. The same is true of static electric fields, where it is called the Stark effect. In both cases, atoms or ions in a gas or free space will show an average effect because of the motion of the particles. However, in a crystal the atom and ion positions are more or less fixed, and the application of either magnetic or electric fields along certain symmetry directions will, in general, cause different degrees of splitting of the levels than the same fields applied along other symmetry directions. Zeeman splitting has been used to change the optical properties of lanthanoid ions in optical-magnetic materials.

Selection Rules

Electron transitions are governed by selection rules that give the *probability* that the transition will occur. Transitions between energy levels derived purely from f orbitals are forbidden by the Laporte selection rule. However, this rule may break down for ions in compounds. The main reason for this is that a degree of mixing between s , p , d , and f orbitals can occur when an ion is not located at a center of symmetry. As s , p , or d to f transitions are allowed, transitions giving rise to color are also allowed, to that a degree corresponding to the amount of orbital mixing achieved. In addition, transitions are only allowed between states of the same multiplicity, called *spin-allowed transitions*. Transitions between states of differing spin can be weakly allowed and in some circumstances can also contribute to observed lanthanoid color. The weak colors exhibited by Ln^{3+} ions are primarily due to these restrictions, especially when compared to typical crystal-field colors of the 3d transition-metal ions (see Transition-Metal Ion Colours).

Some Lanthanoid Absorption Colors

The $4f^0$ ions La^{3+} and Ce^{4+} and f^{14} ions Yb^{4+} and Lu^{3+} have no $f-f$ energy levels and are colorless. The colorless ions Gd^{3+} and Tb^{4+} have a stable $4f^7$ configuration, and there are no energy levels in the appropriate energy range to give rise to color. The same is true for the $4f^{13}$ ion Yb^{3+} .

Ce^{3+} and Eu^{2+}

The lowest energy levels of Ce^{3+} , arising from the single f electron, are $^2F_{7/2}$ and $^2F_{5/2}$. The next higher energy state for Ce^{3+} is the $5d$ level. Due to interaction of the more exposed $5d$ electrons with the surrounding crystal structure, this is broadened into a band of energies, which may also overlap with another broadened band of energies derived from the $6s$ energy level (Fig. 1a). Transitions between the $5d$ band and $4f$ levels are allowed, and the colors produced by transitions of this type are intense. This transition absorbs at violet end of the spectrum, and the absorption band often encroaches into the visible with a consequence that to the eye compounds are perceived as pale yellow.

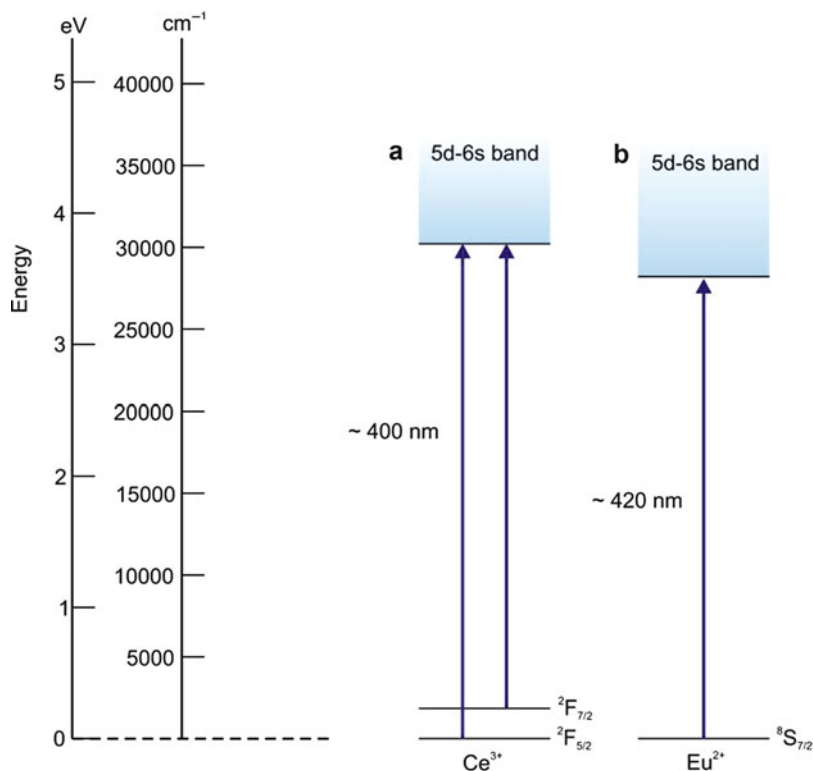
Eu^{2+} , with a configuration $4f^7$, also has a simple energy level diagram because the energy of the state obtained by transferring an f electron to the outer $5d$ orbitals is lower than the other $4f$ energy levels (Fig. 1b). As in the case of Ce^{3+} , transitions from the ground state to the upper energy band are allowed. The energy gap is slightly smaller than in the case of Ce^{3+} , and so the absorption moves slightly deeper into the visible spectrum. Because of this, the color of compounds is described as red brown.

Note that as the d orbitals interact strongly with the surrounding anions, the exact position of the band depends upon the host crystal. Thus, the colors of Ce^{3+} and Eu^{2+} compounds, unusually for lanthanoid ions, vary with host structure.

Pr^{3+} , Tm^{3+} , and Nd^{3+}

The presence of Pr^{3+} ions generally colors the host matrix green. The principal transitions that contribute to the absorption spectrum are from the ground state 3H_4 to 3P_0 , at approximately 485 nm (blue green), to 3P_1 at approximately

Lanthanoid Ion Color,
Fig. 1 Energy level
 diagrams (schematic) for
 Ce^{3+} and Eu^{2+} ions

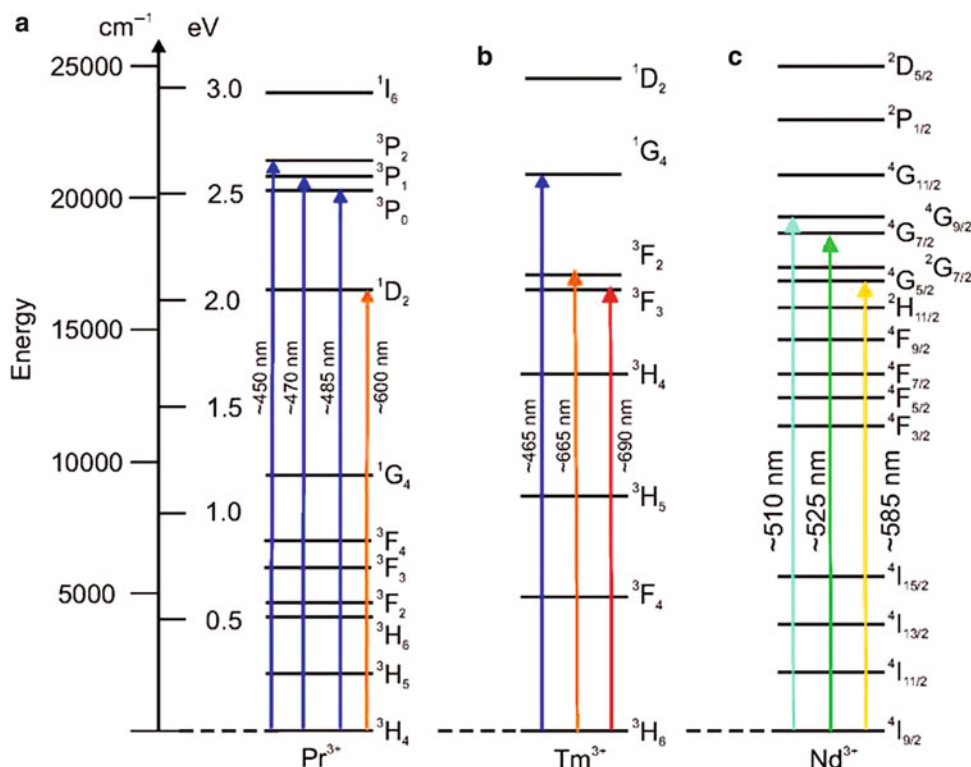


470 nm (blue), and to 3P_2 at approximately 450 nm (blue). A weak spin-forbidden transition to 1D_2 at approximately 590 nm (orange) is also present. Between these absorption peaks is a window of low absorption in the green region of the spectrum (Fig. 2a). The other “green” ion is Tm^{3+} . Here the main transitions are from the ground state 3H_6 to 3F_3 at approximately 690 nm (red) and to 3F_2 at approximately 665 nm (red) and a weak spin-forbidden transition to 1G_4 at approximately 465 nm (blue). Between these absorption bands is a similar green transmission window to Pr^{3+} (Fig. 2b). The Nd^{3+} ion tends to impart a lilac hue to materials. The main transition here is from the ground state $^4I_{9/2}$ to $^4G_{5/2}$, which absorbs strongly at approximately 585 nm (yellow), together with a transition to $^4G_{7/2}$ at approximately 525 nm (green) and $^2G_{9/2}$ at 510 nm (blue green). These absorption peaks remove the middle part of the optical spectrum, leaving both the violet and red extremes, so imparting a lilac hue to compounds.

Some Applications of Lanthanoid Fluorescence Colors

Trichromatic Fluorescent Lamps

Trichromatic (color 80) fluorescent lamps use phosphors with active lanthanoid ions. Generally the lanthanoid ions can absorb a wide range of ultraviolet radiation efficiently, exciting the ions from the 4f-derived ground state to a broad band of energies formed from interaction of 5d and 6s band. Energy is then lost internally, in effect causing the phosphor matrix to warm slightly, until the sharp 4f-derived energy levels are reached. Photon emission between these energy levels then occurs, giving color output. The favored red emitter is Eu^{3+} doped into Y_2O_3 , ($\text{Y}_2\text{O}_3:\text{Eu}$), with the Eu^{3+} ions occupying the Y^{3+} sites. The ground state of Eu^{3+} is 7F_0 . A transition from this state to the higher energy band 5d band absorbs the ultraviolet radiation given off by excited mercury atoms at 254 nm. Subsequent internal energy loss leaves the ion in the 5D_0 level. The main



Lanthanoid Ion Color, Fig. 2 Energy level diagrams (schematic) for (a) Pr^{3+} , (b) Tm^{3+} , and (c) Nd^{3+}

optical transition is between this level and $^7\text{F}_2$, producing emission at 611 nm (Fig. 3a). The green emission is from Tb^{3+} in host matrices $\text{La}(\text{Ce})\text{PO}_4$, $\text{LaMg}(\text{Ce})\text{Al}_{11}\text{O}_{19}$, or $\text{La}(\text{Ce})\text{MgB}_5\text{O}_{10}$, in which the Tb^{3+} ions replace La^{3+} . Tb^{3+} absorbs the mercury emission poorly so is coupled with a sensitizer, usually Ce^{3+} , which is able to absorb the 254 nm wavelength mercury radiation efficiently by way of the 5d band. This absorbed energy is then transferred to the Tb^{3+} ions. The green emission, at wavelength close to 540 nm, is mainly from a $^5\text{D}_4 \rightarrow ^7\text{F}_5$ transition (Fig. 3b). Three other peaks of lesser intensity occur: $^5\text{D}_4 \rightarrow ^7\text{F}_6$, 489 nm; $^5\text{D}_4 \rightarrow ^7\text{F}_4$, 589 nm; and $^5\text{D}_4 \rightarrow ^7\text{F}_3$, 623 nm. The blue emission is produced by Eu^{2+} ions. The excitation and emission is directly to and from the 5d-derived band. The position of this band is strongly influenced by the host structure (see above) and the usual tricolor lamp phosphor, $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}$ is chosen so as to have a suitable blue emission, with a maximum at 450 nm (Fig. 3c).

Plasma Displays

Plasma display panels are made up of a pair of glass plates containing a series of cells each of which acts as a miniature fluorescent lamp as described above. Each lamp is several hundred microns in size, and there are several million such lamps in a display. Each pixel consists of three lamps, giving off red, blue, and green light. The working gas in the cells is a mixture of helium and xenon. When a high voltage is applied across two electrodes above and below a well, the gas is excited into a state that emits ultraviolet radiation, with principal wavelengths of 147 nm and 172 nm. Each well is coated internally with a red, green, or blue phosphor. The main lanthanoid phosphors used at present are a yttrium gadolinium borate doped with europium, $(\text{Y}, \text{Gd})\text{BO}_3:\text{Eu}^{3+}$, which gives a red emission, barium magnesium aluminate doped with europium, $\text{BaMgAl}_{14}\text{O}_{23}:\text{Eu}^{2+}$, for blue emission (see Fig. 3). The green emission utilizes Mn^{2+} rather than a lanthanoid ion.

Lanthanoid Ion Color,

Fig. 3 Energy level diagrams (schematic) for (a) Eu^{3+} , (b) $\text{Ce}^{3+}/\text{Tb}^{3+}$, and (c) Eu^{2+} in trichromatic lamp phosphors. *ET* energy transfer

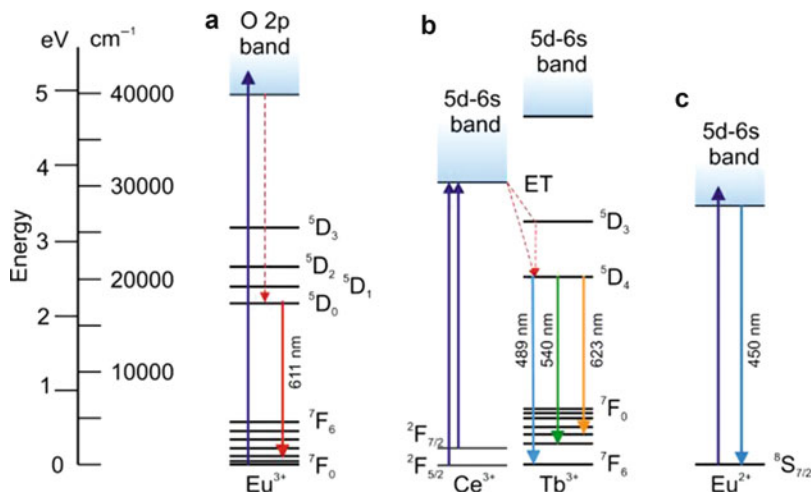
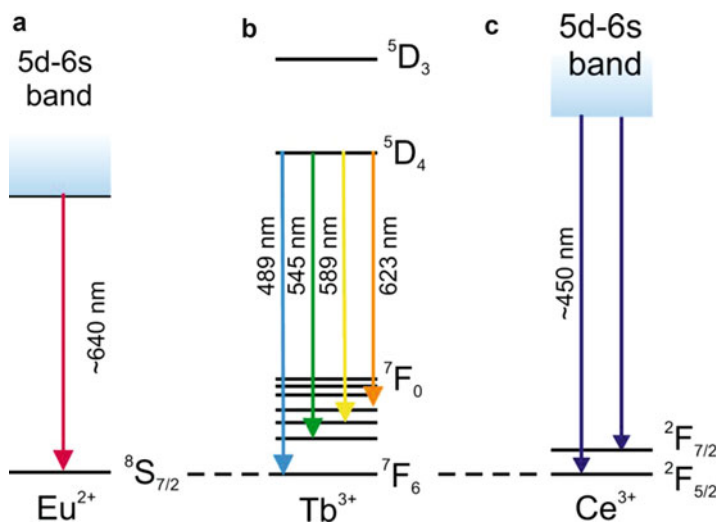
**Lanthanoid Ion Color,**

Fig. 4 Energy level diagrams (schematic) for (a) Eu^{2+} , (b) Tb^{3+} , and (c) Ce^{3+} in ac powered Thin Film Electroluminescent Display (ACTFEL) displays

**Phosphor Electroluminescent Displays**

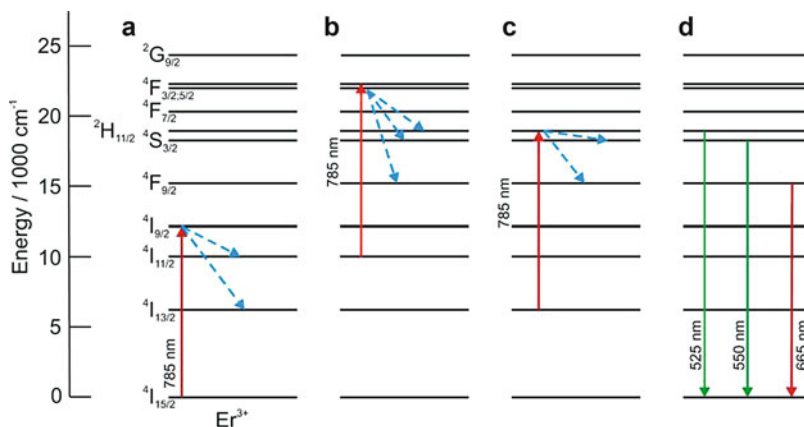
Electroluminescent displays containing a thin film of a phosphor, called thin film electroluminescent (TFEL) displays, find use as flat panel color displays and backlighting in products such as instrument panels. The most promising devices use *ac* supplies in a thin film electroluminescent (ACTFEL) display. Under the influence of an applied electric field, electrons enter the phosphor at the junction with a surface insulating coating. These are accelerated under the influence of the field until they collide with the luminescent centers in the phosphor, transferring energy in the process. The excited luminescent centers then

fall back to the ground state and release energy by light emission.

Red emission is from calcium sulfide doped with europium ($\text{CaS}:\text{Eu}^{2+}$), the color being generated by the transition from the 5d band to the ground state $^8\text{S}_{7/2}$ (Fig. 4a). At first sight this is surprising as the Eu^{2+} derived tricolor lamp phosphor has a blue output. However the position of the upper energy band depends upon the interaction of the d orbitals with the surrounding crystal, and in ZnS the softer bonding gives a broad emission centered close to 640 nm. Green emission is produced by zinc sulfide doped with terbium ($\text{ZnS}:\text{Tb}^{3+}$), with an output at a wavelength close

Lanthanoid Ion Color,

Fig. 5 Upconversion using Er^{3+} doped into CeO_2 , schematic: (a) GSA + relaxation; (b) ESA + relaxation; (c) ESA + relaxation; (d) light emission

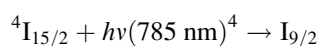


to 545 nm, mainly from a $^5\text{D}_4 \rightarrow ^7\text{F}_5$ transition (Fig. 4b). Three other peaks of lesser intensity occur: $^5\text{D}_4 \rightarrow ^7\text{F}_6$, 489 nm; $^5\text{D}_4 \rightarrow ^7\text{F}_4$, 589 nm; and $^5\text{D}_4 \rightarrow ^7\text{F}_3$, 623 nm. Blue emission still poses a problem for these displays, but the thiogallates CaSr_2S_4 , SrGa_2S_4 , and BaGa_2S_4 doped with the $4f^1$ ion Ce^{3+} are currently favored. The transitions between the 5d band and the $4f^1$ ground state doublet $^2\text{F}_{5/2}$ and $^2\text{F}_{7/2}$ are both in the blue region of the spectrum centered at 459 nm for the Ca compound and 445 nm for the Sr and Ba phases (Fig. 4c).

Upconversion

The conversion of infrared radiation to visible is variously known as upconversion, frequency upconversion, anti-Stokes fluorescence, or cooperative luminescence. The majority of studies of upconversion have involved the lanthanoid ions Er^{3+} , Tm^{3+} , Ho^{3+} , and Yb^{3+} . At present the most efficient upconversion materials for infrared to visible conversion are lanthanoid fluorides such as NaLnF_4 doped with $\text{Yb}^{3+}/\text{Er}^{3+}$ or $\text{Yb}^{3+}/\text{Tm}^{3+}$ couples. The energy for upconversion can be gained by the active ion via several competing energy transfer processes. In principle, the simplest is for the active ion to pick up photons in two distinct steps. The first photon excites the ion from the ground state to an excited energy level, a process referred to as *ground state absorption* (GSA). A subsequent photon is then absorbed to further promote the excited ion to a higher energy level again – *excited state absorption* (ESA). The

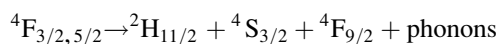
oxide CeO_2 doped with approximately 1 % Er^{3+} exhibits upconversion in this way. Irradiation with near-infrared photons with a wavelength close to 785 nm excites the Er^{3+} ions from the $^4\text{I}_{15/2}$ ground state to the $^4\text{I}_{9/2}$ level:



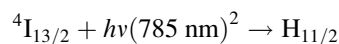
Subsequent internal energy loss drops the energy to the $^4\text{I}_{11/2}$ and $^4\text{I}_{13/2}$ levels (Fig. 5a). The ions are then excited by ESA. Those in the $^4\text{I}_{11/2}$ energy level are excited to the $^4\text{F}_{3/2, 5/2}$ doublet:



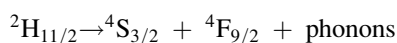
These states subsequently relax to the $^2\text{H}_{11/2}$, $^4\text{S}_{3/2}$ and $^4\text{F}_{9/2}$ levels:



The ions in the $^4\text{I}_{13/2}$ energy level follow a similar path, being excited to the $^2\text{H}_{11/2}$ energy level:

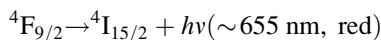
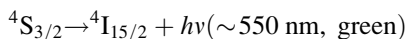
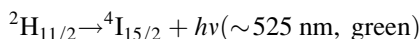


then subsequently relax to the $^4\text{S}_{3/2}$ and $^4\text{F}_{9/2}$ levels:



The result of this is that the levels $^2\text{H}_{11/2}$, $^4\text{S}_{3/2}$ and $^4\text{F}_{9/2}$ are populated to varying degrees, depending

upon the precise details of the excitation and relaxation steps. Subsequent loss of energy from these levels gives rise to green and red emission:

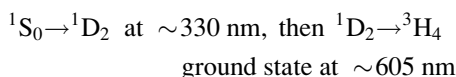
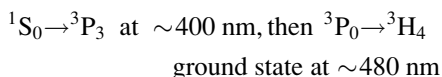


All upconversion spectra from Er^{3+} , including those using different mechanisms, are similar, but the relative intensities of the three peaks vary with concentration of the ions and the nature of the host matrix.

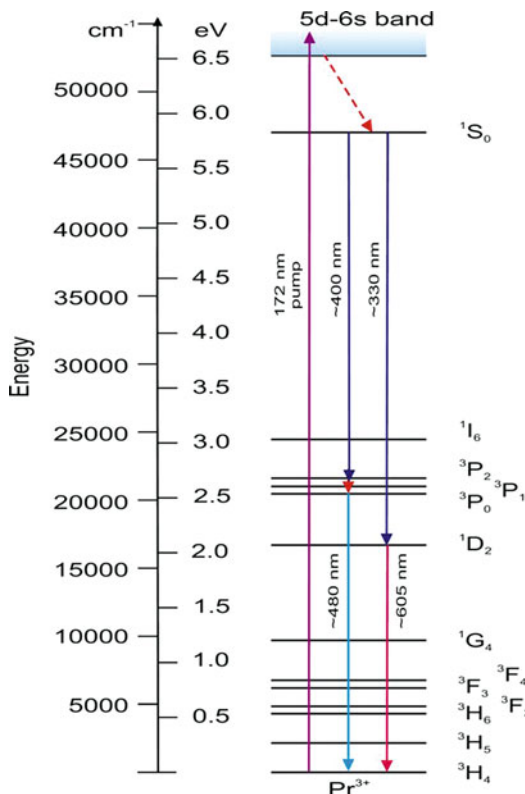
Energy transfer, in which input radiation is picked up by a sensitizer and is then passed to the emitter, is the mechanism of operation of host structures containing the co-dopants $\text{Yb}^{3+}/\text{Er}^{3+}$ which gives a strong green emission and $\text{Yb}^{3+}/\text{Tm}^{3+}$ that gives a blue emission. The ion that absorbs the incoming infrared radiation is the Yb^{3+} ion, which then transfers energy to Er^{3+} or Tm^{3+} centers. Other upconversion mechanisms are also known.

Quantum Cutting

Quantum cutting is the reverse of upconversion, as one high-energy photon is processed (i.e., cut) to give out several lower energy photons, typically transforming ultraviolet to visible. There are several mechanisms for quantum cutting; here photon cascade emission, exhibited by Pr^{3+} ions, is described. Initial absorption of high-energy 185 nm ultraviolet photons results in excitation to the 5d-6s energy band (Fig. 6). Subsequent relaxation takes the ion to the 1S_0 level. Thereafter, the transitions giving rise to visible output are the following:



Other ions such as Tb^{3+} are also candidates for quantum cutting devices, but the mechanisms involved are more complex than that with Pr^{3+} .



Lanthanoid Ion Color, Fig. 6 Visible light emission (quantum cutting) via photon cascade emission from Pr^{3+} ions irradiated by ultraviolet light

Cross-References

► [Transition-Metal Ion Colors](#)

References

1. Tilley, R.J.D.: Chapter 7. In: *Colour and the Optical Properties of Materials*, 2nd edn. Wiley, Chichester (2011)
2. Nassau, K.: Chapter 4. In: *The Physics and Chemistry of Colour*, 2nd edn. Wiley, New York (2001)
3. Huang, C.-H. (ed.): *Rare Earth Coordination Chemistry*. Wiley, Singapore (2010)
4. Linganna, K., Jayasankar, C.K.: *Luminescence Spectroscopy of the Lanthanides*, Scholars Press (2013)
5. Hänenen, P., Härmä, H. (eds.): *Lanthanide Luminescence*. Springer, Heidelberg (2011)
6. Cotton, S.: *Lanthanide and Actinide Chemistry*. Wiley, Chichester (2006)

Layering

► Compositing and Chroma Keying

Le Blon, Jacob Christoph

Rolf G. Kuehni
Charlotte, NC, USA

Biography

Le Blon was born on May 2, 1667 in Frankfurt am Main, Germany, a descendant of Huguenots fleeing France in 1576, having settled there. His grandmother was a daughter of the artist and engraver Matthaus Merian the Elder (1593–1650). Showing an early interest in engraving and painting, he had, sometime between 1696 and 1702, an extended stay in Rome where he is reported to have studied art under the painter and engraver Carlo Maratta

(1625–1713) [1]. Around 1702 Le Blon moved to Amsterdam, where he worked as a miniature painter and engraver. In 1708/1709 he is known to have made colorant mixing experiments in Amsterdam, and in 1710 he made his first color prints with yellow, red, and blue printing plates. In 1717 he moved to London where he received a royal patent for the three-color printing and a related textile weaving process [2]. In 1722 he published a small book on painting, *Coloritto*, in French and English [3]. There he stated that “Painting can represent all visible objects with three colors, yellow, red, and blue.” During his stay in England, he produced several dozen images printed from three or four plates (the fourth for black ink) in multiple copies that initially sold well in England and on the continent. In the long run his enterprise did not succeed, however. Le Blon left England in 1735, moving to Paris where he continued producing prints by his method. In 1740 he began work on a collection of anatomical prints for which he had a solid list of subscribers. He died on May 16, 1741 in Paris. A detailed technical description of Le Blon’s method was published in 1756 by Antoine Gautier de Montdorge who supported him during his final years in Paris [4].



J. C. Le Blon, Head of a woman, ca. 1720 (three-color printing process)

Major Accomplishments/Contributions

The idea of three chromatic primaries, yellow, red, and blue, was quite well established in Le Blon’s time among painters, graphically represented by Aguilonius in 1613 [5] and described by R. Boyle in 1664 (p. 220) [6]. What was new in Le Blon’s work is that he applied this concept to color printing of images in an entirely new fashion making greater and much subtler detailing and coloration possible. It required experience in deconstructing an image in terms of color so that printing multiple copies, based on only three or four plates, produced good quality coloration. It required the ability to mentally resolve the image into its presumed primary chromatic components and understanding and predicting the effects of superimposed printing inks in certain areas, for which extensive trial and error work was required.

Le Blon manually engraved copperplates, using the mezzotint process, with the relative components of the three primary colors printed successively in registration in the sequence blue, yellow, and red onto the paper substrate. As he gained experience, he at times used a fourth plate printing in black to achieve greater tonality and contrast, thus employing an early version of the CMYK process. Le Blon used the pigments Prussian blue, Stil de grain (yellow lake), a mixture of red lake and carmine for red, and a common printer's black ink [4]. The pigments were dispersed in copal tree resin dissolved in copal oil to make the inks. The technical problems associated with the process prevented it from becoming a standard method and lithographic printing of color images from up to a dozen wood engravings or stones per image continued until H. E. Ives' invention of the chromatic halftone printing process ca. 1890.

References

1. Lilien, O.M.: Jacob Christoph Le Blon, 1667–1741: Inventor of three- and four colour printing. Hiersemann, Stuttgart (1985)
2. Lowengard, S.: Jacob Christoph Le Blon's system of three-color printing and weaving. In: The Creation of Color in the 18th century Europe. Columbia University Press, New York (2006). http://www.gutenberg-e.org/lowengard/C_Chap14.html. Accessed 11 May 2015
3. Le Blon, J. C.: Coloritto, or the Harmony of Colouring in Painting: Reduced to Mechanical Practice (with parallel French text). London (ca.1725)
4. Gautier de Montdorge, A.: L'art d'imprimer les tableaux. Traité d'après les écrits, les opérations et les instructions verbales de J.-C. Le Blon. Mercier, Paris (1756)
5. Aguilonius, F.: Opticorum Libri Sex. Plantin, Antwerp (1613)
6. Boyle, R. Experiments and Considerations Touching Colours. Herringman, London (1664)

Light Bounces

► [Global Illumination](#)

Light Depreciation

► [Lumen Depreciation](#)

Light Distribution

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Candela distribution](#); [Luminous intensity distribution](#)

Definition

A luminaire property that indicates the values of the luminous intensities radiated in all relevant directions by the luminaire. Usually the luminous intensities are given as intensities per 1,000 lamp lumen when the lamp or lamps are operated under standard test conditions: I/1,000 lm.

Luminaire Performance Characteristic

The light distribution of a luminaire represents the most important performance characteristic of a luminaire. It is the basis for the determination of photometric characteristics of luminaires such as upward and downward light output ratio, utilization factors for defined areas or zones, beam angles, and glare figures. Luminaire manufacturers produce for their luminaires photometric datasheets which show these photometric characteristics together with a graphical representation of the light distribution itself.

Measurement

Strictly speaking, the light distribution can only be measured for a point source. In practice the measurement inaccuracy will be negligible provided the optical path length of the measuring setup is at least ten times the length of the light emitting surface of the luminaire. For narrow beam projector type of luminaires, greater optical path lengths are required. The instrument used for

measuring light distributions is called a goniophotometer. It exists of a mounting support for the luminaire, a photocell, and a, usually complex, system that rotates the combination of photocell and luminaire such that the intensities from all different light directions can be measured. Mirrors are used to keep the actual dimensions as small as possible while keeping the required optical path length. It is important that the luminaire rotates in such a way that its normal operating position is at all times maintained. The measurements should be carried out under standard test conditions. These conditions refer to ambient temperature, air movement, mounting position, and lamp ballast type. The coordinating systems used for indicating the directions are either the so-called B-beta (usually used for projector type of luminaires) or the C-gamma system. The B angles of the B-beta system can be understood as the angles the pages of a book make when the book axis is held horizontally. The beta angles are then angles in each page measured from the middle of the book axis. The C-gamma system in the same analogy can be understood as a book with its axis held vertically. C angles are the angles the pages make, and gamma angles are angles in the pages from the middle of the book axis.

Intensity Table

All lighting calculations of lighting installations are based on the light distribution of the luminaires used. For this purpose the light distribution is given as a two-dimensional table, the so-called I-table. There are some different formats for these I-tables such as CIE, IESNA, Eulumdat, CIBSE, and some luminaire manufacturer's specific formats. Eulumdat is the most popular format in Europe, whereas the IESNA format is more popular in the USA. The CIE format is an attempt to define a "recommended file format for electronic transfer of luminaire photometric data" on a global level, in order to overcome the regionally different formats currently in use. Many light calculation programs can use some different formats as input. Moreover, free software for conversion

from one format to the other is available as well. The layout of the I-table is different for different type of luminaires. So have road lighting luminaire I-tables more values around the direction of the horizon than for other directions because these values very much determine glare, general lighting interior luminaire I-tables have a somewhat more regular layout because of the more diffuse character of these luminaires and flood and accent lighting luminaires (projector type of luminaires) have I-tables with extra values around the direction of the maximum beam intensity.

Cross-References

► [Luminaires](#)

Light Pollution

Terry McGowan

Lighting Ideas, Inc, Cleveland Hts., OH, USA

Synonyms

[Glare](#); [Sky glow](#); [Unwanted light especially referring to unwanted electric lighting at night](#)

Definition

Outdoor electric lighting at night which is unwanted, unneeded, or wasteful and which results in glare, light trespass and sky glow or other harmful effects on people and the environment.

Light Pollution Recognized

Electric lighting makes travel, working, playing, and other normal living activities possible at night. Indeed, the view of a lighted city from an airplane window or tall building is a hallmark of twenty-first-century civilization, yet little more

than a century and a half ago, there were no such views and human activities outdoors at night were limited to those considered essential. Even then, they were carried out only with difficulty especially tasks involving work or travel. Limitations remained as candles and then gas flames and mantles began to generate more powerful and reliable night lighting, but the benefits were not widespread because of the relatively high cost of providing light especially over large outdoor areas at night.

When electric lighting was invented, one of its first uses was for outdoor illumination. Carbon arc lamps were used initially and then incandescent light sources, which were developed during the 1870s. These became the worldwide standard for all types of lighting until the commercial introduction of the fluorescent lamp in the late 1930s. Outdoor lighting, however, did not dramatically change until mercury lamps, the first of the so-called high-intensity discharge (HID) lamps became widely used for street, parking, and area lighting during the 1960s and then other types of HID lamps – high-pressure sodium (HPS) and metal halide (MH), due to their relatively higher efficacies, transformed the cost and use of outdoor lighting leading to rapid growth and the night time appearance of cities that we see today. Now, however, a new transformation is underway characterized by the growing use of solid state lighting in the form of Light Emitting Diodes (LEDs) which are increasingly being used for outdoor lighting.

Electric lighting has changed the night. And there are negative aspects that were recognized early on by critics who pointed out that bright lights frightened horses, annoyed pedestrians, and produced glare. There were those who also noticed that electric light changed the egg-laying habits of chickens, attracted insects, and modified the leaf and flowering characteristics of certain plants – early examples that hinted at more significant problems that later appeared due to unneeded or unwanted light. But the term “light pollution” did not appear until much later – in the 1980s. That’s when the large, powerful HID lamps began to be used in large quantities for major outdoor lighting systems such as street

lighting, parking areas, signs, sports stadiums, and floodlighting. The stray, uncontrolled light from these new sources as typically used outdoors plus the light from the headlights of cars and light escaping through windows from the interior lighting systems of buildings at night has now transformed the night environment in conspicuous ways including making the night sky less visible. Instead, light into the sky blankets cities with “sky glow,” a fog, or haze of light.

Astronomers, both professional and amateur, were among the first to recognize and talk about sky glow as light pollution as they tried to observe the stars from urban locations. As a first response, professional astronomers moved telescopes and observatories from cities to new locations in remote areas. Then, it was recognized that such places were increasingly rare and costly to develop. Further, most amateur astronomers were not able to move or take advantage of such sites. Members of the public who simply wanted to catch a glimpse of the Milky Way from their backyard or to show their children a comet or star constellation were not able to do so. The result is a growing and widespread reaction to uncontrolled electric lighting outdoors at night.

Efforts began, usually via zoning, planning, or the so-called “nuisance” regulations to limit and control unwanted light especially around astronomical observatories. It was found that light from urban areas could significantly affect the sky brightness 100–200 km away and that minimizing light directed into the sky was a simple and effective way to control such light. The term “light pollution” became familiar and efforts began to understand the various associated problems and to find ways to limit the detrimental effects. A global organization, the International Dark Sky Association (IDA), founded in 1988, focused on the task and continues to work “to preserve and protect the nighttime environment . . . through environmentally responsible outdoor lighting” [1]. Other regional and local organizations also appeared worldwide to raise awareness of the problems, write laws, provide educational materials, and organize local activities such as promoting the use of shielded luminaires. Lighting technical and professional organizations have also

recognized light pollution problems and have sponsored research, issued reports, produced recommended lighting practices, and written model legislation [2-4] in order to control it.

Types of Light Pollution

Light pollution was initially understood as primarily a visual problem caused by unnecessary or uncontrolled outdoor lighting that interfered with the viewing of the sky. But, there are now other considerations, and light pollution has been more broadly defined and classified as follows:

Light trespass – Light generated in one location which, when poorly controlled, travels to another location where it is not wanted. An example would be a streetlight which is intended to light the street surface but also directs light into front yards or bedroom windows along the street.

Glare – Excessive light directed into the eye where it causes visual discomfort or reduces the ability to see. Overly bright unshielded or misaimed luminaires are usually the main sources of glare.

Sky glow – Light directed or reflected into the sky where it is scattered by air molecules, water, and other particles in the atmosphere and reflected diffusely to produce a fog or haze of brightness that reduces the visibility of the night sky.

Light pollution has energy implications since stray or misdirected light is luminous energy which has been converted from electric energy by the light source. That electric energy, the fuel, and other resources which produce it have therefore been wasted, delivering no value to those who pay for those resources.

The IDA estimates that some 30 % of the electric energy used for outdoor lighting goes into the sky. Considering the hours of operation and an average electrical energy rate of \$0.10/kWh (USA), the cost of that wasted energy is well over \$2 billion/year.

Spectral Effects

Electric lighting is the cause of light pollution. However, all light sources are not equally effective at generating light pollution of the types described. The scattering of light in the atmosphere that affects sky glow, for example, is a function of the wavelength of the light (Rayleigh scattering). Such scattering affects how far light pollution will travel. Similarly, the intensity (candlepower) of the luminaire as well as the way the light is aimed and directed matters too.

Astronomers who first sought to alleviate sky glow during the 1970s and 1980s found that the light from low-pressure sodium (LPS) lamps was helpful because all of the light from such lamps is emitted in a narrow range of wavelengths in the yellow part of the spectrum. A telescope equipped with a filter that absorbed those wavelengths and transmitted the rest of the spectrum would therefore see the sky as “dark.” US cities such as Tucson, AZ, and San Jose, CA, installed LPS street and area lighting as a way to maintain dark skies for area observatories while still providing essential lighting for drivers and pedestrians. LPS lamps, however, have poor color rendering and so there were objections to the light. Provisions were made for exceptions so that white light could be used in such applications as outdoor automobile sales lighting and sports facilities. Installations permitted as exceptions usually had other conditions attached that limited total lighting wattage, mandated certain operating hours, or required shielding – often in combinations. Of course, for the amateur astronomer or the casual viewer of the sky, outdoor lighting with LPS lamps and luminaires, which are typically unshielded, offers no advantage since the yellow light is visible to the eye and continues to produce light pollution via glare, light trespass, and sky glow (Fig. 1).

Note that while the spectral characteristics of the light source matter, overall light pollution, including sky glow, is the result of not only how the light is generated but also how it is modified by outdoor surfaces via reflection, refraction, and absorption. However, light pollution which interferes with a view of the sky in any given situation is always directly proportional to one quantity:



Light Pollution, Fig. 1 Example of the effect of light pollution on the visibility of the sky in a residential area near Toronto, Ontario, Canada. *Left:* during an area-wide power failure stars and constellations can be clearly seen.

Right: the same view after the power had been restored with streetlights, floodlights, and other outdoor lighting contributing to sky glow

total luminous power. Measured in lumens, luminous power is radiant power modified by the spectral sensitivity of the human eye. The key to controlling light pollution is to limit or reduce the total lumens being generated from light sources and sent into the night environment. Turning lights off when not needed, shielding luminaires, and reducing illumination levels to the minimum values required are all techniques that can be used to minimize light pollution problems. These are common-sense ideas, and there are many more [5, 6] although what should happen, of course, is that lighting should be designed to incorporate the long-established principles of proper lighting design which include the requirements to control stray light, minimize glare, and not use excessive light for the intended application. The problem is that much outdoor lighting is not “designed” in the sense that major buildings are designed with careful thought given to materials, appearance, function, and purpose. Rather, outdoor lighting

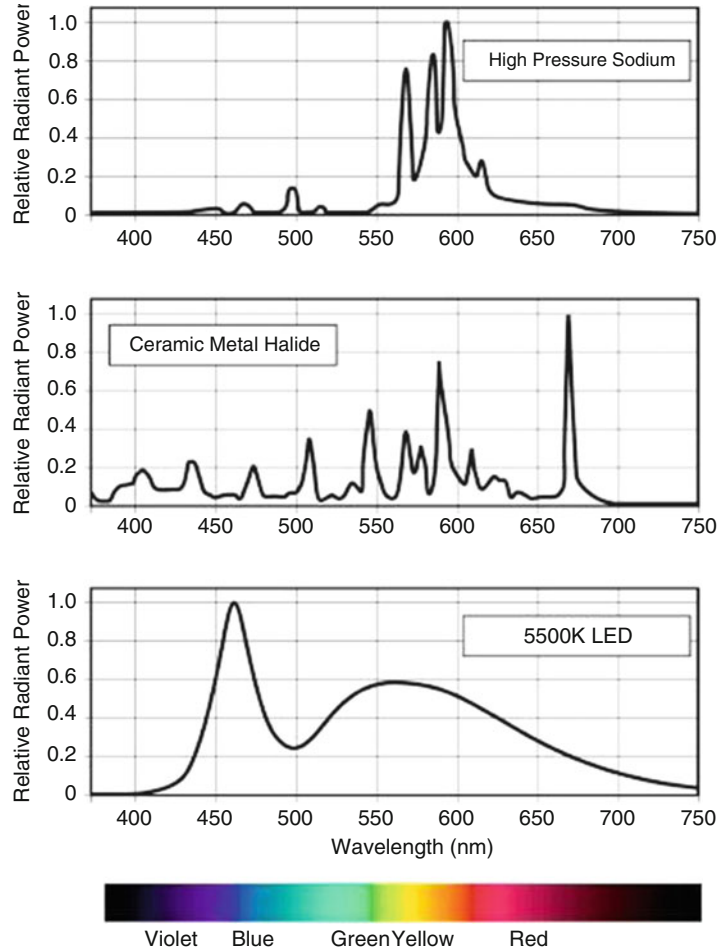
is often a make-do or do-it-yourself affair. It “happens” or factors of cost, expediency, or lack of information drive the installation (Fig. 2).

Other Light Pollution Effects

Research reported during the 1990s has shown that light pollution seriously affects the health and behavior of animals, including mammals, birds, amphibians, and insects. Changes in migration patterns, food gathering and feeding habits, mating, and social behavior have all been documented [7, 8]. However, bird and sea turtle problems have received the most attention.

Birds may become disorientated by lighting to the extent that they die from exhaustion while continuously circling lighted buildings or, confused, they collide with the buildings. According to a Chicago Audubon Society study [9] documented by actual counts of dead birds

Light Pollution, Fig. 2 A comparison of the spectra of light sources commonly used for outdoor lighting. *Top:* high-pressure sodium is the most widely used and is characterized by its familiar orange color. *Middle:* metal halide lamps emit white light but can have a visually “warm” or “cool” color depending upon the chromaticity rating. Emissions, however, are relatively balanced over the visible spectrum. *Bottom:* white-light LED light sources use blue-emitting LEDs to activate phosphors which emit the light shown in the 500 nm and above range. These so-called “blue-rich white-light” LEDs are of concern because light and human health research indicates that eye receptors which control circadian rhythms have a peak sensitivity near 460 nm, similar to the emission curve of the blue LEDs



found at the base of tall buildings, bird deaths due to building collisions were reduced significantly when just the decorative floodlighting of buildings was switched off for most of the night hours during migration seasons.

Baby sea turtles hatch on beaches and have a short time to dig out of the sand and to find their way to the sea before they die by dehydration or are eaten by predators. The turtles appear to use the brightness of the horizon as a visual clue to find the water. Area and streetlighting along beachfront areas confuses the turtles, so efforts have been made to remove, turn off, or modify such systems including changing to amber-colored light sources which, because of their spectral sensitivity, appear less bright to turtle eyes [10].

Human Health Effects

A controversial issue involving light pollution is whether or not outdoor lighting negatively affects human health. Research beginning in the 1970s has indicated that certain uses and kinds of electric light and, particularly, light at night (LAN) when the human body expects and needs darkness for sleep, can be detrimental to health because it interferes with the body’s normal circadian rhythms.

These rhythms mark the periods of alertness and sleep due to the ebb and flow of hormones, such as melatonin, in the body. According to the research, strong circadian rhythms are essential for overall health and mental functioning and particularly for cell repair which helps protect

against major diseases including cancer. The general rule is that people need bright days and dark nights so current research is exploring questions about light intensity, spectrum, timing, and exposure duration to determine the “dose” of light that might result in the interruption of the melatonin cycle and, further, what kind of interruptions lead to an increased risk of disease. The research literature on the subject is already extensive – one compilation now lists over 1,000 citations [11], and how outdoor lighting impacts human health will remain an important subject of research and discussion.

The American Medical Association adopted a resolution in 2009 entitled “Advocating and Support for Light Pollution Control Efforts and Glare Reduction for Both Public Safety and Energy Savings” [12]. Significantly that AMA view was broadened in 2012 when they said, “The (AMA) policy also supports the need for developing lighting technologies that minimize circadian disruption and encourages further research on the risks and benefits of occupational and environmental exposure to light at night” [13].

Outdoor lighting, because such lighting impacts the natural environment and so many people see it, use it, and experience it, has become complicated. It is now a social, technical, medical, and environmental problem that has to be carefully analyzed rather than a traditional exercise in illuminating engineering so that people can see. Light pollution issues contribute to outdoor lighting’s complexity, but light pollution also calls attention to important and growing outdoor lighting problems that have to be addressed and resolved.

References

1. Mission Statement. The International Dark Sky Association (IDA). www.darksky.org. Accessed 13 December 2014
2. Model Lighting Ordinance (MLO): A joint effort of the International Dark Sky Association and the Illuminating Engineering Society of North America. www.darksky.org or www.ies.org (2011). Accessed 13 December 2014
3. International Commission on Illumination: Technical Report – Guide on the Limitation of the Effects of

- Obtrusive Light from Outdoor Lighting Installations. CIE 150. CIE Publications. <http://www.cie.co.at/index.php/Publications> (2003)
4. International Commission on Illumination: Technical Report – Guidelines for Minimizing Sky Glow. CIE 126. CIE Publications. <http://www.cie.co.at/index.php/Publications> (1997)
5. Mizon, B.: *Light Pollution: Responses and Remedies*. Springer, New York (2012, 2nd. Edition) ISBN-13 978-1461438212
6. Rich, C., Longcore, T. (eds.): *Ecological Consequences of Artificial Night Lighting*. Island Press, Washington, DC (2005)
7. The Royal Commission on Environmental Pollution: *Artificial light in the environment*. The Stationery Office, United Kingdom (2009)
8. Chicago Audubon Society: *Lights out after 11:00 p.m. During migration*. <http://www.chicagoaudubon.org/lightsout.shtml> Accessed 20 June 2012
9. National Geographic News: *Saving sea turtles with a lights out policy in Florida*. http://news.nationalgeographic.com/news/2003/03/0310_030310_turtlelight.html Accessed 20 June 2012
10. International Dark Sky Association: *Visibility, environmental and astronomical issues associated with Blue-Rich White Outdoor Lighting*, Tucson, IDA, 2010. Available at: <http://darksky.org/assets/documents/Reports/IDA-Blue-Rich-Light-White-Paper.pdf>
11. Wagner, R.: *Light at night, human health – references with abstracts*. http://www.trianglealumni.org/mcrol/References-With_Abstracts.pdf (2011). Accessed 12 Jun 2012
12. American Medical Association House of Delegates: Resolution: 516 (A-09) http://www.eficienciaenergetica.gub.uy/con_luminica/American%20Medical%20Association%20-%20Resolution%20516.pdf (2009). Accessed 10 June 2012
13. American Medical Association: *AMA Adopts New Policies at Annual Meeting*. <http://www.ama-assn.org/ama/pub/news/news/2012-06-19-ama-adopts-new-policies.page> (2012). Accessed 2 July 2012
14. Bogard, P.: *The End of Night: Searching for Natural Darkness in an Age of Artificial Light*. Little, Brown and Co. New York. First Edition July, 2013

Light Sensors

► Photodetector

Light Stimulus

► Light, Electromagnetic Spectrum

Light, Electromagnetic Spectrum

Joanne Zwinkels

National Research Council Canada, Ottawa, ON,
Canada

Synonyms

Light – **Light stimulus**, **Optical radiation**, **Electromagnetic radiation**; Electromagnetic spectrum – **Optical spectrum**

Definitions

Light is so much a part of our everyday existence contributing to our quality of life, as well as a key enabler in photonics, solar power, and new lighting technologies, that the year 2015 has been declared by the UNESCO as the International Year of Light and Lighting Technologies.

Light has many different meanings depending upon the application.

When it is used to broadly describe *optical radiation*, it is defined as electromagnetic radiation with wavelengths between approximately 10 nm and 1 mm, i.e., a term that is used to describe the ultraviolet, visible, and infrared regions of the electromagnetic spectrum.

When it is used to describe a *light stimulus*, the International Lighting Vocabulary [1] gives the following two definitions: (1) It is a characteristic of all sensations and perceptions that is specific to vision, and (2) it is radiation that is considered from the point of view of its ability to excite the human visual system, i.e., it is limited to electromagnetic radiation with wavelengths in the visible spectral region between approximately 380 and 780 nm.

The term *light* is also sometimes used as a synonym for *electromagnetic radiation* which encompasses not only the ultraviolet, visible, and infrared range but also the X-ray and gamma-ray ranges and the radio range.

Electromagnetic radiation is the emission or transfer of energy in the form of electromagnetic waves and in the form of photons.

Photon is a quantum of electromagnetic radiation that is usually associated with radiation that is characterized by one wavelength or frequency (monochromatic radiation).

Electromagnetic wave is a transverse oscillation of inextricably linked electric and magnetic fields traveling through space; through empty space, it travels at the speed of light in vacuum; through other media, its speed is reduced by the value of the refractive index of the medium.

Electromagnetic spectrum is defined as the graphical representation of electromagnetic waves arranged according to their wavelength.

Overview

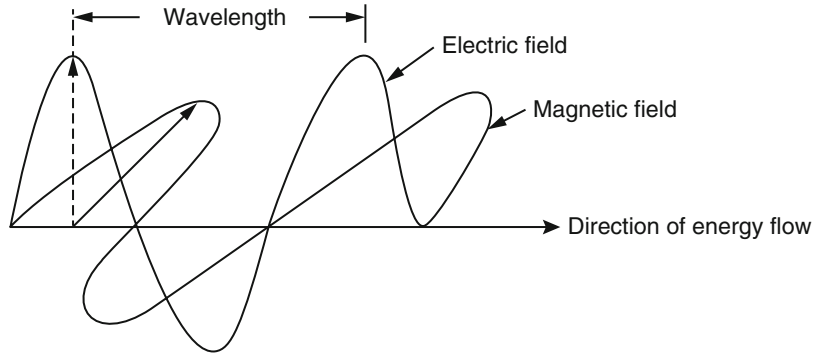
Early Views of Light

The nature of light has intrigued philosophers and scientists for thousands of years. Because light travels in straight lines (*rectilinear propagation*), it was long believed that it was made of particles (corpuscles) that emanated from the source. However, some properties of light, such as the fussiness produced when light passes around obstacles or through openings (*diffraction*), needed a different model of light where it behaved like a type of wave or propagating disturbance. In the early 1700s, both the particle and wave models of light prevailed with Sir Isaac Newton being a staunch supporter of the particle theory and others, such as the Dutch physicist, Christiaan Huygens, being proponents of the wave theory [2]. But, in the early 1800s, when the English physician Thomas Young showed that light diffracted and interfered to produce fringes when it passed through a double slit, the wave model of light became the accepted theory. At the end of the nineteenth century, the particle theory of light was revived by Einstein, who used particle behavior and a new type of theory – quantum physics – to explain the photoelectric effect. It is now widely accepted that light exhibits all three of these behaviors: particle, wave, and quantum, which are used to explain all known properties of light [2].

The history of light has also included interesting debates on how it is propagated. Ancient

Light, Electromagnetic Spectrum,

Fig. 1 Simplified picture of an electromagnetic wave. The oscillations are perpendicular to each other and to the direction of energy flow



astronomers thought that light traveled at infinite speed. The Greek philosopher Aristotle proposed that light propagated through a hypothetical substance, luminiferous ether or light-bearing ether, because it was believed that all types of waves required a medium for propagation such as sound waves in air and water waves in water. However, light is very different from other types of waves. In 1881, in one of the most famous experiments with light, Michelson and Morley tried to measure the rotational motion of the earth in this luminiferous ether [2]. However, they found that the light did not travel at different speeds for different paths. This failed experiment indirectly showed that light does not need a medium and can travel in free space (vacuum) and at constant speed. Michelson also carried out experiments to measure the speed of light by improving on Galileo's echo technique by replacing men carrying lanterns with a series of mirrors and a telescope. By adjusting the rotation speed of the mirrors, he could observe the light in the telescope, and from the rotation of the mirror and the known distance traveled by the light, he was able to determine the speed of light to about six significant figures [2]. Presently, the speed of light is the most accurately known fundamental constant of nature.

Description of Electromagnetic Radiation and Light

Electromagnetic radiation (emr) can also be described as the emission or transfer of energy from vibrating charged particles that creates a

disturbance in the form of oscillating electric and magnetic fields (Fig. 1). The British physicist James Clerk Maxwell [3] was the first to discover that these electric and magnetic waves always occurred together, i.e., were coupled, giving rise to the name, *electromagnetic waves*. It was also found that the lines of force of these electric and magnetic waves were always perpendicular to each other and perpendicular to the direction of wave propagation, i.e., *transverse*, never longitudinal. This *transverse* nature of light gives rise to various phenomena associated with *polarization*.

Polarized electromagnetic radiation is defined as radiation whose electromagnetic field, which is transversal, is oriented in defined directions [1, 4] where this polarization can be linear, elliptic, or circular. Light is totally linearly polarized (or *plane polarized*) if the electric field vectors are all oriented in the same plane, parallel to a fixed direction which is referred to as the *polarization direction*. Light is *unpolarized* when its electric field vectors vibrate randomly in all directions; this can also be considered to be equal amounts of plane-polarized radiation whose vibration directions are perpendicular to each another. The *state of polarization* of the radiation (linear, circular, elliptic) is described by the phase relationship between these two orthogonal components. If this phase difference is zero or 180° , the radiation is *linearly polarized*; if the phase difference is 90° or 270° and both components have the same amplitude, the radiation is *circularly polarized*; and if the phase difference is not $0^\circ, 180^\circ, 90^\circ$, or 270° and/or the amplitudes are different, the radiation is *elliptically polarized*.

Light is *partially polarized* if the electric vectors have a preferred direction. If the electric vector is polarized vertical to the plane of incidence, this is referred to as *s-polarized* (or TE), and if the electric vector is polarized parallel to the plane of incidence, this is referred to as *p-polarized* or (TM).

In addition to its state of polarization, these *electromagnetic waves* are characterized by several other measurable properties, such as the number of oscillations per second, the separation between successive peaks or troughs, and the size or amount of oscillations. These quantities are described by the frequency, ν ; wavelength, λ ; and amplitude or intensity, I , respectively. The *intensity* describes the amount of energy flowing in the electromagnetic wave and is proportional to the square of the *amplitude*. It is also necessary to specify the *direction* that the wave is traveling, which is indicated in Fig. 1 by using an arrow. It can be seen that these waves crest up and down in a sinusoidal fashion, which is described as the *wave front* moving up and down. When more than one wave is traveling, it is important to describe how they differ in phase. Two waves are *in-phase*, if their crests coincide and their troughs coincide. Two waves are *out-of-phase*, if the crest of one wave coincides with the trough of the second wave and vice versa. A *plane wave* is a wave whose surfaces of constant phase are infinite parallel lines transverse to the direction of motion. If all of the waves vibrate in-phase, the emr is referred to as *coherent*. The laser is an example of a coherent light source. If there is no fixed phase pattern between the waves emitted by a light source, the emr is referred to as *incoherent*. Most ordinary light sources, such as an incandescent lamp, are examples of incoherent light sources.

Properties of Electromagnetic Waves

As mentioned above, generally a wave, such as a water wave, is a propagating disturbance of some equilibrium state in a continuous medium. However, no medium is required for propagation of electromagnetic waves. This property of electromagnetic waves makes them unique from other types of waves that can be considered mechanical

in nature and was first demonstrated by Einstein in 1905. He also proposed the idea that light could behave both as a wave and as a particle in order to explain the spectral distribution of the radiation emitted from a hot (incandescing) object. This type of electromagnetic radiation is referred to as *blackbody radiation* (give link to encyclopedia article on blackbody radiation).

Electromagnetic waves of all wavelengths travel exactly at the same speed in vacuum. This fundamental constant of nature is known as the *speed of light in vacuum*, denoted by the symbol, c . It has a value of 299,792,458-metre per second. The frequency, ν , is related to wavelength, λ , by $\nu = c\lambda^{-1}$. Waves are also sometimes described by their *wave number*, σ , which is the inverse of wavelength, defined by $\sigma = \lambda^{-1}$.

When an electromagnetic wave travels through other media such as air, glass, or water, its frequency remains constant, but its wavelength and speed are both reduced by a factor, n , the *refractive index* of the material. For commonly used optical glasses, the value of n ranges from 1.52 to 1.72 and is a function of frequency. For standard air, n has a value of 1.00028 [5] and can be ignored for most practical applications.

The behavior of electromagnetic waves at the boundary from one medium to another giving rise to reflection, transmission, and refraction phenomena is described by a set of wave equations known as *Maxwell's equations* [6].

These four equations have become the fundamental laws in electromagnetics connecting the principles of electricity and magnetism. Solving these equations reveals the wave equation and the form of the electromagnetic waves. The Maxwell's equations are not given here because this is outside the scope of this entry; the interested reader can find a good description in the following references [2, 6].

The smallest unit or quantity of electromagnetic radiation is called a *photon*. The energy of a photon is given by $E = h\nu$ where h is Planck's constant (6.62606×10^{-34} J s). Light can then be described as a stream of individual photons, each with a definite energy and that can interfere with each other like waves and diffract around corners [3]. The motion of these photons is controlled by these same set of Maxwell wave equations [2, 3].

The *electromagnetic spectrum* spans the total range of wavelengths of electromagnetic radiation from the shortest to the longest wavelength that can be generated physically. This range of wavelengths spans practically from zero to near infinity and can be broadly divided into regions as shown in Table 1 [7], which includes radio waves, infrared, visible, ultraviolet, X-rays, and gamma rays. This division is not exact since there is a gradual transition from one region to the next, which is shown schematically in Fig. 2 where the visible and ultraviolet regions are highlighted. This classification of light into regions is also shown as a function of frequency. It is important to note that the only difference between electromagnetic

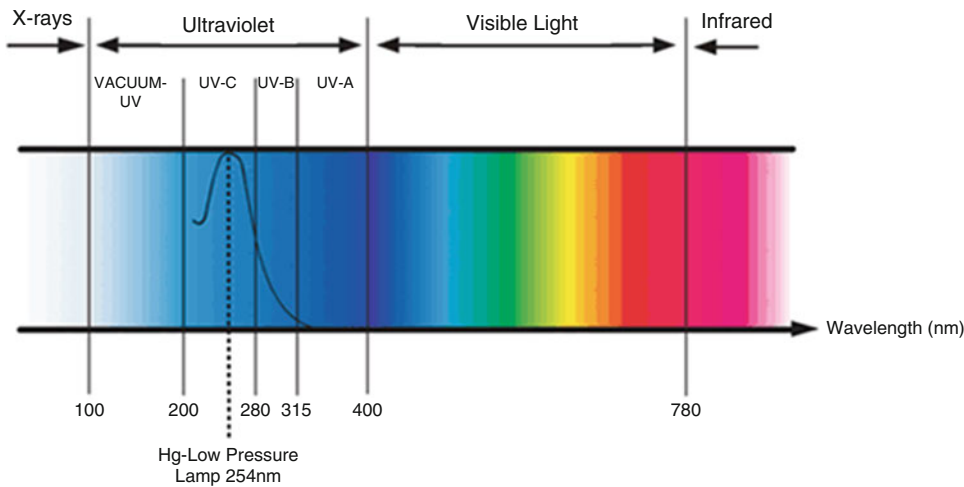
radiations in all these regions is its wavelength (and frequency). It has different descriptions because of the relationship between its frequencies and those that are excited in the various materials that the electromagnetic radiation can interact with. For example, the visual receptors in the human eye are only sensitive to electromagnetic radiation in a very narrow frequency range from 10^{14} to 10^{15} s^{-1} (or 400–700 nm), whereas X-rays are used to excite features in the body that are of the size of an atom (0.1 nm). These different applications of emr are described in more detail below.

Interaction of Light with Matter

Light, Electromagnetic Spectrum, Table 1 Regions of the electromagnetic spectrum

Wavelength range (nm)	Frequency range (s^{-1})	Description
<0.1 nm	$10^{20}\text{--}10^{23}$	Gamma rays
0.1–10 nm	$10^{17}\text{--}10^{20}$	X-rays
10–400 nm	$10^{15}\text{--}10^{17}$	Ultraviolet
400–700 nm	$10^{14}\text{--}10^{15}$	Visible
700 nm to 1 mm	$10^{11}\text{--}10^{14}$	Infrared
1 mm to 1 cm	$10^{10}\text{--}10^{11}$	Microwaves
1 cm to 100 km	$10^3\text{--}10^{10}$	Radio waves
100–1,000 km	$10^2\text{--}10^3$	Audio frequency

When light interacts with matter, different phenomena can occur depending upon the relationship of the wavelength (frequency) of the light with the physical size (resonant frequencies) of the interfering matter. This matter consists of atoms, ions, and molecules. As mentioned above, only light in the visible portion of the spectrum with frequencies from 10^{14} to 10^{15} s^{-1} can stimulate the visual receptors in the human eye. Light can also exhibit wave or particle properties depending upon this relationship between size (or frequency). For light waves that are closely spaced in relation to the spacing of the



Light, Electromagnetic Spectrum, Fig. 2 The regions of the electromagnetic spectrum, highlighting the optical spectrum which includes the visible and ultraviolet regions.

interfering matter, for example, a wave front passing through a slit or an opaque edge, secondary wave fronts are generated. These will interfere with the primary wave front, as well as with each other, and produce *diffraction* patterns. Since these secondary wave fronts were produced from the same primary wave, their phase will change in step. Thus, even for a normally incoherent light source, wave-like behavior such as *diffraction* and *interference* can be seen under certain conditions [8].

Many colorful effects are produced by this combination of light *diffraction* and *interference*. In nature, this can be seen, for example, in the beautiful array of colors displayed by mother of pearl (*opalescence*) or by thin film *iridescence* (add link to iridescence article). For light waves that are very small in relation to the interfering particles, such as the interaction of visible light with dust particles or gas molecules in the atmosphere, *Rayleigh scattering* occurs which produces the blue color of the sky (add link to Rayleigh and Mie scattering article).

However, when light interacts with objects that are large in relationship to their wavelength, such as a brick wall or pane of glass, the light behaves more like a ray (particle) and the laws of geometrical optics can be applied.

Depending upon the transparency or translucency of the material and its surface quality, the light will be reflected, refracted, transmitted, absorbed, or scattered. The type of light interaction that dominates depends upon the precise nature of the matter, e.g., for a smooth opaque colorless solid, specular (mirror-like reflection) will dominate, whereas for an opaque colored material, diffuse reflection or scattering will dominate. The selective modification of the energy distribution of the incident light by the diffusely reflecting material gives rise to different colors.

Light Intensity Distribution and Color

The different wavelengths in the visible spectrum can also directly stimulate different colors in the human visual system. This dependence of color on wavelength is shown in Table 2.

Light, Electromagnetic Spectrum, Table 2 Dependence of color on wavelength

Wavelength range (nm)	Color
400–430 nm	Violet
430–480 nm	Blue
480–560 nm	Green
560–590 nm	Yellow
590–620 nm	Orange
620–700 nm	Red

Radiation of a single wavelength is called *monochromatic*. Except for lasers and certain specialized lamps, most sources of optical radiation emit energy over a broad wavelength region. The curve describing the power at each wavelength is called the *spectral power distribution* (SPD). A color stimulus generally has an SPD that varies with wavelength across the visible spectrum producing a color or sensation that is a shade or mixture of the colors listed in Table 2. Stimuli containing all the visible wavelengths in roughly equal proportions appear white.

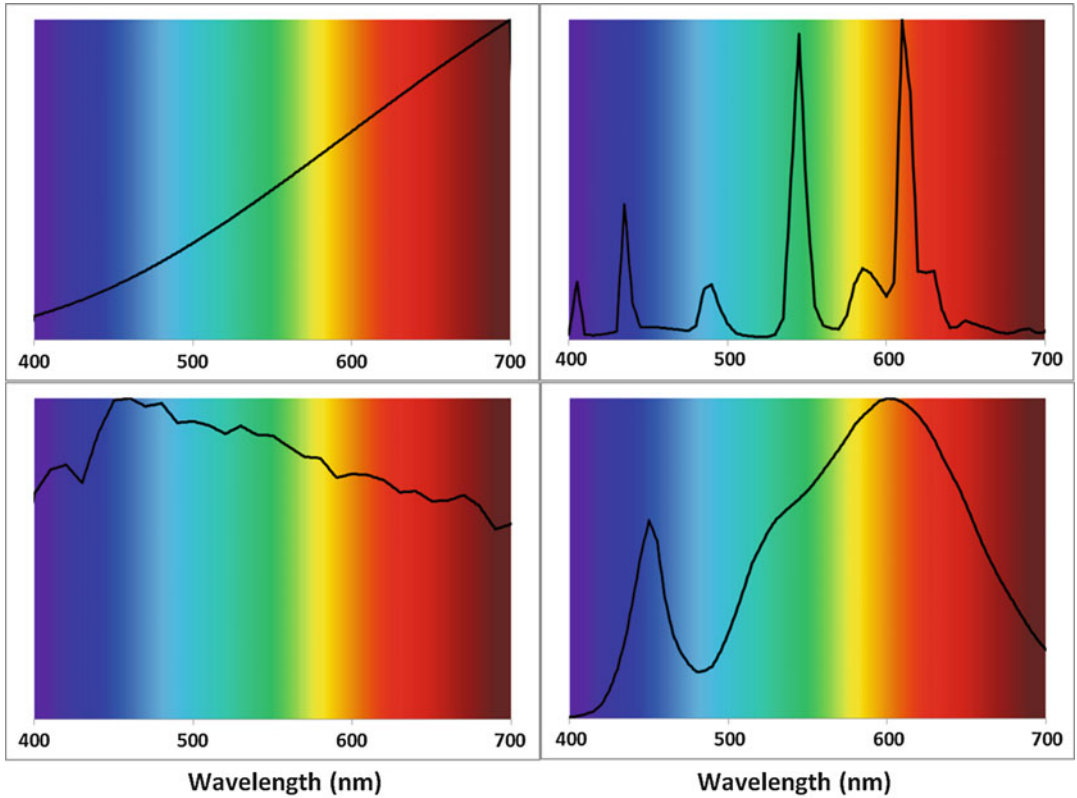
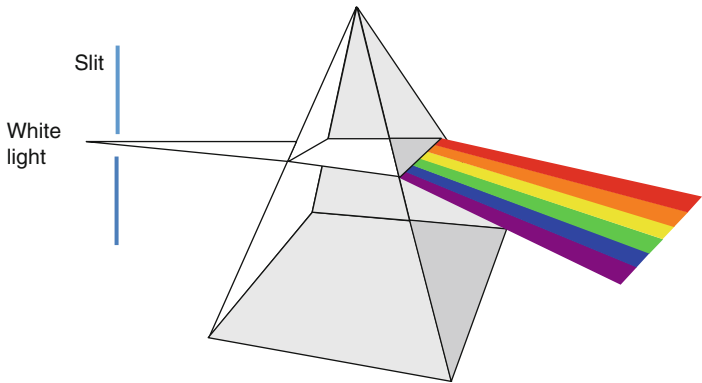
This is demonstrated by passing white light through a glass prism, which spreads the light out into a spectrum of colors (see Fig. 3).

However, white light can also be produced by superimposing discrete monochromatic lights. For example, white light can be produced by superimposing red, blue, and green lights in certain proportions. This was first demonstrated by Thomas Young [3] who used this experimental finding to theorize that the human eye contained three different types of color receptors, which is known as the trichromatic theory of human color vision.

Because of this trichromacy, white light can be produced by many different SPDs. The variation in these SPDs is illustrated in Fig. 4, which shows the SPD of a typical tungsten lamp, a fluorescent lamp, a white LED source produced by a blue LED and yellow phosphor, and noon daylight. The most prevalent source of light in our natural environment - the sun, has an SPD that can change dramatically during the course of the day, from a bluish cast at noon to a reddish-orange cast at sunset. These variations in SPD of the light source are very important for visual evaluations of



Light, Electromagnetic Spectrum, Fig. 3 Use of a prism to separate white lights into a spectrum of colors



Light, Electromagnetic Spectrum, Fig. 4 Relative spectral power distributions (SPDs) of four different sources of white light; CIE illuminant (if applicable)

given in parentheses: *Top Left*: tungsten lamp (CIE A); *Bottom Left*: noon daylight (CIE D65); *Top Right*: fluorescent lamp (FL12); blue LED + yellow phosphor

colored goods. For this reason, the SPDs of different phases of daylight have been standardized by the International Commission on Illumination (CIE) as a series of D illuminants. Shown in Fig. 4 is the SPD of CIE standard illuminant D65, which is average daylight with an approximate correlated color temperature of 6500 K.

Applications of Electromagnetic Radiation

The applications of electromagnetic radiation (emr) cover almost all aspects of our daily lives. The importance of visible light for human vision has already been discussed. Light emanating from

the sun spans the solar wavelength region from about 250 nm in the UV to 2,500 nm in the near-infrared region. This solar radiation is extremely important to the energy balance of the earth and for various building design applications, such as the design of glazing materials. It is also exploited in light-based technologies, such as photovoltaics.

Light in the infrared region of the spectrum is felt as heat and is used in the design of radiant heating, solar panels, and collectors. The frequencies in the infrared region are close to the resonant vibrational frequencies in molecules and can be used for molecular identification by spectroscopic means. Similarly, the frequencies in the microwave region are close to the resonant rotational frequencies in molecules, in particular the water molecule. Since most foods contain a significant amount of water, the microwave radiation heats the food by the increased rotational motion of the interspersed water molecules. These microwaves pass through glass, ceramic, and plastic but reflect from metal which can also be used to advantage in the selection of the food container and walls of the microwave, respectively.

Light in the ultraviolet (UV) region has many important uses. In the early nineteenth century, it was called “chemical rays” because it was found that UV radiation could cause chemical changes. This was used to advantage to harden special glues in inks, coatings, and adhesives more quickly, in a process called UV curing through polymerization. The skin of our bodies needs exposure to UV with wavelengths in the region 280–315 nm (UVB) in order to stimulate the production of Vitamin D. However, too much exposure to UVB can cause skin cancers, so it is important to find the optimum exposure level. UV light with wavelengths in the region 200–280 nm (UVC) is called “germicidal UV” because it has sufficient energy to inactivate bacteria and viruses by disabling their DNA strands. For this reason, it is also used for sterilization of surfaces such as medical equipment. This type of UVC radiation is often produced by a low-pressure mercury vapour lamp (see Fig. 2). UV radiation is also needed to excite a class of molecules referred to as fluorescent-whitening agents (FWAs) or optical brighteners in manufactured white goods, such

as paper, fabrics, detergents, soaps, and cosmetics, which emit in the blue portion of the visible spectrum and produce an enhanced whiteness effect. UV radiation is also of the appropriate frequencies to absorb ozone molecules in the atmosphere so very little UV radiation reaches the earth’s surface.

X-rays have a higher frequency than UV radiation and can pass through the skin and soft tissue, but they do not easily pass through the bone or metal. This type of emr is used to produce photographs of bones in medical diagnostics to check for damage such as fractures. The frequencies of gamma radiation, on the other hand, are more penetrating and can destroy chemical bonds by interacting with the electrons of the constituent atoms or disrupt DNA bonds, resulting in prevention of cellular division. This can be used to advantage to kill microorganisms throughout the material. Gamma sterilization is used, for example, in sterilization of water and food products.

Other important applications of emr are radio, television, and electric current.

Cross-References

- [Blackbody and Blackbody Radiation](#)
- [Iridescence \(Goniochromism\)](#)
- [Rayleigh and Mie Scattering](#)
- [Spectral Power Distribution](#)

References

1. CIE S 017/E: ILV – International Lighting Vocabulary. CIE Central Bureau, Vienna (2011)
2. Jenkins, F.A., White, H.E.: *Fundamentals of Optics*, 4th edn. McGraw-Hill, New York (1976)
3. Overheim, R.D., Wagner, D.L.: *Light and Color*. Wiley, New York (1982)
4. Shurcliff, W.E.: *Polarized Light: Production and Use*. Harvard University Press, Cambridge, MA (1962)
5. Ciddor, P.E.: Refractive index of air: new equations for the visible and near infrared. *Appl. Optics* **35**, 1566–1573 (1996)
6. Stern, F.: In: Seitz F., Turnbull D. (eds.): *Elementary theory of the optical properties of solids*. *Solid State Physics*, vol. 15. Academic, New York, pp 301–304 (1963)
7. Williamson, S.J., Cummins, H.: *Light and Color in Nature and Art*. Wiley, New York (1983)
8. Falk, D., Brill, D., Stork, D.: *Seeing the Light: Optics in Nature, Photography, Color, Vision, and Holography*. Harper & Rowe Publishers, New York (1986)

Light-Emitting Diode, LED

Wout van Bommel
Nuenen, The Netherlands

Synonyms

Solid-state light source; Optoelectronic light source

Definition

Light source that produces light as a result of recombination of positively charged holes and negatively charged electrons at the junction of inorganic, solid-state, p- and n-type semiconductor material.

Solid-State LED Light Sources

LEDs are solid-state radiators where the light is created inside solid-state material [2–4]. Common semiconductor diode chips, used today in so many electrical circuits, all use much the same technology. The light-radiating diode versions are called light-emitting diodes or LEDs. The name LED is commonly used for light-emitting diodes made of inorganic semiconductor material; they are point light sources. Light-emitting diodes made of organic semiconductor material are referred to as OLEDs; they are planar light sources. A separate chapter deals with OLEDs.

Until the mid-1990s of the last century, LEDs had a low lumen output and low efficiency, making them only suitable as small indicator lamps (e.g., in electrical appliances). Today, the efficacy of LEDs is comparable to that of gas discharge lamps. The lumen output of a single LED can be more than 1,000 lm. To distinguish these LEDs

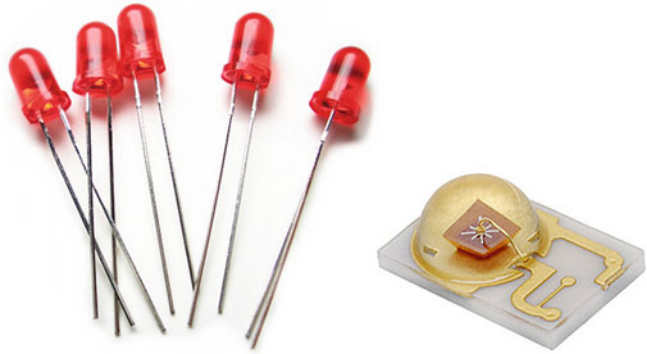
from the indicator type of LEDs, they are referred to as high-brightness or high-power LEDs (Fig. 1). Further improvements in high-brightness LEDs are expected, ultimately leading to efficacies of probably slightly more than 200 lm/W (for white-light LEDs). This is approximately twice the efficacy of today's most efficient white-light gas discharge lamps. With its light-emitting surface of some 0.5–5 mm², an individual LED chip represents the smallest artificial light source. LEDs have a long lifetime and are, given the solid-state material, extremely sturdy. They are available in white and in colored-light versions. The colored versions are extensively used in traffic signs. Colored versions were also the first ones to be used on a large scale for lighting: specifically, the exterior floodlighting of buildings and monuments. Both the efficacy and the color quality of white LEDs have been improved so much that they are now used in many different lighting applications, including road lighting, indoor accent lighting, domestic lighting, and automobile lighting. Examples of the use of LEDs for office lighting and outdoor sports lighting can also be found. Given the potential for a further increase in their efficacy, the number of LED applications is bound to increase. Their small size and their availability in many different colors and the easiness of lighting control, both in terms of dimming and color changing, are properties that permit of completely new applications.

Working Principle

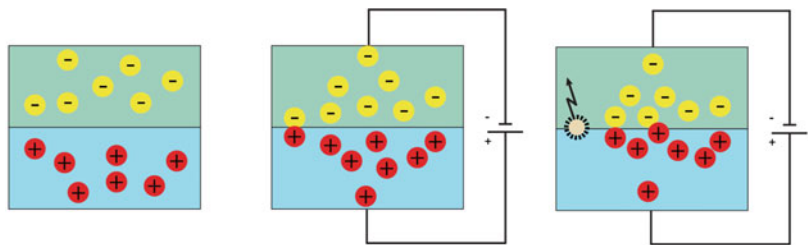
Principle of Solid-State Radiation

Like any diode, an LED consists of layers of p-type and n-type of semiconductor material. The n-type of material has an excess of negatively charged electrons, whereas the p-type material has a deficiency of electrons, viz., positively charged holes. Applying a voltage across the p-n semiconductor layer pushes the n- and p-type atoms towards the junction of the two materials (Fig. 2). Here the n-type of atoms “donate” their excess electrons to a p-type of atom that is deficient in electrons. This process is called recombination. In doing so, the electrons move from a high level of energy to a lower one, the energy

Light-Emitting Diode, LED, Fig. 1 Indicator lamp LEDs (*left*) and a high-power LED (*right*)



Light-Emitting Diode, LED, Fig. 2 Principle of operation of solid-state radiators



difference being emitted as light. The wavelength of the light is dependent on the energy-level difference between the p and n materials, which in turn depends on the semiconductor material used: different semiconductor materials emit different wavelengths, and thus different colors, of light. The process is very much like the process of excited electrons in a gas discharge falling back to their original orbit with a lower energy level while emitting light. In this sense it is surprising that the expression “solid-state discharge” and “solid-state discharge lamp” is hardly ever used. Not all recombinations result in light emission. Some recombinations are non-radiative and just heat up the solid material. This limits the efficiency of light creation. A further limitation of the efficiency is caused by absorption of light in the solid material of the chip itself. Improvements in radiative-recombination efficiency and light-extraction efficiency have been the most important reasons for the dramatic improvements of efficacy of LEDs during the past decade. Further improvements will be sure to further greatly increase the efficacy of LEDs over the coming decade.

Like gas discharge lamps, solid-state lamps cannot function when they are operated direct from the mains supply voltage. Solid-state light sources are low-voltage rectifiers that allow current to pass in one direction only. This means that the AC mains supply has to be transformed to low voltage and then rectified into a DC supply. Small fluctuations in supply voltage cause large variations in current that can damage the light source. High-power LEDs need an electronic driver to obtain a constant current characteristic.

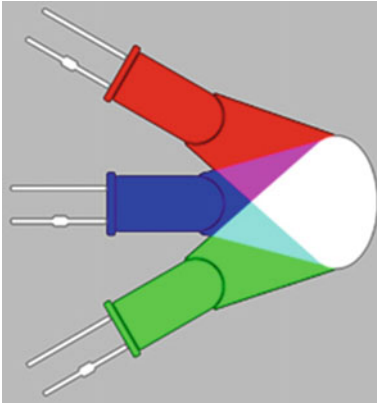
Principle of White LEDs

The spectrum of a single LED is always narrow. Consequently, its light is colored. White LED light can nevertheless be obtained by combining three (or more) differently colored LED chips. A common method is to combine red, green, and blue LED chips into a single module to produce white light (RGB LED, Fig. 3). Sometimes the three chips are mounted in the same LED package (Fig. 4). The color rendering of an “RGB white-light” system is not good, since large areas of the full color spectrum are not included in its light. Sometimes an amber color chip is added to the

RGB combination to improve the color quality of the light (RGBA LED). Research is going on to produce single, multilayer LED chips, each layer producing a specific color of light. A single LED producing red, green, and blue light would therefore result in white light.

Good-quality white light, which is especially important when it comes to providing good color rendering, is obtained by using a blue LED chip in combination with fluorescent material that

converts much of the blue light into light of different wavelengths spread over almost the whole visible spectrum. In LED technology it is customary to call such fluorescent materials, phosphors: hence white LEDs based on this principle are called “white-phosphor LEDs.” By mixing different phosphors in different proportions, white LEDs producing different tints of white light with different color-rendering capabilities can be made.

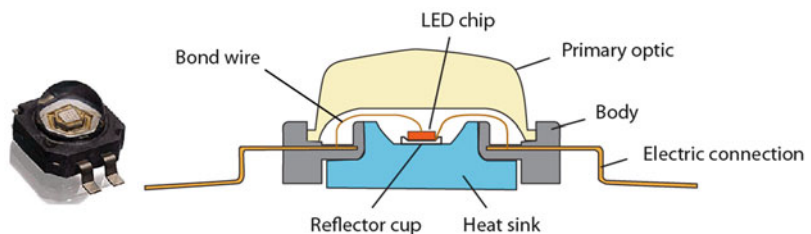


Light-Emitting Diode, LED, Fig. 3 White light by combining red, green, and blue LED light [1]



Light-Emitting Diode, LED, Fig. 4 RGB chips mounted in a same LED package

Light-Emitting Diode, LED, Fig. 5 The main components of a high-brightness LED [1]




Materials and Construction

The LED chip is embedded in a larger structure for mechanical protection, for the electrical connections, for thermal management, and for efficient light out-coupling. The main parts of an LED are (Fig. 5):

- Semiconductor LED chip
- Reflector cup
- Supporting body
- Electrodes and bond wires
- Heat sink
- Primary optics
- Phosphors (for white LEDs)

Semiconductor Chip Material

The p-n semiconductor sandwich forms the heart of the LED and is called the LED chip or die. The semiconductor material used determines the wavelength and thus the color of the light emitted. For LEDs, compound semiconductor material is used, which is composed of different crystalline solids. These are doped with very small quantities of other elements (impurities) to give their typical n and p properties. For the colors blue, green, and cyan, the elements indium, gallium, and nitride (InGaN) are used in different compositions (Fig. 6, left). The elements aluminum, indium,

InGaN Colours					AlInGaP Colours			
								
450 nm Blue	498-500 nm Green-Blue	505 nm Blue-Green	525 nm Green		590 nm Amber	605 nm Orange	615 nm Red-Orange	626 nm Red

Light-Emitting Diode, LED, Fig. 6 The main semiconductor material elements used in high-brightness LEDs, with examples of the corresponding light colors [1]

gallium, and phosphide (AlInGaP) are used to produce the colors amber, orange, and red (Fig. 6, right). It was the Japanese Nakamura who, in 1993, succeeded for the first time in producing a blue LED suitable for mass production. This was the “missing link” that enabled the production of white-phosphor LEDs and the production of white LED light on the basis of mixing the light of red, green, and blue LED-S. As can be seen from Fig. 6, today a very small area of the spectrum, in greenish yellow, is still missing.

Shape of the Chip

Unfortunately, the LED chip is a “photon or light trap”: that is to say, much of the light emitted within the chip is internally reflected by its surfaces (borders between the material and air) and, ultimately, after multi-reflections, absorbed in the material (heating up the material). Only light that hits the outer surface more or less perpendicularly (approximately 20°) can leave the material. By giving the chip a specific shape and by keeping it thin, the so-called light-extraction efficiency can be improved. Figure 7 gives an example of such a specifically shaped chip.

Reflector Cup

In many cases, the chip is placed in a reflector cup which, because of its shape, helps to direct the light in an upwards direction. Highly reflective material is used: for example, metal or ceramic material.



Light-Emitting Diode, LED, Fig. 7 An example of a specifically shaped LED chip that improves light-extraction efficiency. On top of the chip, the anode with its bond wire can be seen [1]

Primary Optics

The silicon lens on top of the LED chip serves as protection for the chip. More importantly, it helps in increasing the light extraction from the chip and as such is essential for a high lumen efficacy of the LED. This is because by introducing a lens medium between the chip and the air with a refractive index value between that of air and that of the chip material, the angle over which light can escape from the chip increases.

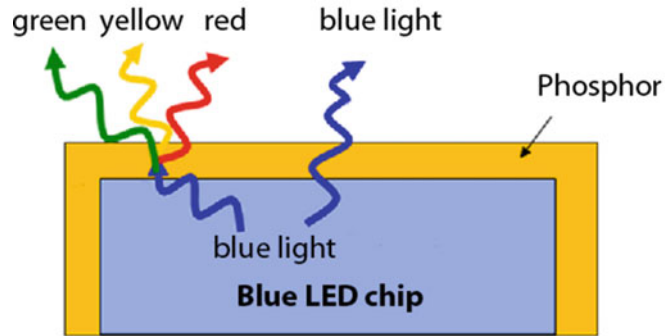
Electrodes and Bond Wires

In order to be able to apply power to the chip, the p and n parts of the chip have metal contacts called electrodes. Bond wires connect the electrodes with the electrical connections. They are usually gold wires. Since the electrodes intercept light leaving the chip, the dimensioning of the

Light-Emitting Diode, LED, Fig. 8 Examples of luminaires with cooling fins



Light-Emitting Diode, LED, Fig. 9 Principle of creating white light with a blue-light chip covered with phosphor



electrodes and bond wires, especially on the side of the main light-escape route, is one of the factors that determines the light efficiency of the LED.

Heat Sink

LEDs do not radiate infrared radiation and consequently give a cool beam of light. However, this does not mean that they do not generate heat. Non-radiative recombinations of electrons and holes in the p-n sandwich, and light trapped in the chip, do heat up the chip. The larger the power of the chip and the lower its luminous efficacy, the higher is this heating effect. The higher the temperature of the p-n junction in the chip, the lower the light output of the chip. A too-high chip temperature also seriously shortens LED life, and it also slightly shifts the emitted wavelength and thus the color of the LED. Effective thermal management is therefore critical for a proper functioning of LEDs. All high-power high-brightness LEDs therefore have a heat sink of high-thermal-conductivity material (like aluminum or copper) on their rear side to conduct the heat away from the chip towards the outside world through the luminaire housing. LED luminaires must therefore incorporate in their design thermal conduction and convection features (such as cooling fins

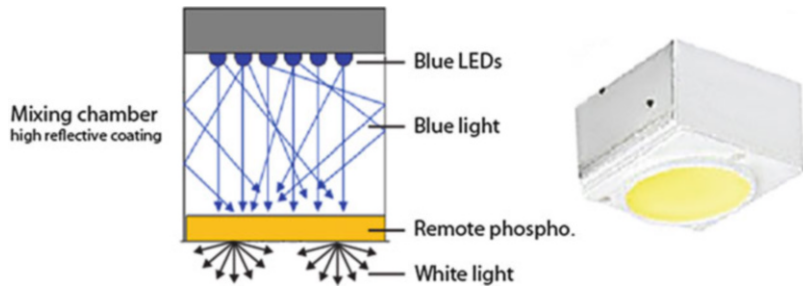
Fig. 8) to dissipate the heat to the immediate surroundings. For retrofit LED lamps (LED bulbs), the size of the heat sink is limited by the size of the bulb. The heat sink therefore has a limited capacity, thus limiting the power of the retrofit LED bulbs.

Phosphors

As mentioned above, the most important method for producing white light with LEDs is by applying a phosphor coating to a blue-light LED that converts part of the blue light into longer-wavelength, green, yellow, and red light. Different compositions of different phosphors are used to produce white light of different color tints. Since the basis of the process is the blue light of the chip, the final efficacy will become higher as more blue is kept in the light. However, this implies a high color temperature (cool-white light) and relatively poor color rendering. If, with a different phosphor composition, more blue light is converted, the color quality will improve at the expense of a somewhat-lower efficacy.

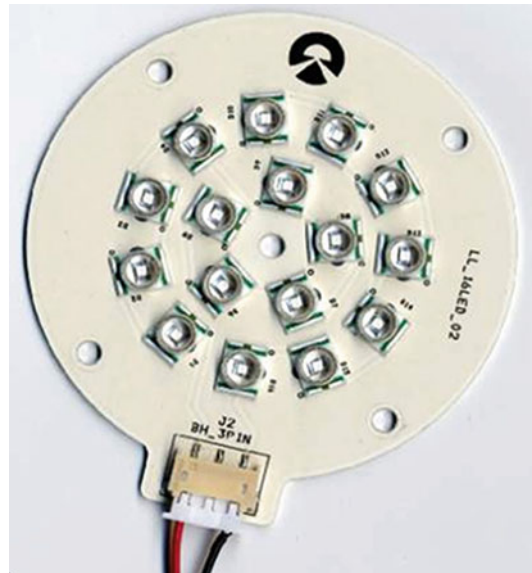
Phosphors Applied Direct to the Chip The phosphor is often applied on or very near to the blue LED chip (Fig. 9). The thickness of the layer

Light-Emitting Diode, LED, Fig. 10 Principle of a remote-phosphor LED module creating white light [1]



has to be very uniform as a variation in thickness of the phosphor layer causes a variation of the color temperature in the light beam.

Remote Phosphor In the case of multi-LED units, the phosphor is sometimes applied at a greater distance from the LEDs. Such modules are called remote-phosphor LED modules (Fig. 10). Here, a number of blue LEDs are placed inside a mixing chamber of high and diffuse reflective material. The phosphor layer, positioned remotely from the LEDs on the bottom of the chamber, converts the blue light of the chips into white light. In this way, thanks to the mixing process, small differences in light output and or color of individual chips are not visible. The risk for disturbing glare is also reduced because the light intensity from the large-sized phosphor layer is much lower than the intensities of small, individual LEDs. In normal phosphor LEDs the hot LED heats up the phosphor that is applied on the LED itself, reducing its conversion efficiency. Heating of remote positioned phosphor is much less and thus the conversion efficiency of remote-phosphor systems is higher. In normal phosphor LEDs a relatively large amount of light from the LED is reflected back from the phosphor towards the LED and absorbed there. In the remote-phosphor situation, most of the light reflected back from the phosphor layer takes part in the mixing process without reaching the LEDs, thus further increasing the final efficiency of the system. The phosphors used for the blue-light conversion appear yellow when they are not activated, that is to say, when the LED is not switched on (see Fig. 10 right). This sometimes makes people think, erroneously, that the blue



Light-Emitting Diode, LED, Fig. 11 LED cluster module with multiple LEDs mounted on a printed circuit board

light is filtered through a yellow filter. It is really wavelength conversion and not light filtering that takes place.

LED Cluster Modules

The luminous flux of one individual LED is quite low compared to that of most conventional light sources. Multiple LEDs are therefore often mounted on a printed circuit board (PCB) to obtain an LED module emitting a high luminous flux (Fig. 11). The PCB establishes the electrical connections between all components and the external electrical driver. The PCB must also conduct the heat from the heat sinks of the LEDs to the outside world.

Properties

Temperature of the Chip Junction

It has already been mentioned that with rising temperature of the p-n chip junction, the performance of LEDs decreases: particularly the light output and lifetime. The performance data are usually specified for a junction temperature (T_j) of 25 °C. However, under normal operating conditions, a junction temperature of 60–90 °C is easily obtained. Depending on LED type, the lumen output falls to 60–90 % when the junction temperature increases from 25 to 80 °C. Amber and red LEDs are the most sensitive to changes in junction temperature, and blue LEDs the least.

Binning

The mass production of LEDs results in LEDs of the same type varying in color, light output, and voltage. In order to ensure that LEDs nevertheless conform to specification, LED manufacturers use a process called binning in their production process. At the end of the manufacturing process, LED properties are measured and LEDs are subsequently sorted into subclasses or “bins” of defined properties. As far as the color quality is concerned, the tolerances in these definitions (based on MacAdam ellipses) are such that visible differences in color between LEDs from the same bin are minimized. As the definition of a bin does not change with time, the same quality is also assured from production run to production run.

With the advancement of knowledge concerning LED materials and the mass-production process, it may be expected that binning will ultimately no longer be required (“binning-free LEDs”).

Energy Balance

The energy balance of an LED is much easier to specify than that of conventional light sources. This is because no energy is radiated in the UV and infrared region of the spectrum, which means that the energy balance comprises only visible radiant energy and heat energy. Today, white LEDs transform 20–30 % of the input power into visible light and the remaining part into

heat. The light percentage comes close to that of fluorescent lamps and will soon supersede it.

System Luminous Efficacy

Sometimes, luminous efficacies are specified for the bare chip. It is evident that the ancillary devices described in the previous sections, which are essential for a proper functioning of LEDs, absorb light. The only realistic thing to do, therefore, is to specify luminous efficacy (and light output as well) for the total LED package. As with most conventional lamps, the luminous efficacy of LEDs is dependent on the power of the LED and on the color quality of the light it produces. Higher-power LEDs have higher efficacies, while those with better color rendering have lower efficacies. Today cool-white LEDs are commercially available in efficacies up to some 150 lm/W and warm-white LEDs with color-rendering indices larger than 80 in efficacies of around 130 lm/W (all lm/W values include driver losses). Retrofit LED bulbs with warm-white light (around 2,700 K) and color-rendering index better than 80 are available in efficacies up to 80 lm/W. Here, too, cooler-white versions are slightly more efficient.

As discussed at the beginning of this chapter, in coming years it may be expected to see further important improvements in luminous efficacies for white LEDs – even up to slightly more than 200 lm/W.

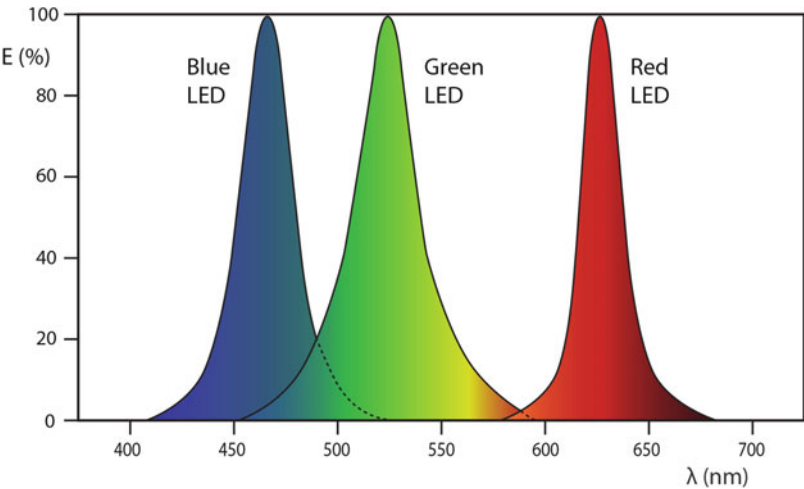
Lumen-Package Range

Today, single LEDs exist in lumen packages varying from a few lumen (indicator lamps) to more than 1,000 lm. In the latter case, severe screening is called for to restrict glare, because so much light comes from such a very small light-emitting surface. By mounting multi-LEDs on a printed circuit board, LED modules can be achieved with much larger lumen packages.

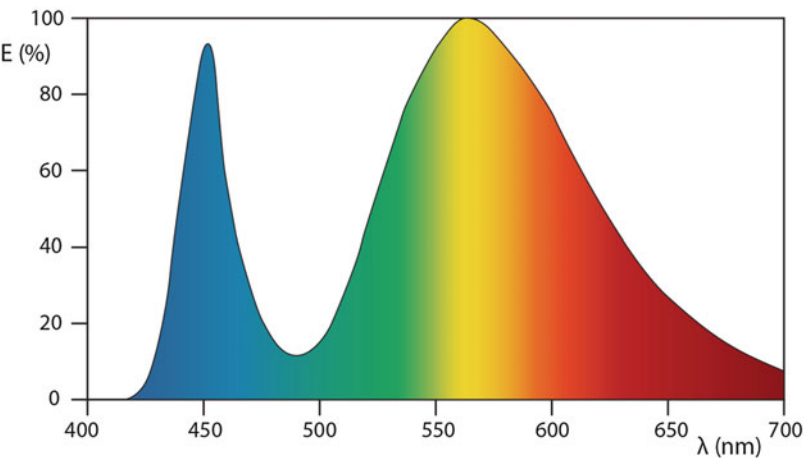
Color Characteristics

Principally, LEDs have a quasi-monochromatic, narrow-band spectrum. Figure 12 shows the spectra of blue, green, and red LEDs. The half-maximum width of the spectrum bands is smaller than ca. 50 nm. It has been already been shown

Light-Emitting Diode, LED, Fig. 12 The relative spectral power distributions of a typical blue, green, and red LED



Light-Emitting Diode, LED, Fig. 13 The relative spectral power distribution of a white-phosphor LED with color temperature $T_k = 4,000$ K and color rendering $R_a = 70$



what colors can be produced by using different semiconductor materials (see Fig. 6). With some LED modules, different colored LEDs can be controlled (dimmed) individually. With such LED modules (e.g., making use of RGB LEDs) the color of the light can be dynamically changed from white to all colors of the spectrum.

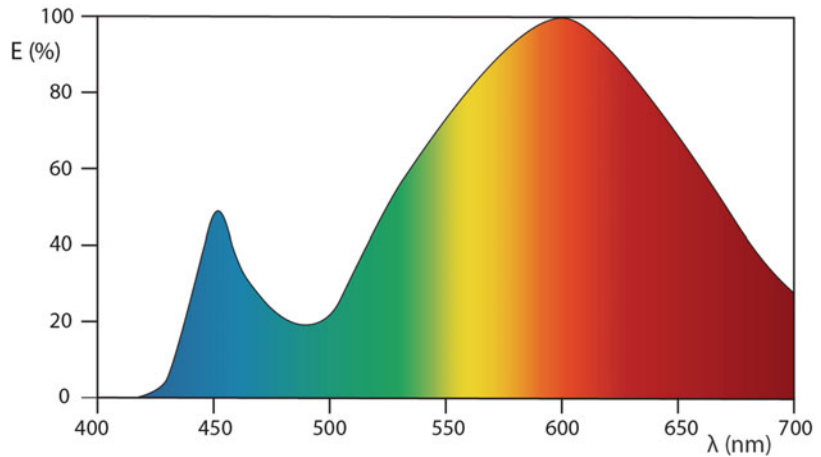
White phosphor LEDs can have a near-continuous spectrum (Figs. 13 and 14). By applying different phosphors, white light within the color temperature range of 2,700–10,000 K can be produced. Often, the higher-color-temperature versions have only moderate color rendering (R_a between 50 and 75). In the lower-color-temperature versions, LEDs are available with

good (R_a larger than 80) to excellent color rendering (R_a larger than 90 or even 95).

Beam Control

For lighting designers, one of the most interesting properties of an LED is its small light-emitting surface. This allows the creation of very accurately defined beams. As an illustration of this, Fig. 15 shows a near-parallel light beam made with an LED-line luminaire that is impossible to create with conventional light sources. The other side of the coin is that small light-emitting surfaces often need professional screening in order to limit excessive glare. Multi-LED luminaires have also multi-light

Light-Emitting Diode, LED, Fig. 14 The relative spectral power distribution of a white-phosphor LED with color temperature $T_k = 2,750\text{ K}$ and color rendering $R_a = 85$



Light-Emitting Diode, LED, Fig. 15 Near-parallel light beam with an LED-line luminaire [1]

beams that may cause multiple shadows because a lighted object is illuminated from many slightly different directions. With RGB color mixing, this may lead to disturbing, multicolored shadows.

Lifetime

In the case of high-performance LEDs, it takes a very long time before they actually fail – usually

considerably more than 50,000 h. Before that time, however, their lumen depreciation is so great that the LED is no longer giving sufficient light for most applications. Therefore, for LED lifetime specifications, the length of time that it takes to reach a certain percentage of its initial light value is used. Based on a depreciation value of 70 %, lifetime values of between 35,000 and 50,000 h are common for high-performance LEDs. LED bulbs, with their limited space for handling heat, have a lifetime of some 25,000–35,000 h (25–35 times longer than an incandescent lamp).

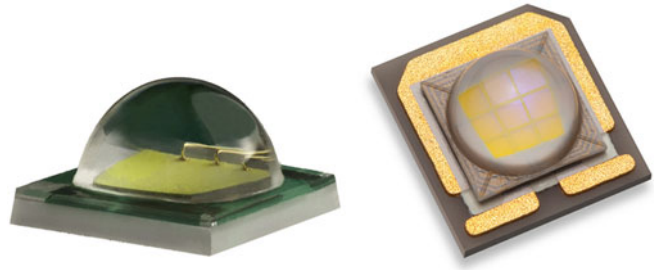
Lumen Depreciation

The electric current passing through the chip's junction, and the heat generated in it, degrades the chip material and is so responsible for light depreciation. Decoloration of the housing and yellowing of the primary lens may be further reasons for lumen depreciation. In white-phosphor LEDs, chemical degradation of the phosphor material also causes lumen depreciation. As mentioned above in the section on LED lifetime, for high-performance LEDs 30 % lumen depreciation is reached at around 35,000–50,000 h.

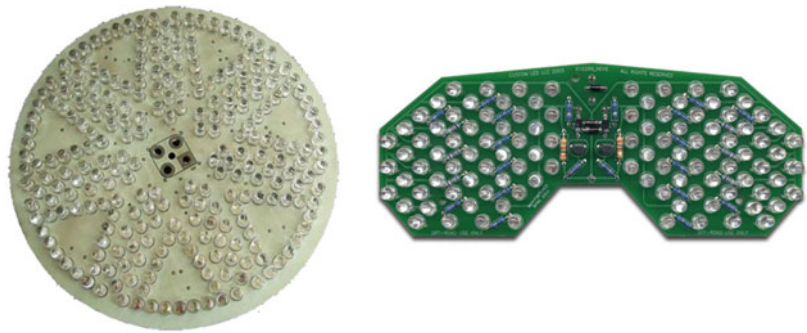
Run-Up and Reignition

LEDs give their full light output immediately after switch on and after reignition.

Light-Emitting Diode, LED, Fig. 16 Single LEDs



Light-Emitting Diode, LED, Fig. 17 PCB-mounted LEDs



Light-Emitting Diode, LED, Fig. 18 LED engines
(bottom: for use in road lighting)



Dimming

LEDs can be dimmed by simple pulse-width modulation down to 5 % of full light. Not all retrofit LED lamps can be dimmed on normal, commercially available dimmers. Special retrofit LED lamps that are designed to be dimmed on such dimmers are so specified on their packaging.

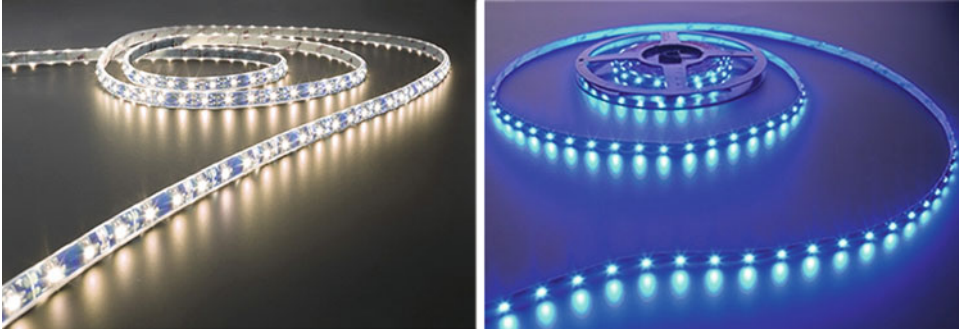
Ambient-Temperature Sensitivity

As has already been mentioned several times, limitation of the junction temperature of the LED chip is essential for the proper functioning of LEDs in terms of lumen output, lumen efficacy,

lamp life, and even color properties. In high-temperature environments the products perform worse, while at low temperature they perform better. The actual influence of the junction temperature is different for the different types of LEDs. In extreme temperature environments, therefore, relevant information for a particular product has to be obtained from the manufacturer.

Mains Voltage Variations

LED drivers are designed to drive LEDs on constant current. In this way the influence of mains voltage variations is not an issue.



Light-Emitting Diode, LED, Fig. 19 Foldable LED string

Light-Emitting Diode, LED, Fig. 20 Retrofit LED lamps for incandescent, halogen, and fluorescent lamps, respectively



UV and IR Component

LEDs radiate only visible radiation. There is no ultraviolet or infrared radiation.

Product Range

LED products are available as:

- Single LEDs (Fig. 16).
- Multiple LEDs on flat or three-dimensional PCB boards (Fig. 17).
- LED modules (or LED engines) with secondary optics and with or without built-in driver that can be used in the same way as lamps. Interfaces are standardized for interchangeability (Fig. 18).
- Multiple LEDs on strings (Fig. 19).
- Retrofit LED lamps with built-in driver and conventional design (Fig. 20); new type of designs are introduced for retrofit LED lamps as well (Fig. 21).

Light-Emitting Diode, LED, Fig. 21 Designer type of retrofit LED lamps



Cross-References

- [Light-Emitting Diode, OLED](#)
- [Phosphors and Fluorescent Powders](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Course Book. Philips Lighting, Eindhoven (2012)
2. Schubert, E.F.: *Light Emitting Diodes*, 2nd edn. Cambridge University Press, Cambridge, UK (2006)
3. Mottier, P.: *Led for Lighting Applications*. Wiley, Hoboken, NJ (2010)
4. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *IES Handbook*. 10th edn. Illuminating Engineering Society of North America, IESNA, New York (2011)

Light-Emitting Diode, OLED

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Optoelectronic light source](#); [Solid-state light source](#)

Modified and reproduced from Ref. [1] by permission of Philips Lighting, Eindhoven, The Netherlands. © 2012 Koninklijke Philips Electronics N.V.

Definition

Light source that produces light as a result of recombination of positively charged holes and negatively charged electrons at the junction of organic, solid-state, p- and n-type planar semiconductor materials.

Solid-State Light Sources

LEDs are solid-state radiators where the light is created inside solid-state material [2–4]. The light-radiating diode versions are called light-emitting diodes or LEDs. Light-emitting diodes made of organic semiconductor material are referred to as OLEDs; they are planar light sources (Fig. 1). Organic material is a semiconductor chemical compound whose molecules contain carbon and hydrogen (C – H bonds). The name LED is commonly used for light-emitting diodes made of inorganic semiconductor material; they are quasi point light sources. A separate chapter deals with LEDs. Serious development of OLEDs only started in the mid-1990s of the last century. OLEDs can be produced to give most colors of the spectrum and white. The technology permits producing OLED windows that are transparent when not switched on (Fig. 2). Currently, the main application for OLEDs is displays for mobile phones and for other display screens including

Light-Emitting Diode, OLED, Fig. 1 Flat OLED light sources



Light-Emitting Diode, OLED, Fig. 2 OLED window [1]



small television screens. OLED lighting products are now gradually also coming onto the market for illuminating purposes. Initially, the most interesting application for OLEDs, in this respect, is architectural, decorative lighting. For general lighting purposes the efficacies have to be improved much further.

these recombinations results in the emission of light. The color of the light is dependent on the composition of the semiconductor material. White light can be obtained by bringing phosphorescent material in the emissive layers.

Materials and Construction

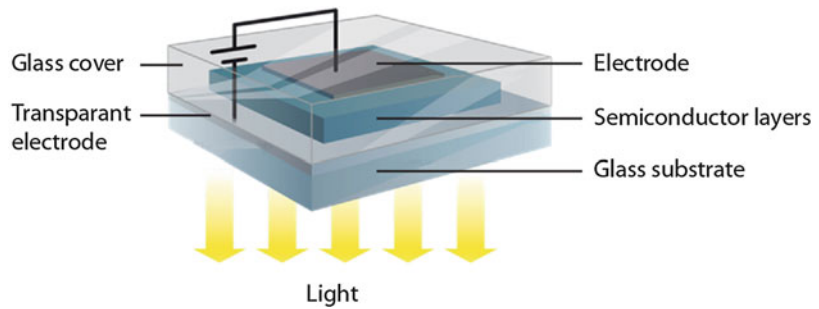
Working Principle

The process that is responsible for the emission of light in OLEDs is very similar to that with LEDs: positively charged holes and negatively charged electrons are pushed through semiconductor layers towards each other and recombine (see entry “► [Light-Emitting Diode, LED](#)”). Part of

The semiconductor organic layers are placed between electrodes, the one on the light-escape side being transparent (Fig. 3). The layers are supported by a glass substrate and, for protection of the organic materials against oxygen and water, are sealed in glass. Development is going on to substitute the glass seal by thin-film encapsulation. This development not only

Light-Emitting Diode, OLED,

Fig. 3 Composition of an OLED [1]



reduces the thickness and weight of OLEDs but also opens up the possibility of bendable OLEDs. The first laboratory examples have already been produced.

Properties

Commercial OLED products for lighting are just recently coming onto the market. Data of the properties of commercial products are only scarcely available. It is expected that they will change relatively fast the coming years. Only a rough indication will therefore be given here.

System Luminous Efficacy and Brightness

Commercially available white OLEDs now have efficacies between 15 and 30 lm/W. Ultimately, white-light, large-sized OLEDs with efficacies up to 150 lm/W would seem to be possible. Products with luminance values of up to 2,000 cd/m² have been shown. Compare this with a luminance value of some 6,000–10,000 cd/m² of fluorescent tubes.

Dimensions

Today sizes go up to some 30 × 30 cm. The expectation is that that sizes of more than 100 × 100 cm will be available in just a few years.

Lifetime

Lifetimes of some 10,000 h have been reported.

Color Properties

As has been mentioned, by choosing different materials of semiconductor material, different

colors are obtained. The quality of white light depends upon the combination of the semiconductor and phosphorescent material used. Spectra similar to those of inorganic LEDs are obtained (see entry “► [Light-Emitting Diode, LEDD](#)”).

Cross-References

- [Light-Emitting Diode, LED](#)
- [Phosphors and Fluorescent Powders](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Course book, Philips Lighting, Eindhoven (2012)
2. Schubert, E.F.: *Light Emitting Diodes*, 2nd edn. Cambridge University Press, Cambridge (2006)
3. Mottier, P.: *Led for Lighting Applications*. Wiley, Hoboken, NJ (2010)
4. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *Illuminating Engineering Society of North America*, New York. 10th edn. (2011)

Lighting Control Systems

- [Lighting Controls](#)

Lighting Controls

Peter Dehoff

Zumtobel Lighting, Dornbirn, Austria

Synonyms

[Lighting control systems](#); [Lighting management](#)

Definition

Lighting controls are intelligent network-based lighting solutions. They incorporate the communication between various system inputs and outputs related to lighting control. Input can be given manually by humans and/or by sensor signals or parameter settings. Part of the control systems are unusually one or more local or central computing devices.

Basics

Lighting control is understood to mean the manual or automatic tuning of a lighting situation inside a building, fixed to a building or in an outdoor area. In its simplest form, the tuning is performed using a light switch or dimmer; in more complex configurations, sensors are used as well as control curves (time lines) and/or control inputs. As a basic requirement, the lighting installation must allow the light sources it incorporates to be varied individually, in groups or all together. As a minimum, this means the ability to alter the brightness of the light sources but can also include light color and/or spectral composition or even colors. The lighting installation incorporates not only the light fittings (luminaires), operator control units, and control devices, but also the blinds or other shading devices which regulate the amount of daylight entering the building. The emergency lighting system can also be integrated. In the case of building management systems, the lighting installation is linked up with other automated services such as heating, ventilation, and air-conditioning.

This type of integrated lighting control is often referred to as lighting management. The purpose of lighting management is to provide the right light in the right place at the right time [1–3].

Overview

The usefulness of lighting control is easy to understand as tuning the light for a particular purpose is not an invention of modern times. Human beings intuitively seek the ideal lighting situation. They spend their lives alternating between periods of activity and relaxation in keeping with the rhythm of day and night. Over the course of the day and the year, they seek and experience the change and variation of natural light.

The invention of electric light brought with it the advantage of having “brightness at the push of a switch,” available at all times. However, there was also an inherent technical limitation: the ability to tune the electric light was initially restricted to switching it on and off. Lighting was a static arrangement. Despite the fact that the incandescent lamp was dimmable from the outset, this feature proved difficult in the case of discharge lamps. The first widely used fluorescent lamps as well as high-intensity discharge lamps proved reluctant when it came to dimming. Both the physics of these lamps and the ballasts required to operate them made it difficult to regulate their light output without considerable effort. Merely switching them on and off therefore became the common practice. Switching different groups of luminaires was the only approach that came anywhere near “lighting control.”

In the meantime, the available technology has vastly improved. Today’s fluorescent lamps can be cost-effectively dimmed using electronic ballasts, and in the case of LED light sources, tuning the brightness was never a problem. High-intensity discharge lamps, on the other hand, continue to hold out against straightforward dimming.

Lighting control is attracting a lot of attention and is considered a necessary component of a lighting solution.

The main reasons for using lighting control are as follows:

To increase energy efficiency: Dimming and switching reduce the amount of energy consumed by the artificial lighting. When users want less or no light, the artificial lighting can be dimmed or switched off. And if sufficient daylight is available, this can partially or entirely replace the artificial lighting.

To improve the light quality: The ability to adjust the lighting environment to suit the activity or individual preference ensures optimum light quality. Light can be used to set the scene in interiors and in outdoor areas.

To enhance safety: Emergency lighting can be integrated.

To ensure flexibility: By readdressing the individual luminaires, it is possible to provide a fast and flexible response to changes in organization and work layouts. Individual lighting scenes (moods) can also be programmed and changed.

Applications

Where should lighting control be used? The emphasis will vary depending on the application:

Shops

Intuitive control points invite users to adapt the lighting scenes to suit the activity – ranging from light for working at the cash desk to attractive color changes in the lounge.

Growing awareness of the need to save energy has paved the way for dimmable luminaires in retail spaces. It is possible to enhance the visual impression of merchandise and architecture by coordinating changes in color temperature and luminance. High comfort, great flexibility, and low maintenance are the characteristic features of these controlled lighting solutions. The light spectrum, for instance, can be optimally adjusted to suit the illuminated object without the complication of having to change filters. Using control points, timelines, or daylight-linked management of the artificial lighting, the general illumination can also be gradually changed in line with expected light levels over the course of the day. The staging of merchandise to create strong emotional appeal and surprise

effects is achieved with the appropriate static and dynamic lighting scenes. Making use of available daylight not only saves energy but also helps to make zones within an interior particularly attractive.

Façades

The use of LED technology in conjunction with lighting control systems has brought about a revolution in façade illumination. The subtlest of messages can be communicated using media façades. Dynamic façade design in particular is aimed at directing the gaze and conveying information. While ecological discussion focuses primarily on the amount of scattered light in the night sky, avoiding the unnecessary use of light by defining sensible operating hours would seem more important. The ultimate purpose of staging façades with light is to attract attention. It therefore makes sense for façades to be lit up exclusively in the evening and morning hours when greater numbers of people are circulating. This approach enables the identity of companies and communities to be highlighted, structures outdoor spaces at night, and assists nighttime perception while simultaneously addressing ecological objectives.

Museums

Lighting control ensures that sensitive exhibits are only exposed to light which is absolutely necessary, i.e., the level of luminance or the light color required for good perception or when visitors are present, e.g., using occupancy sensing. On and off times can be defined for specific hours of the day. Blinds control and daylight sensors ensure that exactly the right amount of daylight is admitted to achieve a balance between the needs of architecture, human well-being, preventing harm to exhibits, and saving energy. The emergency lighting is discreetly and safely integrated into the lighting management system which provides central monitoring and ensures reliable visibility in an emergency situation.

Hotels

In hotel rooms, intuitive control points which allow the guest to make selections according to

individual preference are the first priority. The guest sets the lighting environment. With convenient control of the blinds, the levels of artificial light and daylight are adjusted to suit different activities and visual needs such as watching TV, putting on makeup, or reading. Dynamic lighting moods, with flexible definition based on timelines, or controlled according to weather situation or time of the day, significantly affect the well-being of the guests, particularly in hospitality or wellness areas. In entrance zones, lighting moods based on the outdoor light can also optimize visual adaptation for arriving guests, giving them security and orientation. Defined lighting scenes in conference areas support the use of different media and enable the appropriate light to be provided at the press of a button.

Health and Care

A homely atmosphere and care activities call for entirely different lighting scenes which are available at the press of a button. Intuitive operating features to suit the age and physical abilities of the occupants as well as the accessibility of the control units are the key to successful lighting solutions in hospitals and care homes. The amount of light required by the eye increases with age. For precise visual tasks, the artificial lighting can be individually adjusted by care staff and patients. In addition, the aging eye filters out blue light, affecting biological processes such as the internal clock and cycles of rest and repair. This must be compensated at specific times of the day by spending time outdoors or through biologically effective artificial light with high intensity and a high blue light component. The use of timelines in lighting management systems permits this interplay of artificial light and daylight at appropriate times of the day.

Office

Lighting control optimally addresses individual lighting needs depending on the age and visual task. Concentrated work and increasing communication call for entirely different requirements to be met by the lighting concept. Biologically effective artificial lighting components at the right time of day in addition to daylight help the internal

clock and raise alertness. Lighting management systems with a high level of automation can achieve maximum energy savings as well as flexibility in the case of office moves, thanks to time management, daylight-linked control, and occupancy sensing. Employees readily accept the technology when it gives them the freedom to alter their lighting environment. This means providing adequate means of control and small groups of luminaires with assigned access.

Industry

Industry offers particularly high potential for saving energy. Long periods of lighting use due to shift and night work as well as a lack of daylight in many work areas and rooms mean that investments in lighting management have short payback times. Lighting management ensures the required flexibility in production areas. New lighting installations have to be overdimensioned in order to take into account the effects of deterioration of the installation over time. This additional energy consumption can be counteracted with daylight-linked or constant light control which continuously adjusts lamp output in line with the available daylight or length of service. Integrated lighting solutions allow the interaction between different services. Maintenance and monitoring operations are also optimized through the incorporation of emergency lighting. Interfaces to other services ensure the convenient and cost-effective operation of buildings.

Education

New forms of teaching and new media technologies call for flexible room use and frequent adaptation of the lighting situation. Intuitive control units are used to select defined lighting scenes at the press of a button, i.e., for working in small groups or traditional teaching, reduced lighting levels for presentations using LCD projectors, or higher vertical illuminance for blackboards and flip charts. Comfortable control units enable the user to rapidly adapt the lighting situation.

Daylight activates human beings and enhances their feeling of well-being as well as their performance. With daylight-linked control of the artificial lighting or occupancy sensing, it is possible to

achieve maximum energy savings without restricting the quality of light. Blinds control not only improves the contrasts of presentation media but also increases room comfort because glare and heat gain can be minimized. Conveniently placed control units near the door and on the teacher's desk enable intuitive selection of the appropriate lighting situation.

Corridors and Stairways

Circulation zones are predestined for lighting management. Ideally, the lighting should only be activated if a person approaches. The light should be on once the person has entered the circulation zone. Movement detectors must be suitably positioned. The energy consumption of lighting in corridors and stairways can be further reduced in conjunction with daylight-linked control.

Streets and Public Spaces

Illumination is necessary for reasons of safety and orientation in streets and public spaces. Switching the lighting on and off in conjunction with the available daylight has long been a common practice. Nowadays, it is also possible to reduce brightness levels at hours of the night when traffic density decreases. Control is performed centrally for streets, districts, and public spaces.

Movement detectors can be installed to ensure that the brightness of the lighting anticipates the approach of pedestrians and cyclists. While the general brightness level is lowered, it can be increased at the point of use to ensure safe circulation.

Domestic

Being able to vary the lighting takes on particular emotional importance in the home setting. Well-being and work, social interaction, and intimacy all take place within the confines of a small space. Access to different lighting moods at the press of a button enables the occupants to set the scene for their private space with area, accent, and mood lighting. Blinds are frequently integrated into the control system to regulate solar heat gain. Time control is also a security feature which can be used to make the house appear occupied when the residents are away on holiday.

Individual lighting scenes can be selected from a central control panel. Switches incorporating a reduced number of scenes on doors and in readily accessible locations allow intuitive operation of the lighting installation.

The possibilities and the requirements of the application are taken into account in the lighting concept. The lighting designer incorporates the operator control units, the control devices, and the choice, number, and arrangement of luminaires, blinds, and other actuators.

Interfaces for Lighting Management

Generally, there are interfaces between the user, the environment, and the lighting installation.

The user can actively intervene using different operator control units:

- *Standard or momentary action switch, with or without dimming function:* The simplest form is a switch and/or dimmer frequently fitted near the door or in another readily accessible location in the room.
- *Multifunction switch:* Also near the door or in a readily accessible location, for calling up different lighting scenes.
- *Remote control:* This can be an IR or wireless unit or the user's mobile phone, for calling up individual lighting scenes.
- *Workstation control:* In this case, operation is through the user's own computer or using a control unit directly at the workstation.
- *Central control unit:* At a defined location, trained personnel can select specific scenes or timelines for the lighting installation, often using a touch screen.

The environment can be integrated by means of sensors or defined timelines:

- *Daylight sensors:* Sensors installed outdoors or near daylight openings detect the level of daylight brightness and transmit the appropriate signals to control or regulate the lighting installation (luminaires and blinds).

- *Presence or absence detectors*: These sense whether there is a person within their detection zone. The signals they transmit switch the lighting installation on or off.
- *Time signals*: The lighting can be switched to preselected scenes at set times. The simplest case is to switch off all the lights at specific times. It is also possible to call up timelines to automatically tune the lighting in accordance with programmed strategies.
- *User behavior sensors*: Signals based not only on presence or absence but also on user movement patterns can be used to initiate situation-based tuning of the lighting situation.
- *Environmental signals*: Other predefined data can be used for control purposes, e.g., to maintain constant luminous flux on the basis of set maintenance factor curves or to reduce installation output if the energy demand exceeds a specific limit.

Behind each operator function, there are signals which are processed by the lighting management system. This means, for instance, that if a sensor detects that a person has left the room, switch off can be delayed.

Intelligent control is based on the following logic: for each activity, a lighting scene is defined, saved, called up, and modified as required. Each scene is assigned a name and a symbol which are shown on the operator control units. The lighting scenes are activated automatically by means of presence detectors or simple time inputs. The basic philosophy is always the same: it must be possible to manually override any activated lighting scene.

Strategies for Lighting Management

What strategies can be used to achieve useful objectives with a lighting management system? Firstly, those objectives have to be defined and secondly, acceptance of the usefulness of the objectives put to the test.

One objective is to increase energy efficiency:

1. Individual switch on and off: This simplest objective is dependent on a switch being

located near the user. Users appreciate this possibility but frequently do not use it.

2. Daylight-linked control – switching function: Automatic switch on and off; the control input is the daylight measured by a sensor.
3. Daylight-linked control – switching function: Automatic switch off and manual/individual switch on; this requires a daylight sensor and a switch.
4. Daylight-linked control – dimming: Automatic setting at a constant value; the control input is the daylight measured by a sensor.
5. Daylight-linked control – dimming: Automatic switch off and individual switch on; this requires a daylight sensor and a switch.
6. Occupancy-based control – switching function: Automatic switch on and off; the control input is delivered by a sensor which detects human presence.
7. Occupancy-based control – switching function: Automatic switch off and individual switch on; this requires a presence detector and a switch.
8. Constant light control (without daylight): A constant level of luminous flux is maintained to compensate for the decrease in light output with increasing age of the installation; the control input is delivered by a programmed timeline or a sensor for the emitted luminous flux.
9. Time-based control: Programmed lighting scenes are called up, and the installation is switched on or off at set times.
10. Load shedding: Automatic limitation of energy demand.

Another objective is to improve the quality of light:

11. Activity-related control: Individual setting of a lighting scene to suit an activity; this includes the switching and dimming of individual blinds, luminaires, and groups of luminaires in the adjacent area.
12. Individual daylight control: Individual operation of blinds to avoid disturbances such as glare and heat.

13. Automatic daylight control: Automatic control of blinds to reduce heat gain and glare from direct sunlight and excessive outdoor brightness; this requires a daylight sensor.
14. Algorithmic lighting: Automatic sequence of variations in the lighting based on programmed rules.
15. Room-related scene setting: Selection of pre-set static or dynamic lighting scenes.

User Acceptance and Energy Efficiency

While the various strategies might bring great benefits from a planning perspective, they also have to be accepted by the users. Many field studies have looked at the question of user acceptance.

One of the essential requirements is the ability to manually override each activated lighting scene. In comparison with automated processes, however, this makes it difficult to calculate the parameters for possible energy savings, the determination of which is a major challenge in view of the many unknown factors.

The behavior of energy-conscious occupants receives too little attention in calculations of this kind. Making the groups of luminaires as small as possible provides greater opportunity to change the lighting situation manually and individually to suit requirements. Large rooms with several groups of luminaires do not lend themselves to this approach as it is difficult for large numbers of users to agree the preferred setting of the lighting installation. For this reason, a smaller group of luminaires and operator control units must be installed in close proximity to enable the user to make spontaneous changes.

Changing weather conditions and unforeseeable user behavior make it difficult to predict energy usage. The time delay for presence detectors also plays a role.

In view of their complexity, the dynamic changes taking place in the world of work and in building use are continually being analyzed and measured in field studies carried out by different industries. In the context of a lighting management system, human behavior will always be more difficult to predict than technically defined parameters (Table 1).

Technical Requirements

The transmission of information from the operator control units and control devices to the actuators is at the heart of any lighting management system. Actuators are luminaires, blinds, or other equipment integrated into the lighting management system.

Information transmission means the sending and receiving of signals, which can be achieved using various means:

- The signal “on/off/dim” is sent via the power line.
- Signal transmission between operator control units and control devices takes place using the power line.
- Signal transmission between operator control units and control devices and to the powered luminaire/actuator uses a separate control line with standardized control signals (DALI, DMX)
- Operator control units and control devices (e.g., computers) are networked via data lines. Communication is by means of bus (LON, KNX, Luxmate, etc.) or TCP/IP protocols.

Definitions

Bus protocols such as LON, KNX, and Luxmate: Bus systems network all technical services in the building and provide central control for heating, ventilation, window blinds, and security systems.

TCP/IP: Transmission Control Protocol/Internet Protocol (TCP/IP) is the set of communication protocols used for the Internet and is consequently also referred to as the Internet protocol suite.

DALI stands for Digital Addressable Lighting Interface and is a standardized digital interface for control gear and electronic ballasts. DALI can be used with a small number of lines to address large numbers of control circuits over large distances.

DMX is an acronym for Digital Multiplex, a standard for digital communication networks commonly used to control stage and event lighting. Nowadays, the DMX protocol is also frequently used by architects and lighting designers as it enables over 500 channels to be controlled individually with rapid signal sequences from a single central control unit.

Lighting Controls, Table 1 The table shows the effect of individual strategies on energy efficiency and light quality and assesses the associated user acceptance

Strategy	Operation	Energy efficiency	Light quality	Acceptance
1 Individual		Low	High	High
2 Daylight-linked control – switching	Auto off, auto on	Medium	Medium	Low
3 Daylight-linked control – switching	Auto off, manual on	High	Medium	Medium
4 Daylight-linked control – dimming	Auto off, auto on	High	Medium	Medium
5 Daylight-linked control – dimming	Auto off, manual on	Very high	Medium	High
6 Occupancy-based control – switching	Auto off, auto on	High	Medium	Medium
7 Occupancy-based control – switching	Auto off, manual on	Very high	Medium	High
8 Constant light	Auto	High	Medium	Medium
9 Time-based control	Auto	Very high	Medium	High
10 Load shedding	Auto	High	Medium	Medium
11 Activity-related control	Manual	Medium	Very high	Very high
12 Individual daylight control	Manual	Medium	High	Very high
13 Automatic daylight control	Auto	High	High	High
14 Dynamic lighting	Auto	High	Very high	Very high
15 Scene setting	Manual	Medium	Very high	Very high

Auto automatic

Luminaires, blinds, and other actuators integrated into the lighting management system have to be compatible with the existing communication protocols and understand the signals. The actuators are assigned individual or group addresses so that they can be uniquely identified within the network and to enable them to communicate.

In the case of classic lighting installations equipped with fluorescent lamps, dimmable ballasts have to be used. High-intensity discharge lamps, on the other hand, can rarely be dimmed, at best switched in stages. As a rule, LED lighting incorporates dimming. With relatively simple means, it also offers the opportunity to vary the light color (with a range from 2,700 to over 8,000 K) or even to set different colors using RGB diodes.

Lighting Design

The decision regarding a lighting management system should be made at an early stage of the project. Lighting management is an integral part

of the overall building management system. For this reason, the lighting management system will often come under a different budget to the actual lighting installation itself.

The choice and layout of the luminaires should allow different lighting scenes to be selected.

The location of operator control units, the layout of the luminaires, and in particular the integration of daylight should be part of a holistic design approach. Service and long-term support for the lighting management system will ensure that the operators obtain the expected benefits.

Factors Driving Lighting Management

The growing demands for lighting installations to be energy efficient are currently a strong driver for the use of lighting management. European green building directives are leading to national regulations for energy requirements in buildings, which cannot be met without the use of lighting management. Daylight-linked control and presence detectors are therefore a necessity.

Green building certification schemes also look to reduce environmental impact with the aid of intelligent lighting management. Well-known schemes include LEED, BREEAM, DGNB, and CASBEE, among others. As well as helping to reduce energy consumption, the integration of lighting management can improve the quality of light for the users.

Cross-References

- [Architectural Lighting](#)
- [Electrical Control Gear for Lamps](#)
- [Interior Lighting](#)
- [Road Lighting](#)

References

1. Craig, D.L.: *Lighting Controls Handbook*. The Fairmount Press, Lilburn (2007)
2. Simpson, R.S.: *Lighting Control*. Focal Press, Oxford (2003)
3. David, D.L., Houser, K.W., Mistrich, R.G., Steffy, G. R.: *The Lighting Handbook*, 10th edn. IES, New York (2011)

Lighting Design

- [Architectural Lighting](#)

Lighting for Indoor Spaces

- [Interior Lighting](#)

Lighting for Interior Spaces

- [Interior Lighting](#)

Lighting Management

- [Lighting Controls](#)

Linguistic Influences on Color Perception

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

Linguistic Relativism

- [Color Category Learning in Naming-Game Simulations](#)

Lippmann, Jonas Ferdinand Gabriel

Mark D. Fairchild
College of Science, Rochester Institute of
Technology, Rochester, NY, USA



Gabriel Lippmann was a French inventor and physicist (born in Luxembourg) who created the first color photographs using what was certainly

the first spectral imaging system. Lippmann's system of color photography was conceived in 1886 and then refined for several years due to the complex nature of its theory and implementation. The system works by placing a very fine-grain photographic plate in contact with mercury that acts as a mirror. Light waves pass through the emulsion, reflect from the mercury backing, and then create an interference pattern within the emulsion. The developed plate then has an interference filter built into it due to the properly spaced layers of silver in the emulsion (created by the interference pattern exposure). The plates are then viewed with directional lighting, and the observer sees the same wavelengths that were present in the scene – a spectral image reproduction. Lippmann presented the first color photograph using his system in 1891 and then presented several nearly flawless photographs made by Auguste and Louis Lumière created with the process. The process was difficult and time-consuming, and few have been able to replicate the stunning photographs. Lippmann's work certainly presaged modern photographic and holographic processes, and for it he received the 1908 Nobel Prize in Physics.

Despite ending up a full professor at the Sorbonne (University of Paris), Lippmann had no formal education beyond high school. He was a student at the École Normale, but failed the examination that would have qualified him as a teacher due to his penchant for concentrating only on the work that interested him and neglecting the rest. However, he was appointed to a government scientific mission to Germany where he was able to work with the likes of Kirchhoff and Helmholtz. At about the same time, in 1873, he invented the Lippmann capillary electrometer for precise measurements of extremely small electrical voltages. It served as the basis for early echocardiographs. Lippmann joined the Faculty of Science in Paris in 1878, became Professor of Mathematical Physics in 1883, and was later appointed Professor of Experimental Physics and Director of the Research Laboratory. He made many contributions in various fields of physics including electricity, thermodynamics, optics, and photochemistry.

In addition to his Nobel Prize for the Lippmann process of full-color photography, Gabriel

Lippmann served as Marie Curie's thesis advisor at the Sorbonne and let her use his laboratory for her thesis work in radioactivity and helped her find other sources of support. Lippmann died at sea in 1921 while returning from a voyage to Canada. There is no record of the cause of death.

Lippmann's Nobel lecture on color photography can be found at the following link:

http://www.nobelprize.org/nobel_prizes/physics/laureates/1908/lippmann-lecture.html

Lorenz-Mie Theory

► [Rayleigh and Mie Scattering](#)

Lorenz-Mie-Debye Theory

► [Rayleigh and Mie Scattering](#)

Lovibond, Joseph Williams

Mark D. Fairchild

College of Science, Rochester Institute of Technology, Rochester, NY, USA



Joseph Lovibond was a British chemist and brewer and is credited with inventing the commercial colorimetry, the Lovibond Tintometer. In Lovibond's *Light and Colour Theories* [1], we read "the writer was formerly a brewer, and this work had its origin in an observation that the finest flavour of beer was always associated with a colour technically called 'golden amber,' and that, as the flavour deteriorated, so the colour assumed a reddish hue." Such observations of the relationships between beer quality and observed color led Lovibond to his work on color standards as a reliable means of reference and to the development of an instrument, the visual colorimeter, in which such standards could be systematically and objectively applied.

The general form of the Lovibond Tintometer was a visual colorimeter in which a split field is viewed. One half of the split field represented the test sample, perhaps a cuvette of beer placed in a beam of light. The other side of the split field was composed of subtractive primaries of adjustable density (e.g., a series of cyan, magenta, yellow, and neutral glass filters) that could be adjusted to select the density and overlapped one another in the adjacent beam of light. By adjusting the density of the filters, observers would match the test color stimulus and record the densities of the standard filters required for the match. Note again that the Tintometer system was one of subtractive color mixing of standard materials rather than the more common additive color mixing of standard lights typically found in visual colorimetry.

While brewing beer might have been the motivation for the Tintometer, by 1914, the system was in use in a wide variety of additional industries including tanning, wine and spirits, dyeing and printing, paint, water chemistry, ceramics, various oils, and hematology. In many cases, specific versions of the instrument and the reference standards were produced for a given application. The system was also presented with a number of awards by international juries (including two gold, five silver, and two bronze medals) along with significant recognition from ten scientific societies (one gold, three silver, and five bronze medals and a diploma). Lovibond worked

tirelessly in promoting his system through lectures and demonstrations around the world. He created a technically successful system that also met with commercial success. In fact, The Tintometer Limited still exists to this day with products such as color standards and scales, visual colorimeters (including versions of the Lovibond Tintometer), and photoelectric colorimeters and spectrophotometers. For example, one can still purchase instruments and standard color scales designed for American (ASBC) and European (EBC) methods of specifying beer color. Lovibond's initial inspiration is still being addressed by the progeny of his instruments using his very techniques.

In the history of The Tintometer Limited, it is stated that "The Company was founded in 1885 by Joseph Lovibond, a prominent brewery owner who developed the 'colorimeter' as a means of ensuring the quality of his beer. By 1893 he had perfected his research and introduced the first instruments. Much has developed since then. Today, the company is bringing colour measurement to the next generation. While still recognizing the importance of traditional methods, The Tintometer Ltd is introducing new techniques to bring measurement and quality control to an even higher level, developing creative solutions to ensure the continued reputation of the Lovibond® brand" [2].

References

1. Lovibond, J.W.: *Light and Colour Theories and Their Relation to Light and Colour Standardization*. E. & F. N Spon, Haymarket, London (1921)
2. www.lovibond.com

Low-Pressure Glow Discharge Lamps

► [Neon Lamp](#)

Low-Pressure Mercury Gas Discharge Lamps

► [Tubular and Compact Fluorescent Lamp](#)

Lumen Depreciation

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Light depreciation](#)

Definition

Light losses during the use of a lighting installation, caused by the decrease of the luminous flux of lamps, the dirt accumulation on and loss of reflectivity and or transparency of luminaires, and dirt accumulation on or discoloration of room surfaces (in interiors).

Lamp Lumen Depreciation

The light output of virtually all lamp types declines gradually during operation. The causes of light output depreciation are numerous. With incandescent lamps, it is especially the blackening of the bulb – caused by evaporation of the filament. Discharge lamps also suffer from blackening, in this case due to scattering of the electrode material, which settles on the wall of the discharge tube. With fluorescent lamps, high-pressure mercury lamps, and solid state, LED, and OLED light sources with a fluorescent coating, the major cause of light output depreciation is a gradual exhaustion of the fluorescent powder, which slowly loses its effectiveness.

Because of the different causes of lamp lumen depreciation, the actual rate of decline of lumen output is different per lamp type. Lamp manufacturers should be able to supply lamp lumen depreciation curves for their lamps.

Luminaire Depreciation

The gradual reduction of output of a luminaire is caused by dirt that is gradually deposited on lamps

and on or in the luminaires and by a gradual loss of reflectivity of reflectors and mirrors or loss of transparency of refractors due to corrosion and discoloration. Where the effects of dirt accumulation can be offset by regular cleaning, the output loss due to corrosion and discoloration cannot be regained. The rate of output reduction depends on the materials used in the luminaire, the construction of the luminaire, and on the nature of airborne dirt in the atmosphere. The degree of protection provided by the construction of the luminaire is classified according to the international protection code, IP, as described in an international IEC standard.

Room Surface Depreciation

Gradual dirt collection on room surfaces and/or discoloration of these surfaces gradually reduces the interreflected component of light from the installation. This may especially have a noticeable effect in interiors of smaller dimensions where, because of the relative large wall area, interreflection components are relatively large.

Maintenance Factor

The maintenance factor is the ratio of illuminance or luminance produced by the lighting installation at the moment that maintenance operations are being carried out to the illuminance or luminance produced by that installation when new. The maintenance factor takes into account all light losses described above together with the average light loss of the installation due to failed lamps not being replaced. The actual value is of course also dependent on length of the period between maintenance operations. The term light loss factor has sometimes been used instead of maintenance factor. Depreciation factor has been formerly used to designate the reciprocal of the above ratio.

The recommended lighting level (in terms of illuminance or luminance) for a lighting design is based on “maintained lighting level” which is the average lighting level at the moment that maintenance is being carried out. This implies that the initial lighting level has to be:

$$L_{\text{initial}} = L_{\text{maintained}} / \text{Maintenance factor}$$

It is the responsibility of the lighting designer to make an economical balance between the recommended length of the maintenance interval and the extra light required for the initial installation. Shorter maintenance intervals increase the maintenance cost but allow for lower initial lighting levels with lower initial investment and lower energy costs.

Cross-References

- [Interior Lighting](#)
- [Luminaires](#)
- [Road Lighting](#)
- [Sports Lighting](#)

Luminaires

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Fitting](#) (Anglo-Saxon English); [Fixture](#) (American English)

Definition

Devices that control the distribution of the light emitted by a lamp or lamps and that include all the items necessary for fixing and protecting the lamps (and sometimes the gear, too) and for connecting them to the electricity-supply circuit [1].

Luminaire Characteristics

The principal characteristics of luminaires can be listed under the following headings:

- Optical
- Mechanical
- Electrical
- Thermal
- Aesthetics

Optical Characteristics

The optical characteristics of a luminaire determine the shape of its light beam, or light-output distribution. The light distribution of a luminaire defines how the luminous flux radiated by the lamp or lamps is distributed in the various directions within the space around it. Different lighting applications require different light distributions and thus different luminaires.

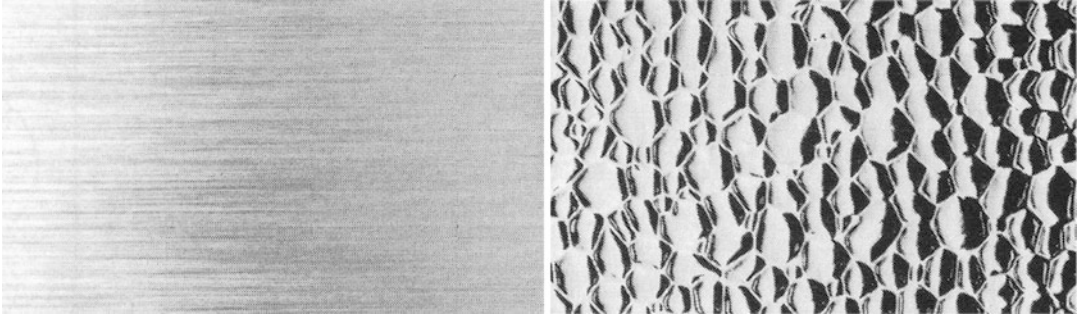
The desired light distribution of a luminaire is obtained through the application of one or more of the physical phenomena: reflection, refraction, and diffuse transmission. Many luminaires also make use of shielding in one form or another, principally to obtain the required degree of glare control and to limit light pollution. The shielding function may be performed by refractors or diffusers or by mirror reflectors, by white-painted surfaces, or, where very stringent glare control is required, by black surfaces.

Reflectors

Many conventional luminaires are provided with a reflector (sometimes in conjunction with another light-control element) in order to create the appropriate light distribution. The reflecting material that is used for reflectors can be specular, spread, or diffuse.

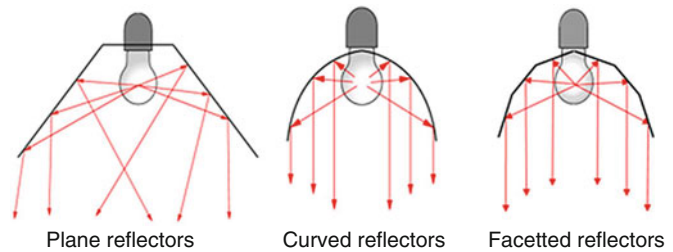
Specular Reflectors

Specular reflectors (also called high-gloss mirror reflectors) are used when a precise form of light distribution is required, as in floodlights, spotlights, and road-lighting luminaires. The reflector creates multiple images of the light source. The most widely used material is sheet aluminum, which has the strength needed to produce a stable reflector. Reflectance values are around 0.70. Alternatively, commercial-grade aluminum can be clad with a thin layer of super-purity aluminum



Luminaires, Fig. 1 A spread finish as produced by brushing (*left*) or by hammering (*right*) a specular surface

Luminaires, Fig. 2 Basic reflector forms



or silver. With aluminum, reflectance values of up to 0.80 can be obtained, while with silver a reflectance of more than 0.90 is possible. Finally, there is vacuum metalizing, in which a specular layer of aluminum is deposited on a suitably smooth substrate (metal, glass, or plastics). The resulting reflectance, which is somewhere between 0.80 and 0.90, is dependent on both the substrate material and the quality of the metalizing process.

Spread Reflectors

With spread reflectors (sometimes also called half-matt reflectors), there is no sharp mirror image of the light source. They are employed where a moderate degree of optical control is required, with the emphasis on producing a beam with smooth transitions. Such reflectors also help to smooth out discontinuities in the light distribution caused by inaccuracies in the shape of the reflector. Spread reflection is produced by brushing or etching or by hammering very small dimples and bumps into a specular surface (Fig. 1).

Diffuse Reflectors

At the other extreme from specular reflection is diffuse reflection, which is also called matt reflection. Here light incident on the reflector is scattered in all directions, so there is no mirror image of the light source. Matt reflectors cannot provide sharp beam control, but are employed where diffuse or non-focused light distributions are required. Matt-finished metals and white glossy paints on metal or glossy-white plastics provide near-diffuse reflection. The small specular component due to the gloss is of no practical optical significance; the gloss merely serves to facilitate cleaning. Reflectance values can be in the range 0.85–0.90. Ceramic materials or finishes have completely diffuse reflection characteristics with extremely high reflectances of up to 0.98.

Reflector Forms

There are three basic reflector forms: plane, curved, and faceted (Fig. 2).

Plane Reflectors

When using a simple plane (or straight-sided) reflector, the light emitted by the light source is

reflected according to the material of the reflecting surface, viz., specularly or diffusely.

Plane reflectors are often used to screen off the direct light from the light source. Accurate beam shaping is scarcely possible with plane reflectors, but by changing the symmetry of the reflectors, the direction in which the bulk of the light is emitted can be changed.

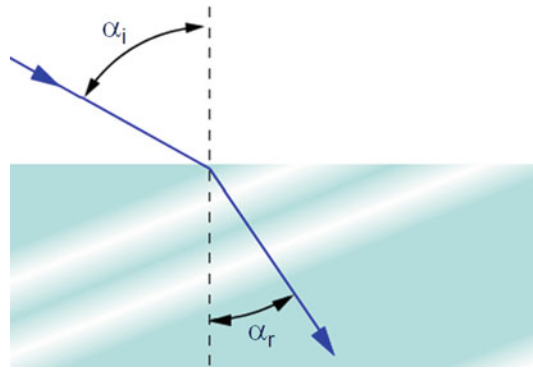
Curved Reflectors

The best optical performance is obtained when using a curved reflector. Depending on the curvature, many different types of beams can be created. A curved reflector may be cylindrical, parabolic, elliptical, hyperbolic, or some other contour to suit a particular application. The circular and parabolically shaped reflectors are the ones most commonly used.

The most important optical property of a parabolic reflector is that a point source of light placed in its focus will produce a parallel beam of reflected rays with the greatest intensity in its center. If the light source is not at the focus but in front or behind it, the reflected rays are no longer parallel. Thus, by choosing the position of the light source relative to the focus point, the desired beam shape (narrow to wide) can be created. Since a lamp is never a real point source, deviations from the theoretical beam shape for a point light source as sketched above will always occur. The smaller the light source relative to the size of the reflector, the more accurately can the beam be shaped.

Faceted Reflectors

Smooth-curved reflectors have to be produced to a high degree of accuracy, because even small deviations from their intended shape will produce undesirable discontinuities in their light distribution (striations). This will not occur with a faceted reflector. A faceted reflector consists of a number of adjacent, plane or curved, facets that together approximate a curve that is an approximation of a parabolic curve. The width of beam produced by the faceted reflector is somewhat greater than that of a smooth-curved reflector.



Luminaires, Fig. 3 Bending of light by refraction from the incident angle to the refraction angle

Refractors

Refractors are used to create the desired luminaire light distribution by passing the light from the source through a refractor (Fig. 3). The angle through which the light is bent is dependent on both the shape of the refractive material and its refractive index (Snell's law).

$$\frac{\sin(\alpha_i)}{\sin(\alpha_r)} = \frac{n_{air}}{n_{material}}$$

Refracting devices are either lenses or prisms. The type of refractor most commonly employed in indoor lighting is the lens found in tubular fluorescent-lamp luminaires intended for general lighting. It consists of a horizontal plastic panel which is mounted just below the lamp. The panel is flat on the top and has a special pyramidal (prism) or lens structure on the underside, which directs the light in certain directions and reduces the brightness under specific angles. Where in the past prismatic controllers were used, we see today more advanced lens-type refractors that give more accurate possibilities to shape the light distribution. These types of refractors are also used to produce LED luminaires for general indoor lighting and for road lighting (Fig. 4). Refracting glass bowls were in the past sometimes used for high-pressure mercury and sodium road-lighting luminaires. They have become obsolete because they are heavy, but more so because lighting control in the upward direction, and therefore control of light pollution, is not easily attainable.

Diffusers

Translucent diffusers enlarge the apparent size of the light source. They scatter the light of the lamp in all directions without defining its light distribution. They serve mainly to reduce the brightness of the luminaire and thus the glare created by it. Diffusers are made of opal glass or translucent plastic, commonly acrylic or polycarbonate. The material should be such that it scatters the light while producing the minimum amount of absorption.

Screenings

Screening the Lamp from Direct View

Screening is employed to hide the bright lamp or lamps from direct view. The higher the brightness



Luminaires, Fig. 4 Micro-lens type of refractor for a LED luminaire for general indoor lighting

(luminance) of the lamp, the stricter the requirements for the shielding.

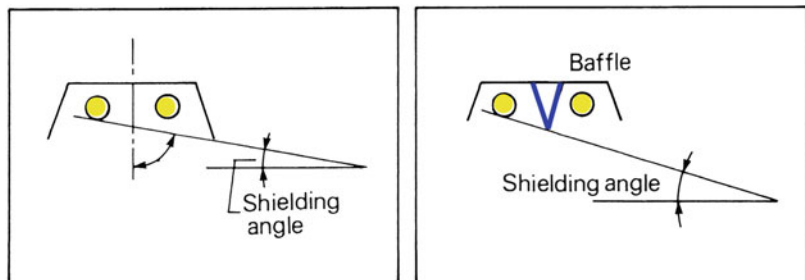
The luminaire reflector housing itself, or a built-in baffle, can provide the screening function (Fig. 5). When the sole purpose of the louvre is to shield the lamp from view, diffuse-reflecting material is used, such as a white-plastic louvre or, in the case of floodlights, matt-black metal rings (Fig. 6).

Shielding devices are often combined with the function of defining the light distribution, in which case highly reflective material is used for the louvre.

Color Filters

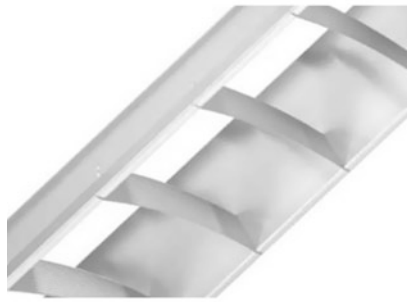
In certain lighting applications, in particular display lighting and decorative floodlighting, color is sometimes used to help achieve the desired aesthetic effect. In the past, color filters attached to luminaires containing white light sources were extensively used for this purpose. Both absorption and dichroic (interference) filters were used, although absorption filters in particular (Fig. 7) lower the efficacy of the total lighting system. Typical transmittance values are as follows: for blue absorption filters 5 %, for green absorption filters 15 %, and for red absorption filters 20 %. The consequence of the light absorption is that these filters become warm, which with high-power floodlights may damage the filter. The solution where such floodlights are employed is to use dichroic filters, which are a more expensive alternative. Today, colored LED light sources are normally employed where colored lighting is required. Here the color comes directly from the

Luminaires, Fig. 5 Lamp shielding by the reflector itself (*left*) and by an internal baffle (*right*)



Luminaires,

Fig. 6 Simple louvre (*left*) to shield the lamp in a fluorescent-lamp luminaire and (*right*) a floodlight louvre



Luminaires, Fig. 7 Absorption-type color filters

lamp itself, so the efficacy of the lighting system is much higher.

Light-Distribution Characteristics

The light distribution of a luminaire defines how the luminous flux radiated by the luminaire is distributed in the various directions within the space around it. It is the result of the optical devices used in the luminaire as described above. This is also called luminous intensity distribution, since it is specified in terms of luminous intensities in all the directions in which the luminaire radiates its light (Fig. 8). The luminous intensity diagram is in fact the light fingerprint

of a luminaire, in digital form (I-Table), and is the basis of all lighting calculations.

Basic photometric data that can be calculated from the light distribution are the beam spread and the luminaire light-output ratio. For all types of luminaires and for all types of application, these data provide an insight into the photometric quality of the luminaire.

Mechanical Characteristics

The mechanical function of the luminaire housing is threefold: it accommodates the various component parts of the luminaire, such as the optical system and the various components of the electrical system; protects these against external influences; and provides the means of mounting the luminaire in the installation.

Material

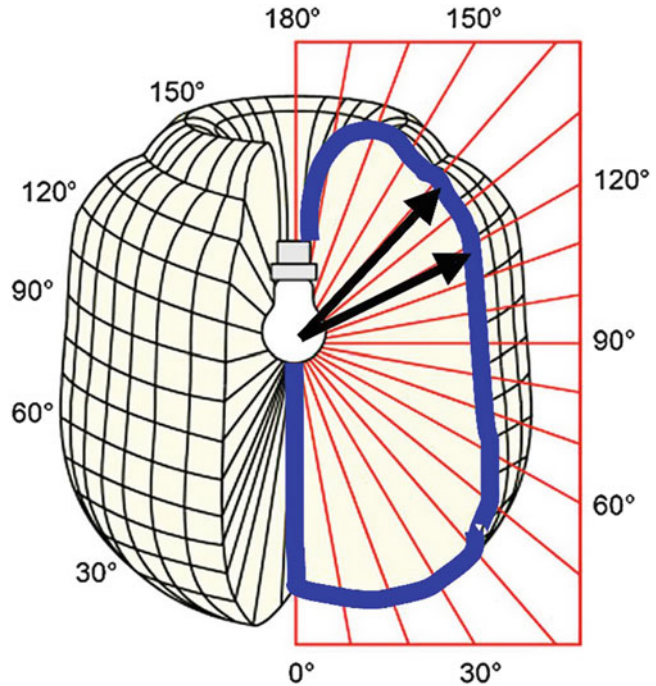
Sheet Steel

Sheet steel is generally chosen for the manufacture of tubular fluorescent luminaire housings for use indoors. The pre-painted sheet steel from the roll is white with diffuse reflection properties. Thus, after having been shaped in the luminaire factory into the desired luminaire form, no finishing-off operations are required.

Stainless Steel

Stainless steel is widely used for many of the small luminaire components, such as clips, hinges, mounting brackets, nuts, and bolts, that have to remain corrosion free.

Luminaires, Fig. 8 Light distribution of a luminaire given by its luminous intensity diagram. The arrows represent the luminous intensities in the directions specified. Here the light distribution is given for all planes, although it is usually only given for one (e.g., the *blue curve*) or two, mutually perpendicular, planes



Aluminum Alloys

Aluminum alloys, in which other elements have been added to the pure aluminum to improve their mechanical, physical, and chemical protective properties, are used to manufacture cast, extruded, and sheet-metal luminaires. Cast aluminum refers to the process in which molten aluminum alloy is poured (cast) in a mold. Extrusion is the process in which softened aluminum alloy is pressed through the openings of a die. Cast and extruded aluminum alloys are much used in housings for floodlighting and road- and tunnel-lighting luminaires because they can be employed in humid and damp atmospheres without having to add protective finishes. Sheet aluminum is chiefly employed in luminaires for reflectors (Fig. 9). The reflectors are anodized to improve their reflection properties and to protect them from becoming matt.

Plastics

Plastics are used for complete luminaire housings, for transparent or translucent luminaire covers, and for many smaller component parts. All-plastic houses can of course only be employed



Luminaires, Fig. 9 Mirror reflector of sheet aluminum protected by a plastic film, which must be removed before use

for light sources that have a relatively low operating temperature.

Plastic covers are of methacrylate or polycarbonate. Methacrylate maintains its high light transmission properties over a long period, but

its impact resistance is relatively low. The impact resistance of polycarbonates is very high and thus offers a high degree of protection against vandalism. It can be chemically treated to protect it from yellowing under the influence of ultraviolet radiation.

Glass

Although glass is heavy, glass covers are used where these have to be positioned close to a light source having a high operating temperature. This is the case, for example, with HID flat-cover road-lighting luminaires and with most floodlighting luminaires. Two sorts of glass are used:

- Normal glass, where no special demands are placed on heat resistance.
- Hard glass, where heat resistance, chemical stability, and resistance to shock are important. Should hard glass break, it will disintegrate into small pieces.

Luminaires made completely out of glass are extremely heavy and nowadays are seldom employed.

Ceramics

Ceramic material is used in compact housings that are exposed to very high temperatures such as very compact halogen lamp luminaires.

Strength

All luminaires should have housings of sufficient rigidity to withstand normal handling, installation, and use. With indoor-lighting luminaires for fluorescent lamps, stiffness and rigidity of construction is particularly important, since these lamps are relatively large and awkward to handle. Perhaps the most critical part of a luminaire as far as strength is concerned are the mounting brackets. The strength required here is covered by a safety factor: the mounting bracket (s) must be able to support at least five times the weight of the luminaire itself. With road-lighting and outdoor floodlighting luminaires, the mounting brackets must also be strong enough to withstand the highest conceivable wind loading for the location. Here a good aerodynamic shape for the

luminaire (characterized by its so-called shape factor) can be advantageous, as it also serves to reduce the strength required for the lighting mast.

Under some circumstances, the impact resistance of the luminaire itself is also important, particularly where protection against vandalism is called for.

Resistance to Pollution and Humidity

The atmosphere can contain many potentially corrosive gases which, in the presence of moisture vapor, will form highly corrosive compounds. In all areas where this danger exists (notably in outdoor applications, indoor swimming pools, and certain industrial premises), luminaires made from corrosion-resistant materials or having protective finishes should be used. In such areas, the luminaire should protect the optical and electrical components it houses. It should, of course, be fully enclosed. The degree of protection provided by the luminaire is classified according to the International Protection code (IP code) as described in an international IEC standard [2]. The IP code consists of two numerals: IP • •.

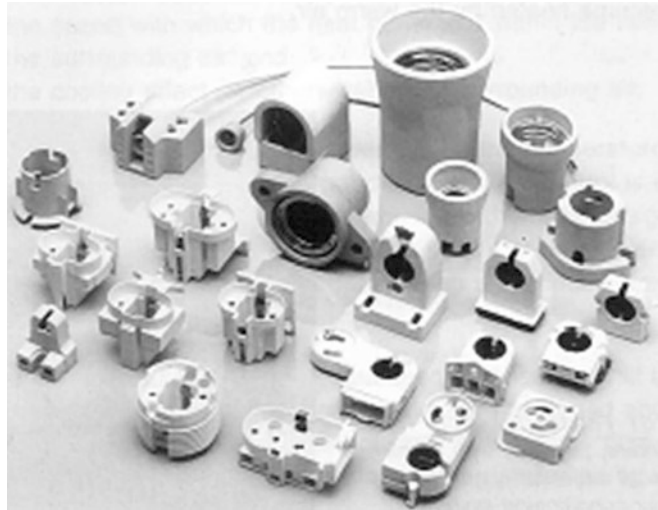
- The first numeral classifies the degree of protection against the ingress of solid foreign bodies (ranging from fingers and tools to fine dust) and protection against access to hazardous parts.
- The second numeral classifies the degree of protection against the ingress of moisture.

The higher the IP values, the better the protection and thus the lower the dirt accumulation and the lumen depreciation.

Ease of Installation and Maintenance

Many luminaires are of such a shape, size, and weight as to make mounting them a difficult and time-consuming operation. Mounting, but also relamping and cleaning, must usually be carried out high above ground level. So the ergonomic design of the luminaire should be such as to make these operations as easy and as safe as possible to perform. For example, covers should be hinged so that the electrician has his hands free to work on the lamp and gear. A good, ergonomically

Luminaires, Fig. 10 A variety of lamp holders made of both plastics and porcelain



designed luminaire is one that can be mounted in stages: first the empty housing or a simple mounting plate, which is light and easily handled, then the remaining parts.

Electrical Characteristics

The electrical function of a luminaire is to provide the correct voltage and current for the proper functioning of the lamp in such a way as to ensure the electrical safety of the luminaire.

Lamp Holders

The most usual types of holder are the Edison screw, the bayonet, and the pin (Fig. 10). Most Edison screw and bayonet holders are made of plastics or porcelain, with metal parts for carrying the current. Porcelain is resistant to high temperatures and has a high voltage-breakdown resistance, which is important considering the high ignition voltage of HID lamps. The pin lamp holders for tubular fluorescent and compact fluorescent lamps are nearly always made of plastic. The metal contacts are spring loaded to ensure a constant contact pressure.

Electrical Wiring

The electrical wiring in a luminaire must be such as to ensure electrical safety. This necessitates great care in the choice of wire used and its

installation. There are a great many different types of wire available, in both single-core (solid) and multi-core (stranded) versions, all with various cross-sectional areas and clad with various thicknesses and qualities of insulation. The cross-sectional area (thickness) of the wire must be matched to the strength of the current flowing through it. The insulation of the wire used must be resistant to the high air temperature in the luminaire and the temperatures of the luminaire materials with which it is in direct contact. This is true not only under normal conditions of operation, but also in the presence of a fault condition.

Mains Connection

The method used to connect a luminaire to the power supply must be both quick and safe. The practice generally adopted is to incorporate a connection block in the housing, although prewired luminaires in which the electrical connection to the mains is automatically made when the unit is placed in position are also available.

Electrical Safety Classification

The electrical safety classification drawn up by the IEC embraces four luminaire classes [2]:

- Class 0: Applicable to ordinary luminaires only. These are luminaires having functional insulation, but not double or reinforced

insulation throughout and with no provision for earthing.

- Class I: Luminaires in this class, besides being electrically insulated, are also provided with an earthing point connecting all those exposed metal parts that could become live in the presence of a fault condition.
- Class II: This class embraces luminaires that are so designed and constructed that exposed metal parts cannot become live. This can be achieved by means of either reinforced or double insulation. They have no provision for earthing.
- Class III: Luminaires in this class are those designed for connection to extra-low-voltage circuits (referred to as Safety Extra-Low Voltage, SELV). They have no provision for earthing.

Thermal Characteristics

Temperature Control

A considerable amount of the electrical energy supplied to the lamp is converted into heat. The ballast adds to this heating effect within the luminaire. With very high-powered lamps, the ballast should be placed outside the luminaire in a special ballast box.

For a given lamp/ballast combination, the working temperature reached by the luminaire is dependent upon three factors:

- The volume of the luminaire. The greater the volume, the lower will be the temperature rise inside the luminaire.
- The ease with which the heat generated within the luminaire can be conducted through it to the surrounding air. One way of promoting airflow through the housing is to make use of heat-conducting materials in its construction. Most metals are good in this respect, while plastics, on the other hand, are thermal insulators and cannot therefore be employed as housing materials where high-power lamps are involved.
- The cooling effect of the surrounding air. Good heat dissipation calls for large surface areas to



Luminaires, Fig. 11 Metal halide high-bay luminaire (*left*) and recessed LED downlight with cooling fins

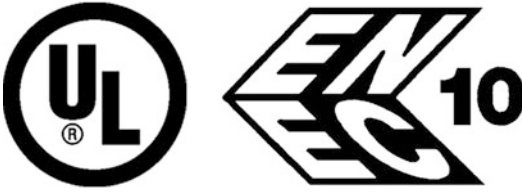
be in contact with the surrounding air. Luminaires for high-power lamps, such as high-bay luminaires, floodlights, and some LED luminaires that are very sensitive to high temperatures, are therefore provided with cooling fins (Fig. 11). Some types of industrial luminaires are provided with air vents in the top of the housing to allow the warm air to escape.

Protection Against Flammability

The flammability of a luminaire operating under fault conditions is an issue with luminaires made of plastics. But the combustion behavior of such luminaires is not just material dependent; it also depends on the shape and thickness of the luminaire housing.

Aesthetics

No less important than the functional characteristics of a luminaire is what is termed its aesthetical appeal, that is to say its appearance and styling. In interiors, all non-recessed luminaires are clearly visible, and so whether switched on or not, their design should be in harmony with that of the interior. In outdoor lighting, it is usually only the daytime appearance of the luminaires, when these are clearly visible, that is important: particularly in



Luminaires, Fig. 12 UL and ENEC certification marks. The number in the ENEC mark indicates the country of the institute that has given the European approval

built-up areas, their design can make a positive contribution to the attractiveness of the locality.

Approval

Luminaires always have to comply with the appropriate safety rules. UL (Underwriters Laboratories) is the US mark for demonstrating compliance with all US safety standards, and ENEC (European Norms Electrical Certification) is the similar European mark (Fig. 12).

Cross-References

- [Interior Lighting](#)
- [Light Distribution](#)
- [Light Pollution](#)
- [Lumen Depreciation](#)
- [Road Lighting](#)
- [Sports Lighting](#)

References

1. CIE Publication: S 017/E: 2011, International lighting vocabulary. International Lighting Commission, CIE, Vienna (2011)
2. IEC International Standard, 60598-1: ed 7.0, Luminaires – part 1: General requirements and tests. (2008)

Luminance Meter

Bor-Jiunn Wen

Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

Synonyms

[Luminance measurement device](#)

Definition

A luminance meter is a device used to measure the photometric unit, luminance, in a particular direction at a solid angle from a surface. The simplest devices measure the luminance in a single direction, while imaging luminance meters measure luminance in much the same way that a digital camera records color images.

Introduction

The luminance of a light source (integrating sphere with known output aperture) is determined from the measurement geometry and illuminance measured by a photometer. The luminance of the light source is obtained as

$$L_v = \frac{E_v D^2}{A} \quad (1)$$

where E_v is the illuminance at a distance D between the aperture plane of the source and the photometer; and A is the area of the source aperture. The distance A depends on the radius of the limiting aperture r_l , the radius of the source r_2 , and the physical distance d between the source and the aperture according to Eq. 2.

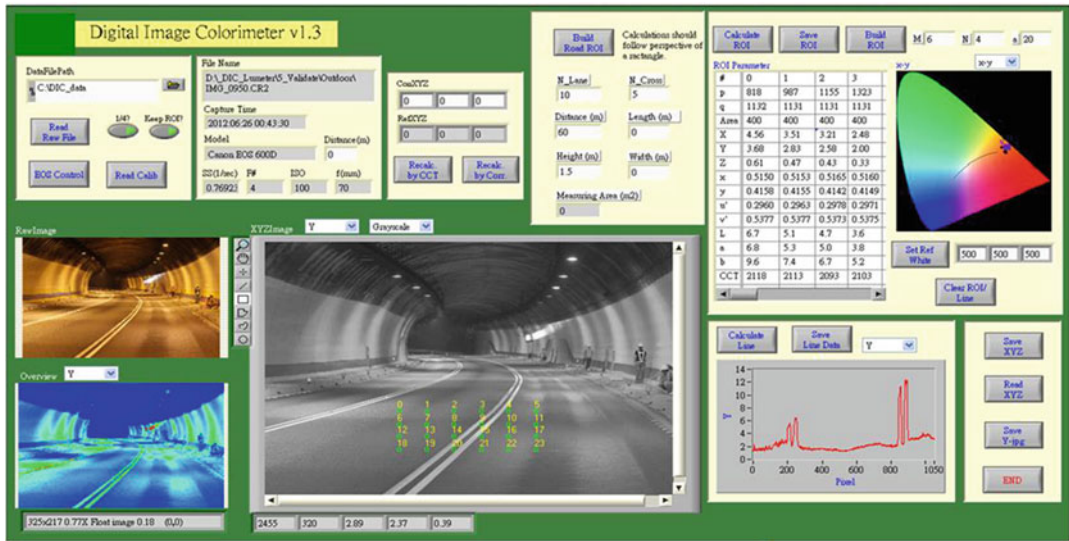
$$D_2 = r_1^2 + r_2^2 + d \quad (2)$$

Equation 2 is accurate within $\pm 0.01\%$ for distances that are more than one decade greater than the radius [1].

Luminance Measurement Device

- [Luminance Meter](#)

Luminance Meter,
Fig. 1 Sharpness
measurement of FPD by
using a ILMD



Luminance Meter, Fig. 2 Automatic measurement and analysis program for road and tunnel lighting systems

Applications

An imaging luminance measurement device (ILMD) is capable of measuring illuminance in an image. It can automatically quantify the light distribution or light uniformity. This is much more powerful than the conventional single-point luminance

meter. For example, an ILMD is utilized for the fast imaging luminance measurement of flat panel displays [2]. Figure 1 depicts measurement of the luminance of vertical or horizontal patterns on a display with an ILMD in a slightly tilted position [2]. The subsequent image processing includes calculating the spatial frequency response (SFR) of the

image using a slanted-edge algorithm, obtaining resolution at a specified decrease (e.g., 50 %) in the SFR of the device under test (DUT), and finally, specifying resolution in the sharpness of the DUT.

In addition, an automatic measurement system based on an ILMD also measures the photometric properties of road and tunnel lighting systems [3, 4]. Figure 2 depicts an automatic system for this purpose. Compared with a single-point luminance meter, an ILMD is more precise and time-saving for measuring road and tunnel lighting.

Since the rapid growth of light-emitting diode (LED) technologies, the adjustable lighting on billboards using LED have become increasingly popular. Many of these products are pushed to much higher contrast in spatial and/or temporal configurations to attract attention. However, the glare and/or flicker of the LED sources may produce an uncomfortable visual experience. Therefore, Hsu et al. [3] proposed a measurement system for regulating between flicker and glare by using perceptual ratings of LED billboards under various conditions. In this research the authors performed both objective and subjective evaluations of flashing LED billboards in interior spaces. The objective flicker and glare evaluations were carried out by taking temporal and spatial measurements respectively. The properties of the flicker and glare values obtained are reproduced under physical conditions. The visual results were modelled by simple equations as a function of objective measurements in terms of a low-pass flicker index and unified glare rating. Thus, an effective method was developed for regulating the degree of flashing LED lighting using an ILMD.

Cross-References

- [Standard Measurement Geometries](#)

References

1. Kostkowski, H.J.: *Reliable Spectroradiometry*. Spectroradiometry Consulting, La Plata (1997)
2. SID IDMS Information Display Measurements Standard, v1.03, 1 June 2012
3. Hsu, S.W., Chung, T.Y., Pong, B.J., Chen, Y.C., Hsieh, P.H., Lin, M.W.: Relations between flicker, glare, and perceptual ratings of LED billboards under various conditions. In: CIE Centenary Conference, Paris (2013)
4. CIE 194: On Site Measurement of the Photometric Properties of Road and Tunnel Lighting (2011)

Luminescent Materials

- [Phosphors and Fluorescent Powders](#)

Luminous Efficiency Function

- [Spectral Luminous Efficiency](#)

Luminous Intensity Distribution

- [Light Distribution](#)

Luminous Intensity Meter

- [Instrument: Photometer](#)

Luxmeter

- [Illuminance Meter](#)

M

M, P, and K Pathways

► [Magno-, Parvo-, Konio-cellular Pathways](#)

MacAdam, David L

Michael H. Brill¹ and Rolf G. Kuehni²

¹Datacolor, Lawrenceville, NJ, USA

²Charlotte, NC, USA



David Lewis MacAdam (1910–1998) was an American physicist and color scientist who made important contributions to colorimetry, color discrimination, color photography and television, and color order.

MacAdam grew up in Philadelphia, attended Lehigh University, and in 1936 received a PhD in physics from MIT. Under Prof. Arthur C. Hardy, he originated the first course in color measurement and assisted Hardy in the preparation of *Handbook of Colorimetry*, published in 1936.

Upon graduation, MacAdam joined the Research Laboratories of the Eastman Kodak company in Rochester, NY, from where he retired as a Senior Research Associate in 1975. Subsequently, he was named Adjunct Professor at the University of Rochester, Institute of Optics, where he remained active until 1995. At Eastman Kodak, among many other things, he helped to establish the theoretical basis for color photography, including color masking as compensation for unwanted dye layer absorptions [1].

Optimal Object Color Limits

While still studying, MacAdam published in 1935 two papers in which he offered a geometric proof of the optimal-color theorem and calculated the optimal object color solid raised over the CIE chromaticity diagram, using the newly established CIE standard observer and illuminant C and A data from 1932 [2, 3].

MacAdam Ellipses

One of MacAdam's best-known contributions was in support of technological color control. Assuming that the basis of color difference perception was the statistical error in matching the appearance of a given color stimulus, he conducted an extensive experiment with one observer, the result of which was expressed in the CIE chromaticity diagram in form of statistically derived ellipses, published in 1942 [4]. However, the resulting Friele–MacAdam–Chickering color difference formula proved less effective in predicting perceived color differences than formulas derived on other bases. In attempting to convert the ellipses to circles of equal size, MacAdam encountered the non-Euclidean nature of psychophysical color space.

Instrumentation and Computation

In the mid-1940s MacAdam pioneered the use of computers in colorimetric computations, established Hardy's reflectance spectrophotometer as a reliable industrial measuring instrument, and invented a tristimulus integrator as an accessory.

Principal-Component Analysis of Daylight

With Deane B. Judd and Günter Wyszecki, MacAdam performed the first principal-component analysis of phases of daylight of various correlated color temperatures, demonstrating that they can be represented as linear combinations of a limited number of spectral components [5].

Optical Society of America Uniform Color Scales

MacAdam was a leading member of the committee of the Optical Society of America that in 1947, at the suggestion of the US National Research

Council, began work on a perceptually uniform colorimetric model of the color solid. Upon the retirement of its first chairman D. B. Judd, MacAdam was elected chair. The result of the committee's work was published in 1974 as "Uniform Color Scales," [6] with OSA publishing a related color atlas with 558 samples in 1977.

MacAdam was a major contributor to OSA's 1953 book *The Science of Color* [7]. His interest in the history resulted in publication in 1970 of *Sources of Color Science* [8], a compilation of 26 seminal papers on color science, from Plato to Le Gros Clark, several translated for the first time into English. In 1981 he published *Color Measurement: Theme and Variations*, a presentation of fields of color science in which he made important contributions [9]. He also authored some 100 peer-reviewed journal articles.

MacAdam was president of the Optical Society of America in 1963 and editor of the *Journal of the Optical Society of America* from 1964 to 1975. He was active in the Inter-Society Color Council and the Commission Internationale de l'Eclairage (CIE). He received honors from many societies, including the Frederic Ives Medal of the Optical Society of America in 1974.

References

1. MacAdam, D.L.: Subtractive color mixture and color reproduction. *J. Opt. Soc. Am.* **28**, 466–480 (1938)
2. MacAdam, D.L.: The theory of maximum visual efficiency of colored materials. *J. Opt. Soc. Am.* **25**, 249–252 (1935)
3. MacAdam, D.L.: Maximum visual efficiency of colored materials. *J. Opt. Soc. Am.* **25**, 361–367 (1935)
4. MacAdam, D.L.: Visual sensitivities to color differences in daylight. *J. Opt. Soc. Am.* **32**, 247–274 (1942)
5. Judd, D.B., MacAdam, D.L., Wyszecki, G.: Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031–1036 (1964)
6. MacAdam, D.L.: Uniform color scales. *J. Opt. Soc. Am.* **64**, 1591–1702 (1974)
7. Optical Society of America Committee on Colorimetry: *The Science of Color*. Crowell, New York (1953)
8. MacAdam, D.L.: *Sources of Color Science*. MIT Press, Cambridge, MA (1970)
9. MacAdam, D.L.: *Color Measurement: Theme and Variations*. Springer, New York (1981)

Mach Bands

Sunčica Zdravković

Department of Psychology, University of Novi Sad, NoviSad, Serbia

Laboratory for Experimental Psychology, University of Belgrade, Belgrade, Serbia

Definition

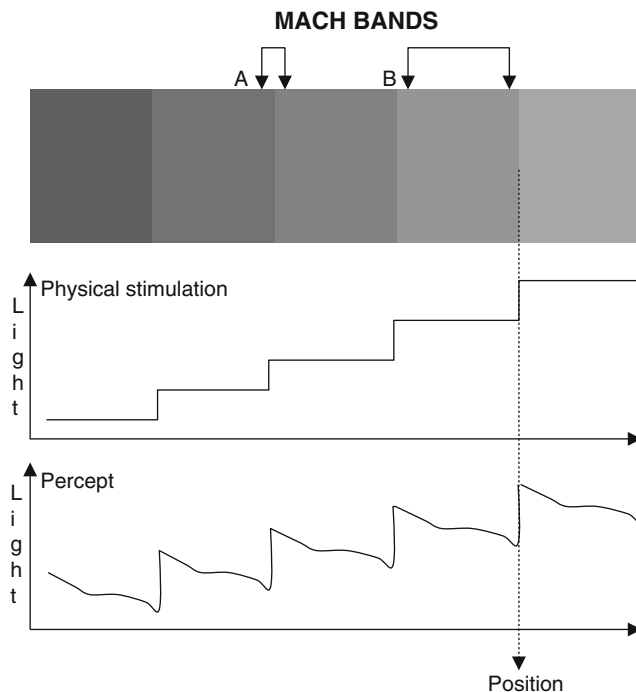
In this brightness illusion the physical contrast on the edge between the adjacent shades of gray is exaggerated (Fig. 1). Austrian physicist and philosopher Ernst Mach (1838–1916) introduced this, now famous, optical illusion in 1865.

Theoretical Importance

Mach bands are a common visual phenomenon and can be observed both on reflectance edges (as presented on Fig. 1) and on illumination edges, where a penumbra is created. They served in a number of vision paradigms because they are a robust phenomenon easily created in the laboratory and because they demonstrate important theoretical issues. Certainly the most general issue is the easily measurable difference between the physical stimulation and an emerging percept.

History of Research

It became clear very early that this simple display yields a physiological explanation. The first neural mechanism offered as an explanation was *lateral inhibition*. Mach himself envisioned a similar mechanism, a potential neural response that

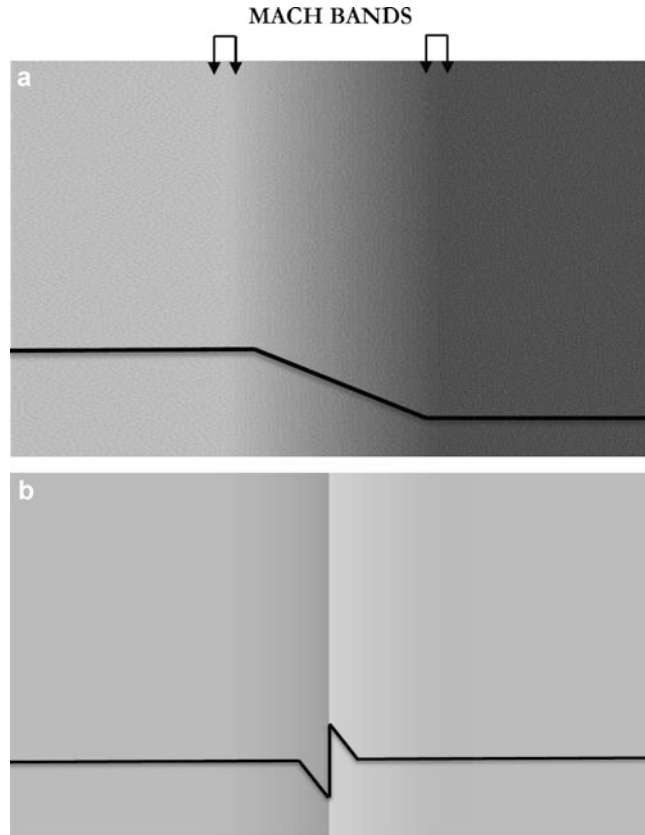


Mach Bands, Fig. 1 The achromatic contrast effect is noticeable at the edges where the two similar gray rectangles meet each other. (a) The darker of the two regions (i.e., gray squares) seems to have another even darker band near the border, while an extra light band appears on the lighter region. (b) The gray rectangles therefore do not appear to

be of uniform color. This illusion is one of the clearest demonstrations of the difference between physical stimulation and the perception. The two graphs represent actual luminance profile and (below) perceived lightness profile on the position of an edge. The underlying mechanism serves an important function of enhancing edge detection

Mach Bands, Fig. 2 (a)

Original version of the Mach band display, with luminance ramp in the middle. The illusory bands are marked on the *top*. The *black line* in the *bottom* represents luminance profile of the physical stimulus showing the position of the ramp. (b) Craik-O'Brien-Cornsweet illusion. *Left panel* appears darker. In fact both panels have the same luminance, and the illusion is created by the central luminance ramp. The *black line* in the *bottom* represents luminance profile of the physical stimulus showing the position of the ramp



would enhance one side of the edge and inhibit the other. When the actual retinal cells performing lateral inhibition were found in the Limulus eye [1], the connection seemed obvious. This elegant explanation was dominant for about a hundred years. During that time experimental studies in perception mostly covered the role of physical characteristics of the gray regions: their reflectance, slope, width of the bands, etc. Mathematical models were developed based on these psychophysical data. They were essentially based on distance-dependent excitation and surround inhibition, which takes place on the retina (presumably in ganglion cells).

Unfortunately, lateral inhibition did not make predictions that would later be confirmed in the experiments. The most obvious problem was the fact that the bands were most pronounced on the version with a ramp (Fig. 2b), while lateral inhibition should predict luminance steps as the strongest case (Fig. 1).

Vision research conducted between 1950 and 1990 established several important facts concerning Mach bands.

1. Bands are more pronounced on the lighter-gray field [2], and this effect is procedure dependent [3].
2. The gradient of the slope determines the size and intensity of the bands: steeper slopes produce brighter and thinner light bands and darker dim bands [4].
3. The Dark-adapted eye does not perceive bands, and the appearance of light bands depends on light adaptation [2].
4. Edge orientation is not important.
5. Bands produced by luminance ramps are thinner [5].
6. Visibility of bands is altered if luminance bars and ramps are placed next to the bands [6].
7. The shape of the gray regions is not important, while ramps of intermediate width produce

most visible bands. Sensitivity is lower for high spatial frequencies [7].

8. Phase relationships among Fourier components of the underlying waveforms of the gray regions influence the appearance of the bands [8].
9. The results on symmetry of light and dark bands are inconsistent.

The accumulated data revealed that, though important, retinal mechanisms alone cannot account for the phenomena. Primary visual cortex contained cells that proved to be another good candidate for a possible neural foundation of the described phenomenon [9]. In cat BA 17, there are cells that better respond to luminance ramps (even-symmetric cells) and cells that better respond to abrupt luminance steps (odd-symmetric cells).

Explanations

Several models were offered as a possible explanation for the accumulated data.

1. Tolhurst [10], based on his adaptation experiments, postulated “edge” detectors (odd-symmetric operators) and “bar” detectors (even-symmetric operators). Their spatially limited mutual inhibition would explain Mach bands on the ramps and the lack of bands in luminance steps. This theory was neat and simple but could not explain all the conditions. For example, it cannot explain why the width of the adjacent stimulus is not important.
2. The proposal offered by Morrone and Burr [11] was that the appropriate combination of the outputs for even-symmetric and odd-symmetric operators might better account for the experimental data. This, however, requires two sets of operators to indicate the position of “salient features” such as lines, edges, and bars. If the position of so-called maximum local energy of the feature corresponds to an odd-symmetric filter, the discovered feature is an edge. In the case of an even-symmetric filter, it is a bar (line). The model very precisely and quantitatively matched a large amount of existing data and predicted some later obtained results. However, the model could not account for intensity differences, i.e., for the change in appearance due to the change in contrast.
3. A Completely different approach was taken, at about the same time, by Watt and Morgan [12] who turned to rule-based rather than feature-based solutions. There are three rules (edge rule, bar rule, and null rule, corresponding to a luminance plateau) that interpret the output obtained by five computational stages, starting with the filtering by even-symmetric operators. The results are categorized into zero-bounded responses and regions of inactivity. After the rules are applied, it is possible to infer about luminance levels that would influence brightness. The model well predicts everything the previous models did (ramp vs. step, width of bands). Unfortunately, the model has a problem with spatial frequency that dictates the appearance of Mach bands.
4. A simpler model postulated only a single filter, with two channels, one selective for high- and one selective for low spatial frequencies [13]. The first channel does not respond to the ramp and only signals the existence of dark and light bars on the edge of the ramp. The second channel responds to the ramp and to the luminance edge. The two channels compose the final percept signaling the ramp and the two bars (i.e., Mach bands). The problem is the lack of a formalized rule specifying stimuli conditions that would elicit certain responses (bar or edge).
5. The next model tried to mimic the responses of V1 cells to lines and edges [14]. In the first phase the model produces Gaussian lines and error-function-shaped edges. Based on these, the reconstruction process recovers lines and edges in the scene. The model nicely predicts most of the data but was not developed enough to allow for proper testing.
6. The model proposed by Kingdom and Moulden [15] was designed to explain brightness perception in general. It starts with light adaptation, continues with multi-spatial scale filtering and a power-law transformation. After that, interpretation rules are used to calculate

brightness for each spatial scale. Finally, outputs from all the scales are averaged. This advanced model can now properly deal with the spatial frequency issue as well as most other experimental results from the brightness domain.

7. The latest attempts are filling-in models. They stress the importance of contours but without special interest in surface brightness. The model by Pessoa [16] can still explain Mach bands using special boundary computations responsive to luminance gradients and steps.

All of the presented models are concerned with identification of bars, lines, and edges mimicking only early vision processes. Nevertheless, the set of primitives even in early stages should be richer and able to include intensity changes. This would allow models to account more specifically for the already obtained data and enable more sophisticated predictions (including gradients of brightness, differences in intensities, temporal effects, etc.).

Filling-in type come closest to an explanation of Mach bands. Newer versions include the allocation of the edges followed by the filling-in process. However, it is still the edge that determines the outcome. The remaining issue for these models is the brightness of the areas within the edges.

Related Phenomena

The Chevreul illusion was introduced a few years before Mach bands (in 1839) and is nowadays presented in most textbooks (and our Fig. 1) as Mach bands. The original Mach illusion (Fig. 2a) has a ramp instead of uniformly gray surfaces, which are typical only for the Chevreul illusion.

Vision research the demonstrated that the Chevreul illusion is just a special case of a much broader phenomenon known as Mach bands.

Brightness/lightness contrast. The two effects were often linked together, probably because lateral inhibition was the first explanation offered in both cases. Later it became clear that lateral inhibition performed by retinal cells cannot be an explanation for either of these phenomena.

Craik-O'Brien-Cornsweet illusion (Fig. 2b). The same, more central mechanisms, such as simple V1 cells (BA 17), were proposed as an explanatory neural mechanism for both illusions.

Cross-References

- [Anchoring Theory of Lightness](#)
- [Color Contrast](#)

References

1. Ratliff, F., Hartline, H.: The responses of Limulus optic nerve fibers to patterns of illumination on the receptor mosaic. *J. Gen. Physiol.* **42**, 1241–1255 (1959)
2. von Bekeesy, G.: Mach- and Hering-type lateral inhibition in vision. *Vision Res.* **8**, 1483–1499 (1968)
3. Davidson, M.: A perturbation analysis of spatial brightness interaction in flashed visual fields. Ph.D. dissertation (unpublished). University of California, Berkeley (1966)
4. Fiorentini, A.: Foveal and extra foveal contrast threshold data point of a non uniform field. *Atti della Fondazione Giorgio Ronchi* **12**, 180–186 (1957)
5. Heinemann, E.: Simultaneous brightness induction. In: Jameson, D., Hurvich, L. (eds.) *Handbook of Sensory Physiology*, vol. VII-4, pp. 146–169. Springer, Berlin (1972)
6. Ratliff, F.: Why Mach bands are not seen at the edges of a step? *Vision Res.* **24**, 163–165 (1984)
7. Ross, J., Morrone, M., Burr, D.: The conditions under which Mach bands are visible. *Vision Res.* **29**, 699–715 (1989)
8. Morrone, M., Ross, J., Burr, D., Owens, R.: Mach bands depend on spatial phase. *Nature* **324**, 250–253 (1986)
9. Syrkin, G., Yinon, U., Gur, M.: Simple cells may be the physiological basis of the Mach band phenomenon: evidence from early monocularly deprived cats. *Soc. Neurosci. Abstr.* **20**, 312 (1994)
10. Tolhurst, D.: On the possible existence of edge detector neurons in the human visual system. *Vision Res.* **12**, 797–804 (1972)
11. Morrone, M., Burr, D.: Feature detection in human vision: a phase-dependent energy model. *Proc. R. Soc. Lond. B* **235**, 221–245 (1988)
12. Watt, R., Morgan, M.: A theory of the primitive spatial code in human vision. *Vision Res.* **25**, 1661–1674 (1985)
13. Fiorentini, A., Baumgartner, G., Magnussen, S., Schiller, P., Thomas, J.: The perception of brightness and darkness: relations to neuronal receptive fields. In: Spillmann, L., Werner, J. (eds.) *Visual Perception*:

- The Neurophysiological Foundations, pp. 129–161. Academic, San Diego (1990)
14. du Buf, J.: Ramp edges, Mach bands, and the functional significance of the simple cell assembly. *Biol. Cybern.* **69**, 449–461 (1994)
 15. Kingdom, F., Moulden, B.: A multi-channel approach to brightness coding. *Vision Res.* **32**, 1565–1582 (1992)
 16. Pessoa, L.: Mach band attenuation by adjacent stimuli: experiment and filling-in simulations. *Perception* **25**, 425–442 (1996)

Magno-, Parvo-, Koniocellular Pathways

Jasna Martinovic

School of Psychology, University of Aberdeen,
Aberdeen, UK

Synonyms

[Geniculate pathways](#); [M, P, and K pathways](#); [Parallel visual pathways](#); [Retinogeniculate pathways](#); [Retinogeniculocortical pathways](#)

Definition

Magno-, parvo-, and koniocellular pathways are the three visual pathways in primates. These pathways are established at the level of the lateral geniculate nucleus (LGN) of the thalamus. They are formed of morphologically distinct cellular layers that receive information from different types of retinal ganglion cells and project to different layers in the primary visual cortex.

Anatomical Considerations

The LGN layers of each of the three visual pathways have a specific cytoarchitectonic structure. The names of the pathways are derived from these structural characteristics. Magnocellular (M) cells have relatively large bodies (lat. *Magnus*: large) and are found in the lowest two layers (layers 1 and 2) of the LGN. Parvocellular (P) cells have

smaller bodies (lat. *Parvus*: small) and are found in the top four layers of the LGN (layers 3, 4, 5, and 6). Koniocellular (K) cells (gr. *Konios*: dust) are very small and form six thin layers. The first K layer is positioned ventral to the first M layer, while the others are intermediate to the six LGN layers. The M cell axons terminate primarily in layer 4C α of the primary visual cortex (V1) while the P cell axons terminate primarily in layers 4C β and 4A of V1 [1]. The six K layers are further subdivided on the basis of their projections [2]. The two dorsal K layers project into layer 1 of the primary visual cortex. The two intermediate K layers receive information from small bistratified retinal ganglion cells and project it into layers 3 and 4A of the V1. The two ventral K layers project into the superior colliculus.

Structural differences between the three pathways emerge at the level of the retina as each pathway receives projections from a somewhat different group of retinal ganglion cells. Although retinal ganglion cells that ultimately project to the P layers are heterogeneous, midrange ganglion cells form the bulk of the P pathway. The receptive fields of P cells which receive inputs from midrange ganglion cells oppose inputs from Long- (L) and Middle-wavelength (M) sensitive cones (L-M; reddish-greenish dimension). On the other hand, the M pathway is based on inputs from parasol ganglion cells. Receptive fields of these cells combine inputs from L and M cones (L + M; black-white dimension). The two intermediate K layers receive projections from small bistratified cells and from several other cell types with Short-wavelength (S) cone input. The receptive fields of such K cells oppose the excitatory input from S cones with the inhibitory input from L and M cones (S-(L + M); bluish-yellowish dimension). The distinction between the processing of luminance signals (L + M) and the processing of two types of chromatic signals (L-M and S-(L + M)) by geniculate cells had led to their conceptualization as the three orthogonal visual mechanisms. Such a conceptualization is implemented in the DKL space, a physiological color space based on the seminal study of macaque LGN cell response properties conducted by Derrington, Krauskopf, and Lennie in 1984 [3].

Historical Overview

Based on his observations of three different types of axons in the rabbit optic nerve Bishop speculated in 1933 that visual information is likely to be processed along morphologically distinct pathways that operate in parallel [4]. An in-depth historical overview of research into parallel visual pathways is provided in the review by Casagrande and Xu [1]. By the end of the 1980s, pathway-tracing studies on primates confirmed Bishop's speculation, establishing that different retinal ganglion cell classes projected to separate layers in the LGN, which subsequently projected to separate cortical layers. Numerous physiological studies uncovered marked differences in spatial, temporal, and contrast sensitivities between various cell types that were found in the retina, LGN, and the primary visual cortex. Early studies of primate LGN cell properties attempted to relate their findings to extensive research performed on cats since the 1950s. By the 1970s, cat retinal ganglion cells had been classified into X, Y, and W types based on their physiological properties and responsiveness to different types of stimuli: for example, X cells, which were the most numerous, were responsive to spatial detail, and the less numerous Y cells were responsive to motion. The W cells were found to have heterogeneous properties and were the least understood. Researchers linked different LGN layers to X, Y, and W cell types on the basis of observed similarities in their spatiotemporal response properties and sensitivities to various visual attributes. At the same time, attempts were also made to relate the newly found geniculate visual pathways to the ventral or "*what*" cortical stream, which was proposed to subserve the processing of objects' visual attributes, and the dorsal or "*where*" cortical stream, which was proposed to subserve the processing of objects' locations. These streams were proposed on the basis of findings from lesion studies on monkeys conducted by Ungerleider and Mishkin in the early 1980s [5]. In 1988, Livingstone and Wiesel proposed an integrative model of retinogeniculocortical pathways [6], linking the M pathway to the processing of motion due to the transient responses and high achromatic

sensitivity of its neurons and the P pathway to the processing of form and color due to its neurons' sustained response to chromatic contrast and weak response to achromatic contrast. Livingstone and Hubel's model epitomized the modular approach that was popular at the time, relating discrete brain structures to cognitive or perceptual functions. The koniocellular (K) pathway was ignored by Livingstone and Hubel's model. At that time, the K layers in primates were considered to be too thin to contribute substantially to any single cortical module. Between the 1950s and 1970s, the K pathway was generally studied in bushbabies and lorises, which are nocturnal prosimian primates [2]. Subsequent studies confirmed that K cells in these species were quite similar to those in other primates. Since then, a range of studies was performed on the primate K pathway, revealing diverse functional properties of K cells, including their role in color vision. At the same time, much more has become known of retinal processing, with 20 different types of retinal ganglion cells identified [7]. Cortical projections of M, P, and K cells have also been mapped more extensively, as will be discussed in the next section. These latest findings have brought into contention the actual number of parallel visual pathways and the degree to which they should be perceived as functionally distinct.

Functional Considerations

Numerous neuroanatomical and neurophysiological studies have investigated the properties of the three parallel pathways. As noted in the historical overview, continuous attempts have been made to relate the pathways to the processing of different visual properties: motion processing, form processing, luminance and color processing, and so on. As evidence accumulated it became more and more clear that simple and direct associations between physiological responses and perceptual properties are difficult to establish [8]. In fact, M, P, and K pathways can only tentatively be treated as singular entities: for example, the M pathway contains cells that project information into dorsal cortical areas subserving motion processing,

which fits Livingstone and Hubel's model, but it also contains cells that project information into ventral areas subserving object representation [1].

The most robust and the least disputed evidence for differential processing between M, P, and K pathways comes from primate color vision research. Strong links have been established between the M pathway and the luminance (L + M) channel. The chromatic L-M channel, which computes the difference between L and M cone signals providing a reddish-greenish dimension of color, maps on to the cells of the P pathway. Furthermore, the L-M cone-opponent mechanism provides signals that can be transmitted in combination with luminance information [9]. The view that such "multiplexing" through the L-M channel underlies achromatic vision was based on the fact that P cells are much more numerous than M cells and was initially strengthened by observations of poor spatial acuity of M cells, although recent data showed much smaller differences in M and P cell receptive field center sizes than originally found. In fact, there was an opposing view that strict subcortical segregation of luminance and L-M signals would be beneficial, since it would optimize signal transmission in a noisy, bandwidth-limited system such as primate vision [10].

Unlike P cells, K cells that receive S-cone input from small bistratified cells do not generally contribute to luminance processing. Their receptive fields are larger than those of P cells, compromising their usefulness for high-acuity spatial vision. However, S-cone sensitive K cells may still contribute to functions other than just color appearance. They project directly to motion-sensitive cortical area (MT/V5), and their signal is also amplified cortically, being weakly present in many cells from V1 onward. Projection of S-cone signals into area MT/V5 could underlie their contribution to motion processing [8].

While it is not possible to fully equate geniculate pathways with psychophysically identified luminance and chromatic channels, a certain degree of functional overlap between them is evident and has been used in many studies that rely on luminance or chromatic contrast to attempt to isolate or bias visual processing toward different

pathways. As M pathway dysfunctions have been hypothesized to underlie schizophrenia, autism spectrum disorders, dyscalculia, and dyslexia, the ability to isolate this pathway through carefully designed stimuli is of keen interest to clinical neuroscience. Studies in this field do not only compare responses to luminance stimuli with those to isoluminant stimuli but also often rely on other well-known processing differences between parallel visual pathways. Fast temporal frequency, low luminance contrast, and low spatial frequency are generally thought to bias processing toward the M pathway, while low temporal frequency, high chromatic contrast, and high spatial frequency are assumed to bias processing toward the P pathway. It is paramount to be careful when interpreting the findings of such studies. For example, besides transient response neurons, the M pathway also contains some neurons with sustained responses. Similarly, while contrast sensitivity functions for luminance and chromatic stimuli do indeed differ, the majority of spatial frequencies that lead to larger excitation in one pathway happen to concurrently cause some excitation in the other pathway. Therefore, the level of bias to one or the other pathway is very difficult to quantify. Finally, responsiveness of M, P, and K pathways to various visual features cannot be assumed to translate directly into perceptual sensitivity for these features. For example, P cells with receptive fields in the parafovea are sensitive to a wide range of temporal frequencies, but human observer performance reveals that the sensitivity of the visual system does not match the sensitivity that is present in this geniculate circuitry [10]. Thus, some of the sensitivity evident in P cells is not utilized at subsequent stages of processing that drive behavior.

Isolating subcortical processing streams or biasing processing toward a certain pathway is ultimately complicated by the fact that the separation of signals from different visual pathways is not maintained in the cortex. Receptive fields in V1 are larger than in the LGN, indicating that each V1 cell receives input from several LGN cells. S-cone signals become more prominent in the cortex, with weak S-cone signals being found in many V1 neurons. Finally, LGN cells with

chromatic sensitivity show preferential responses to stimuli that align with the S-(L + M) and L-M mechanisms, but V1 cells have much more widely distributed color preferences. This indicates that S-(L + M), L-M, and L + M signals start interacting in the cortex. In fact, most of the early visual cortex contains neurons that are tuned to both color and luminance, while also being sensitive to certain spatial properties, e.g., orientation [11].

Summary

Magno-, Parvo-, and Koniocellular pathways are established at the level of the LGN. They receive inputs from different retinal ganglion cells and project to different layers of the V1, leading to parallel processing of visual information. Parallel visual pathways have been extensively researched both in terms of their structure and in terms of their functional properties. Geniculate pathways are of interest for color scientists as the retinal projections they receive differ in chromoluminance content: while the M pathway is solely devoted to the processing of luminance information, the P and K pathways also subserve chromatic processing. Other differences between pathways concern contrast sensitivity and spatial and temporal processing, with the M pathway being preferentially responsive to high temporal but low spatial frequencies and to low luminance contrast. In spite of these recognized differences, studies that attempted to relate these pathways to specific visual functions have met with limited success. Such research is complicated by the fact that data on physiological responses of M, P, and K cells does not always correspond to behavioral data obtained from human observers under the same conditions. This is at least partly caused by the fact that visual signals undergo much further processing in the cortex, as signals from different pathways get recombined from primary visual cortex onward. There is growing evidence that contributions of M, P, and K pathways to various perceptual functions are not as segregated as was once thought. Recent research has also revealed that subcortically there are likely to be more than

just three parallel visual pathways, based on the evidence that the processing in the retina itself is already very complex and driven by a wide array of cells with different inputs and connectivity.

Cross-References

► [Ganglion cells](#)

References

1. Casagrande, V.A., Xu, X.: Parallel visual pathways: a comparative perspective. In: Werner, J.S., Chalupa, L.S. (eds.) *The New Visual Neurosciences*, pp. 494–506. MIT Press, Cambridge (2004)
2. Hendry, S.H.C., Reid, R.C.: The koniocellular pathway in primate vision. *Annu. Rev. Neurosci.* **23**, 127–153 (2000)
3. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (1984)
4. Bishop, G.H.: Fiber groups in the optic nerves. *Am. J. Physiol.* **106**, 460–470 (1933)
5. Ungerleider, L.G., Mishkin, M.: Two cortical visual systems. In: Goodale, M.A., Mansfield, R.J.W. (eds.) *Analysis of Visual Behavior*, pp. 549–586. MIT Press, Cambridge (1982)
6. Livingstone, M.S., Hubel, D.H.: Segregation of form, color, movement and depth: anatomy, physiology and perception. *Science* **240**, 740–749 (1988)
7. Conway, B.R., et al.: Advances in color science: from retina to behavior. *J. Neurosci.* **30**(45), 14955–14963 (2010)
8. Kaplan, E.: The M, P and K pathways of the primate visual system revisited. In: Werner, J.S., Chalupa, L.S. (eds.) *The New Visual Neurosciences*. MIT Press, Cambridge (2012)
9. Stockman, A., Brainard, D.H.: Color vision mechanisms. In: Bass, M. (ed.) *OSA Handbook of Optics*, 3rd edn, pp. 11.1–11.104. McGraw-Hill, New York (2010)
10. Lee, B.B.: Visual pathways and psychophysical channels in the primate. *J. Physiol. Lond.* **589**(1), 41–47 (2011)
11. Solomon, S.G., Lennie, P.: The machinery of colour vision. *Nat. Rev. Neurosci.* **8**(4), 276–286 (2007)

Mass Pigmenting

► [Coloration, Textile](#)

Matting

► Compositing and Chroma Keying

Maxwell, James Clerk

Renzo Shamey

Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA



Maxwell, James Clerk,

Biography

James Clerk Maxwell was born on June 13, 1831, in Edinburgh, Scotland, to a family of comfortable means and was the only child of his parents. He is considered to be a pioneer in several fields of science. Maxwell contributed greatly to the field of optics and the study of color vision and helped

lay the foundations for practical color photography. Over a period of 17 years from 1855 to 1872, he published a series of papers concerning the perception of color, color blindness, and color theory [1]. He died in Cambridge at the age of 48 of abdominal cancer in 1879 [2].

Maxwell had a keen intellect and an unquenchable curiosity from childhood. Maxwell's formal schooling began unsuccessfully, and it is reported that he was treated harshly by his private tutor for being slow and disobedient [3]. He was then sent to the prestigious Edinburgh Academy to continue his education. At the age of 13, he won the school's mathematical medal as well as the first prize for English and poetry. At the age of 14, he wrote his first academic paper on *Oval Curves* which was presented, on his behalf, at the Royal Society of Edinburgh.

When 16 years old Maxwell was enrolled at the University of Edinburgh. There, among many other things, he learned about disk color mixture (Fig. 1) from one of his professors, James D. Forbes (1809–1868), who was working on a color classification system. During his undergraduate studies Maxwell also examined the properties of polarized light [4]. At age 18, he contributed two papers to the Transactions of the Royal Society of Edinburgh. One of these, *On the Equilibrium of Elastic Solids*, laid the foundation for an important discovery on the temporary double refraction produced in viscous liquids by shear stress [5]. In October 1850, Maxwell moved to the University of Cambridge's Trinity College where he graduated in 1854 with a degree in mathematics. He stayed at the Trinity College after graduation until 1856 when he accepted the position of professor of natural philosophy at Marischal College in Aberdeen [6]. In 1858 he married Katherine Mary Dewar who regularly assisted him in his experimental work. In 1860 he was granted the chair of natural philosophy at King's College in London. He stayed there until 1865, working primarily on electromagnetism. He and his wife returned to his inherited estate in Glenlair, Scotland, until he became the first Cavendish Professor of Physics at Cambridge University, where he stayed until his untimely death. Maxwell's famed equations on electro magnetism,



Maxwell, James Clerk, Fig. 1 Young Maxwell demonstrating one of his spinning color wheels

“Dynamical theory of the electromagnetic field” were delivered to Michael Faraday at his residence at the Royal Society in 1864, where as a young man Faraday himself had tried on several occasions to get his work noticed. While the two men were not close, they strongly respected each other. He is considered to be the “father of electromagnetics.” He is also the inventor of a thought experiment on thermodynamics, resulting in what is known as “Maxwell’s demon.”

Maxwell and Color

The nature of perception of color was one of Maxwell’s interests which had begun in Scotland. Using an improved version of a disk mixture apparatus, he demonstrated that white light results from a mixture of red, green, and blue light. His paper *Experiments on Colour* was a fundamental study of the color-mixing principles and was presented to the Royal Society of Edinburgh in 1855 [3]. In the following years he built a visual spectrometer in which he could mix and adjust spectral lights. The results of his own and his wife’s mixture data, published in 1860, demonstrated that only three primary lights are



Maxwell, James Clerk, Fig. 2 Reproduction of the Scottish tartan image using the photographic images and color filters employed by Maxwell and Sutton in their demonstration in 1861

required to match any spectral or composite light. In the same year he was awarded the Royal Society’s Rumford Medal “for his researches on the composition of colours and other optical papers.”

Based on his trichromatic theory, Maxwell proposed in the late 1850s a method for practical color photography. He suggested that if a scene is photographed three times separately but using red, green, and blue filters, and the resulting black and white transparent images are then superimposed on a screen using projectors equipped with similar filters, the result would be a full color reproduction of the image as seen by the human eye. Together with the photographer Thomas Sutton, he demonstrated this in 1861 at the Royal Society. As the image of the Scottish tartan shows, due to the fact that the filters were less than optimal, the outcome was less than perfect, and commercial color photography required another 45 years of invention (Fig. 2).

References

1. Johnson, K.: *Colour Vision*. University of St. Andrews (2012). Retrieved 8 Oct 2014
2. Harman, P.M.: *Oxford Dictionary of National Biography*, vol. 37. Oxford University Press, Oxford (2004). ISBN 0-19-861411-X

3. Tolstoy, I.: James Clerk Maxwell: A Biography. University of Chicago Press, Chicago (1982). ISBN 0-226-80787-8
4. Mahon, B.: The Man Who Changed Everything – The Life of James Clerk Maxwell. Wiley, Hoboken (2003). ISBN 0-470-86171-1
5. Timoshenko, S.: History of Strength of Materials. Courier Dover Publications, New York (1983). ISBN 978-0-486-61187-7
6. Glazebrook, R.T.: James Clerk Maxwell and Modern Physics. 811951455. OCLC 811951455 (1896)

Mayer, Tobias

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Tobias Mayer, born February 17, 1723, in Marbach, Germany, a mathematician, astronomer, cartographer, and physicist, was the only child of a fountain builder and his wife. Mayer was 10 years old when his father passed away, and he grew up in impoverished circumstances in the nearby city of Esslingen, spending some time in

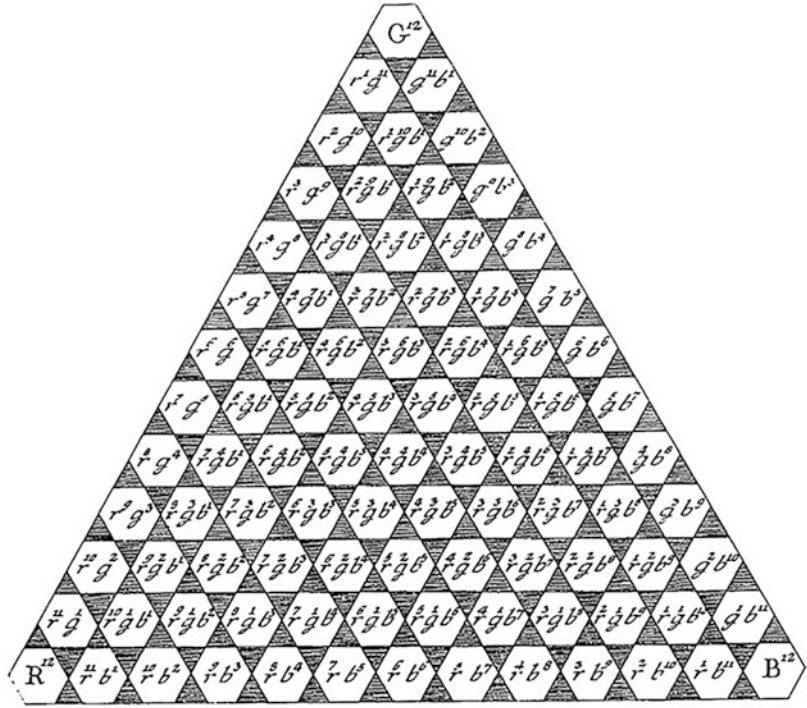
an orphanage. He taught himself mathematics and in his later teens earned some money teaching it. Mayer also showed early interest and capabilities in drawing and painting. He moved to Augsburg to obtain more training, working for the engraver and publisher J. A. Pfeffel. At 18 he wrote and published an elementary book on mathematics and at 22 a much more detailed work on the same subject [1]. His capabilities and the fact that he had designed an accurate map of Esslingen, later published, provided for him in 1746 an offer to join the firm of the cartographer Homann in Nürnberg, at the time perhaps the most important map publisher in Europe. In 1751 he married Maria Gnüge, and a year later their son Johann Tobias was born, in his later life a mathematician/physicist. In the same year Tobias Mayer's reputation as a scientist resulted in an offer of a professorship in economy and mathematics at the University of Göttingen, where he remained until his untimely death on February 20, 1762 due to a typhus infection. In 1754 he also became the supervisor of the Royal Observatory of Göttingen, built some 10 years earlier. His major scientific achievements are a data table of the moon and highly detailed drawings of the surface of the moon based on a new methodology to achieve high accuracy and the development of a new much more accurate methodology for the determination of longitude. The latter effort resulted in his widow receiving a 3,000-pound Sterling Award from the British Parliament. One of the craters of the moon is named "T Mayer" [2].

Major Accomplishments/Contributions

Mayer's combined interests in mathematics and painting resulted in an attempt to develop a mathematical model of the relationships between colors. Not accepting, on practical grounds, Newton's idea of seven main colors in the spectrum, Mayer decided to base his model on the painter's primaries yellow, red, and blue. He placed these at the corners of an equilateral triangle and divided the lines between them into 12 presumably perceptually equal parts which, he assumed, could be represented by corresponding weights of the

Mayer, Tobias,

Fig. 1 Central plane of
Mayer's double triangular
pyramid color solid



pigments used as primaries. He believed that these grades represented perceptually visible and comparable differences. The interior of the triangle is filled with colors that are mixtures of all three primaries. All of them are identified numerically, such as $r^8g^2b^2$ consisting of eight parts of the R primary and two parts each of the G and B primaries (Fig. 1) [3]. The 91 colors of the triangle of Fig. 1 can also be lightened or darkened, by addition of white or black pigments, again in twelve increments each. As these scales must end in single white or black, he gave the resulting solid the form of a double triangular pyramid, with a total of 819 defined colors [3]. In his German translation of the original Latin text, H. Lang defined Mayer's system as attempting to meet five then novel conditions: (1) All possible object colors are represented. (2) All colors are the result of mixtures of three primary colors. (3) Each color is represented by a triple number indicating the content of the three primaries. (4) The totality of all colors in the system is contained in a three-dimensional color

solid. (5) The differences between neighboring colors correspond to perceptual differences [4].

Mayer presented his ideas in 1758 in a public meeting of the Society of Sciences in Göttingen, a report of which was published in the newspaper *Göttingische Anzeigen von gelehrten Sachen* (Göttingen reports on scholarly matters) some three weeks later [5]. This report was the basis of J. H. Lambert's work on his triangular color pyramid [6]. Mayer had not done much experimental work to implement his system. As pointed out by Lambert, Mayer had not been aware of the varying coloristic strength of the three primaries, requiring consideration for the purpose of obtaining perceptual uniformity.

In 1958 Mayer also invented a new coloration method for prints. He proposed, and produced an example of, the coloration to consist of sections of wax containing different amounts of pigments to achieve different colors [7]. Multiple prints could then be produced from the wax collage. The idea proved to be too complex to be practical and was not pursued after his death.

References

1. Mayer, T.: Mathematischer Atlas. Pfeffel, Augsburg (1745)
2. Forbes, E.G.: The life and works of Tobias Mayer. *Q. J. R. Astron. Soc.* **8**, 227–251 (1967)
3. Mayer, T.: De affinitate colorum commentatio (On the relationship between colors), published posthumously in G. C. Lichtenberg, *Opera inedita Tobiae Mayeri* (Unpublished works of Tobias Mayer). Dieterich, Göttingen (1775). English translation: On the relationships between colors (A. Fiorentini, transl.), *Color Research and Application* **25**, 66–73 (2000)
4. Lang, H.: Tobias Mayers Abhandlung über die Verwandtschaft der Farbe. *Die Farbe* **28**, 1–34 (1980)
5. Göttingische Anzeigen von gelehrten Sachen, part 147, p. 1385-9 (1758)
6. Lambert, J. H.: Beschreibung einer mit dem Calauischen Wachse ausgemalten Farbenpyramide. Haude und Spener, Berlin (1772). English translation available at www.iscc.org
7. Günther, G.C.: Praktische Anweisung zur Pastellmalerei, pp. 129–132. Schneider, Nürnberg (1792)

Melanopsin Retinal Ganglion Cells

Robert Lucas

Faculty of Life Sciences, University of Manchester, Manchester, UK

Synonyms

Intrinsically photosensitive retinal ganglion cells; ipRGCs; mRGCs; Photosensitive retinal ganglion cells; pRGCs

Definition

A subset of retinal ganglion cells characterized by their expression of melanopsin and their ability to respond directly to light. Melanopsin is an opsin protein structurally and phylogenetically related to rod and cone opsins. It binds retinaldehyde as a chromophore and activates a G-protein signaling cascade upon photon absorption. As a result, melanopsin retinal ganglion cells are depolarized by light even when physically, pharmacologically,

or genetically isolated from rod and cone influences. This allows them to function as an independent origin of visual information. mRGCs are known to encode environmental brightness for such visual reflexes as circadian photoentrainment and regulating pupil size. They likely also make as yet ill-defined contributions to other visual processes including perceptual vision.

Discovery and Functions

Until around the turn of the millennium it was considered self-evident that all aspects of human vision could be ultimately attributed to the activity of rod and/or cone photoreceptors. With hindsight it is possible to find isolated examples of findings in several branches of vision science that questioned this assumption, but the concerted effort that led to the discovery of a third photoreceptor type came from researchers interested in a particular visual function – circadian entrainment [1].

Like almost all life on earth, humans have an internal circadian clock that oscillates with a period close to 24 h. In order to fulfill its function of coordinating physiology and behavior to the varying demands of the astronomical day this clock needs to be accurately set (or “entrained”) to local time. Entrainment relies upon a representation of the light:dark cycle provided by a dedicated retinal projection to the hypothalamic site of the circadian oscillator, the suprachiasmatic nuclei (or SCN). Studies of human subjects and rodent models of retinal dystrophy revealing that circadian entrainment was buffered against massive loss of rods and cones culminated in the demonstration that this function was retained even when these conventional photoreceptors were completely absent [2]. An explanation for this finding was provided by the discovery that the retinal ganglion cells projecting to the SCN are in fact directly photoreceptive [3].

Melanopsin is one of a number of opsin proteins known to be expressed outside of the eyes in nonmammalian vertebrates. It was initially discovered in the photosensitive dermal melanophores of *Xenopus laevis* [4]. Orthologues were

subsequently identified in mammalian genomes and found to be expressed in the retinal ganglion cells innervating the SCN. These ganglion cells were shown to lack their intrinsic photosensitivity in mice lacking melanopsin [5], while inclusion of melanopsin was shown to be sufficient to induce photosensitivity in a variety of non-light-responsive cell types [6–8].

Although initially discovered in the search to understand circadian entrainment, melanopsin retinal ganglion cells are now thought to contribute to a wide variety of visual responses [9]. mRGCs project not just to the SCN but also to other parts of the hypothalamus, the pretectum, the superior colliculus, the visual thalamus, and the nucleus of the optic tract [10, 11]. Their established functions extend to providing the pupil light reflex, regulating sleep/alertness, and suppressing pineal melatonin production. There is also evidence that they regulate mood and retinal development and influence retinal physiology by controlling local circadian clocks and the activity of dopaminergic amacrine cells.

At present, the nature and extent of mRGC contributions to perceptual vision is unclear. However, the appearance of mRGC fibers in projections to the dorsal lateral geniculate nucleus and established functional connections to other cell types in the retina via gap junctions and intraretinal axonal projections raise the possibility that mRGCs could have both direct and modulatory influences on perceptual vision. In primates, at least some ipRGCs receive antagonistic input from short- and long/medium-wavelength cones and may contribute to blue:yellow discrimination [12].

Anatomy and Physiology

The defining feature of mRGCs is their expression of melanopsin and consequent intrinsic photosensitivity. In rodents, at least five morphologically distinct classes of retinal ganglion cell (termed M1 to 5) meet this criterion [13]. The best-characterized of these is the M1 class, which has the highest melanopsin expression (and the most prominent intrinsic light response) and is responsible for both circadian entrainment and the pupil

light reflex. Other classes differ in their central projections, the nature and spatial extent of dendritic arborization, and their degree of melanopsin expression [13]. mRGCs described in primates thus far are more anatomically homogeneous, with all having large sparsely arborized dendritic fields located in either the inner or outer portions of the inner plexiform layer [12].

The dendrites of mRGCs are sites of melanopsin-driven phototransduction but also of synaptic inputs from bipolar and amacrine cells. As a result, in the intact retina, the spike-firing pattern of mRGCs is defined in part by melanopsin photoreception and in part by synaptic influences originating in signals from rods and cones. As mRGCs lack the membranous disks used by rods and cones to maximize photopigment concentration, they have very poor photon capture efficiency. As a result, melanopsin has an appreciable impact on mRGC firing only at rather high light intensities (several decades above the threshold for cone activation). Even at these light levels the rate of melanopsin photoisomerization is low, and its phototransduction cascade has to have high gain [14]. This is achieved by having a long activation time, with the consequence that the melanopsin light response has poor temporal resolution. Under some circumstances this phenomenon can be extreme with light-induced firing of mRGCs and melanopsin-driven pupil constriction lasting several tens of seconds after the light has been turned off. Under more natural conditions, these sensory characteristics imply that melanopsin provides a low spatiotemporal resolution signal of environmental brightness, with mRGCs relying upon input from rods and cones to encode lower light levels and to track higher frequency changes in light intensity [15].

Melanopsin, like all animal opsins, is a member of the G-protein coupled receptor superfamily of proteins. The molecular identity of the critical elements of its phototransduction cascade is unknown. However, the available evidence indicates that melanopsin couples to a G-protein of the Gq/11 class upon light exposure, which results in activation of phospholipase C and subsequent membrane depolarization by opening TRP channels [16].

Spectral Sensitivity

Measuring the spectral absorbance properties of melanopsin directly has proved challenging. It is not found in sufficient concentrations to allow direct *in vivo* microspectrophotometry. While it has been possible to purify melanopsin from harvested retinas and from heterologous cell-based expression systems, the quality of these preparations falls far short of that achieved with rod or cone opsins. As a result, a precise description of melanopsin's spectral absorbance in the dark state, and following light absorption, remains elusive. Nonetheless, action spectra for melanopsin-driven activation of mRGC activation and downstream responses in a variety of mammalian species are consistent with the hypothesis that in the dark state melanopsin's spectral sensitivity can be well described by the standard opsin: vitamin A1 template peaking around 480 nm [15]. This puts it close to the spectral sensitivity of human rod opsin but far from that of the cone photoreceptors that are conventionally considered to dominate vision at moderate and high light levels. Importantly, its spectral sensitivity is also significantly short wavelength shifted compared to $V(\lambda)$ (the spectral efficiency function of photopic vision), which is used in quantifying (il) luminance. This implies that spectrally distinct natural and artificial light sources might have quite different effective intensity for melanopsin even when matched for illuminance. There is thus a need to update measurement metrics and to reconsider the spectral content of artificial sources to take account of melanopsin's discovery [15].

References

1. Bailes, H.J., Lucas, R.J.: Melanopsin and inner retinal photoreception. *Cell. Mol. Life Sci.* **67**, 99–111 (2010)
2. Foster, R.G., Hankins, M.W.: Non-rod, non-cone photoreception in the vertebrates. *Prog. Retin. Eye Res.* **21**, 507–527 (2002)
3. Berson, D.M., Dunn, F.A., Takao, M.: Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002)
4. Provencio, I., Jiang, G., De Grip, W.J., Hayes, W.P., Rollag, M.D.: Melanopsin: an opsin in melanophores,

brain, and eye. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 340–345 (1998)

5. Lucas, R.J., et al.: Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003)
6. Melyan, Z., Tarttelin, E.E., Bellingham, J., Lucas, R. J., Hankins, M.W.: Addition of human melanopsin renders mammalian cells photoresponsive. *Nature* **433**, 741–745 (2005)
7. Qiu, X., et al.: Induction of photosensitivity by heterologous expression of melanopsin. *Nature* **433**, 745–749 (2005)
8. Panda, S., et al.: Illumination of the melanopsin signaling pathway. *Science* **307**, 600–604 (2005)
9. Schmidt, T.M., et al.: Melanopsin-positive intrinsically photosensitive retinal ganglion cells: from form to function. *J. Neurosci.* **31**, 16094–16101 (2011)
10. Ecker, J.L., et al.: Melanopsin-expressing retinal ganglion-cell photoreceptors: cellular diversity and role in pattern vision. *Neuron* **67**, 49–60 (2010)
11. Brown, T.M., et al.: Melanopsin contributions to irradiance coding in the thalamo-cortical visual system. *PLoS Biol.* **8**, e1000558 (2010)
12. Dacey, D.M., et al.: Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. *Nature* **433**, 749–754 (2005)
13. Sand, A., Schmidt, T.M., Kofuji, P.: Diverse types of ganglion cell photoreceptors in the mammalian retina. *Prog. Retin. Eye Res.* **31**, 287–302 (2012)
14. Do, M.T., Yau, K.W.: Intrinsically photosensitive retinal ganglion cells. *Physiol. Rev.* **90**, 1547–1581 (2011)
15. Lucas, R.J., et al.: Measuring and using light in the melanopsin age. *Trends Neurosci.* **37**, 1–9 (2014)
16. Hughes, S., Hankins, M.W., Foster, R.G., Peirson, S. N.: Melanopsin phototransduction: slowly emerging from the dark. *Prog. Brain Res.* **199**, 19–40 (2012)

Memory Color

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

Synonyms

Canonical color; Diagnostic color

Definition

A memory color is the typical color of an object that an observer acquires through their experience with

that object. For example, most people know that a ripe banana is typically yellow; this knowledge about the typical color constitutes a memory color.

Conceptual Clarifications

A memory color is an observer's knowledge of a typical object color. The typicality of the memorized color implies that the observer considers the memory color to be representative or "canonical" for the range of colors, in which the respective object occurs. In this way, it determines the observer's expectation about an object's color based on her or his prior knowledge. For example, an observer that is only familiar with the yellowish color shades of the ripe common bananas (i.e. Cavendish bananas) would be surprised to encounter a Red Dacca banana because it is not in line with this observer memory color.

The concept of memory color is historically closely related to the idea of a memory color effect on color appearance. Moreover, it partly overlaps with the notion of "color diagnosticity". Finally, it ultimately refers to the same phenomenon as "canonical color," though with different emphases and connotations. This section will clarify these conceptual relationships.

Memory Color Effect

The notion of a "memory color" was coined by Ewald Hering in 1878 [1]. He associated it with the idea that knowledge about typical colors affects the perception of the actual color of given objects (e.g., [2]). However, the fact that an observer knows about the typical colors of certain objects (i.e. memory colors) does not necessarily imply that this knowledge influences the way they see the actual object colors. For this reason, it is sensible to disentangle the concept of a memory color from the idea of a memory color effect on color appearance. Hence not all research on memory color is about the memory color effect.

Color Diagnosticity

Like memory colors, color diagnosticity refers to the association between an object and its typical color. An object is color diagnostic when it only

occurs in a particular range of color shades, and this range of shades may be summarized by a typical color as the best example of this object's colors (objective color diagnosticity). For example, ripe bananas occur in a limited range of yellowish shapes, and this range may be summarized by a yellow that is typical for ripe bananas. In contrast, cars are not color diagnostic (or color neutral) in an objective sense because they occur in arbitrary colors. However, in order for an object to be color diagnostic for a particular observer, the observer needs to know what the range of color shades are, and what the most typical colors are (subjective color diagnosticity). This knowledge consists of memory colors, and hence, the memory color for an object constitutes its subjective color diagnosticity. In most cases, it may be assumed that subjective color diagnosticity converges towards objective color diagnosticity through experience, so that memory colors approach the most common or typical color of an object. However, the objective co-occurrences of objects and colors and the subjective knowledge about them are not a priori the same.

Canonical and Typical Color

While the term "memory color" has been employed in psychophysics, the alternative term "canonical color" is primarily used in developmental research [3, 4]. Here, the research questions are more focused on the developmental order than on the subjective nature of memory colors. Hence, in these studies the term "canonical color" is often used in a way that abstracts from the difference between subjective and objective color diagnosticity and refers directly to the color-object association determined by the experimenter. However, since these studies concern the object-color association of the observers and not the regularities in the environment, the term "canonical color" is ultimately interchangeable with the one of "memory color". While "memory color" emphasizes the psychological, subjective aspect of perception and memorization, "canonical color" highlights the convergence to a shared knowledge that reflects the statistics of the environment.

The notion of “typical” or “prototypical object color” seems to merely refer to the physical occurrence of object colors. However, the idea of typicality requires – at least implicitly – an assumption about a representation of the range of actually occurring object colors through a typical color. For example, the typical color may be represented by the mean of all occurring colors, by the color that occurs in most cases (mode), by the color that is “purest” (i.e., maximally saturated), or simply by a whole region of colors (exemplar-based typicality). So, the mere idea of typicality refers to a real or virtual observer (i.e., a human being or a simulating algorithm) that integrates the single exemplars encountered hitherto into a typical representation. For this reason, it is difficult to dissociate the notion of typical object color from the integration process done by the real or virtual observer and hence from their memory colors.

Methods to Determine Memory Colors

Since memory colors imply a relationship between objects and colors, the assessment of memory colors needs to take three constitutive determinants into consideration. Firstly, memory colors require familiarity with the respective objects. Secondly, they require an idea about the range of colors that are typical. Finally, memory colors are determined by the link between the object and its typical color, which constitutes the color diagnosticity of the object.

Object Recognition

In order to measure an observer’s memory colors for an object, the observer must be able to recognize the object. This does not only require that the observer is familiar with the respective kind of object; it also depends on how well the exemplar used for the measurement can be identified as the respective kind of object. In particular, many measurements of memory colors use two-dimensional images, and the question arises whether the pictorial representations are recognizable and representative. For example, outline shapes of fruits, such as oranges or carrots, are barely recognizable even

though people are highly familiar with these objects (cf. [2]). As a consequence, the measurement of memory colors may be affected by particularities of the image sampling. For this reason, many studies use image databases that also provide familiarity indices for the images, such as the Snodgrass line drawings (e.g., [5]).

Color Typicality

In order to assess the range and typicality of colors associated with an object, color adjustment or rating techniques may be used. In the former, observers adjust the color of the objects to the typical color, and the distribution of these adjustments informs about the nature and the precision of the typical color (cf. e.g., [6]). In rating tasks observer evaluate how well a particular color matches the typical color of an object [7].

Color Diagnosticity

For most studies, the most important measurement is the assessment of the strength of the object-color association, i.e. the color diagnosticity. A fundamental distinction can be made between techniques that refer to conceptual measures and those that measure color diagnosticity perceptually. A typical conceptual measure involves verbal elicitation. Observers are asked to name the most important object features, and color diagnosticity is assessed by the priority with which color is named among these features. Other techniques require observers to rate how important color is to describe the object. Some studies have also used indices such as the Nelson word association norms to assess color diagnosticity [5]. The problem of these techniques may consist of how well these conceptual measures may be transferred to the particular instances used as stimuli. For example, the verbal name of red cabbage may well be associated with the color red. Nevertheless, a grayscale (i.e., achromatic) image of red cabbage is not unambiguous enough to refer to its typical color.

A perceptual approach consists of showing the respective objects (mostly their images) without color and asking the observer to indicate their typical color. The accuracy and speed of the answer gives insight into the strength of the

object-color association. In particular, the speed may also reflect the automaticity of the memory color knowledge. Moreover, this technique also ascertains the validity of the measures for the particular stimuli. This, however, is also the drawback of this technique, since measurements cannot directly be generalized to all instances of the respective objects [2].

Characteristics of Memory Colors

Memory colors may be characterized by how typicality is represented in memory and when and how the first object-color associations develop during childhood. Finally, memory colors have also been used as references to measure color constancy and the subjective appreciation of the quality of the illuminant.

Typicality

Memory colors consist of the most typical color that is associated with the respective object and a region of tolerance, within which concrete colors may still be accepted as instances of the memory color. The most typical memory colors tend to exaggerate the actual hue of the color-diagnostic objects. This implies that memory colors are oversaturated for objects that typically have chromatic colors, such as a banana, and undersaturated for objects that are typically achromatic, such as a cauliflower. Moreover, the regions of tolerance are wider for variations in saturation than in hue, implying that memory colors are more clearly defined in hue than in saturation [7].

Development

Preschool children (up to 5 years) may still have difficulties, when they are asked to identify the appropriate color for very common color-diagnostic objects (e.g., [3]). Some approaches assessed object-color knowledge through their effects on perceptual performance (also see section “[The Memory Color Effect](#)”). Interference effects that reveal memory color knowledge have been found at the age of 3.5 years with an object Stroop paradigm [8], and for 2–3 year old

children in a free-looking paradigm (e.g., [9]). A study that used a simple preferential-looking procedure has even found some evidence that infants possess memory colors as early as at the age of 6 months [4].

Constancy

Memory colors also provide a reference for perceived colors. For this reason, memory colors may be used to measure color constancy under different illuminations; however, these measurements are coarse due to the overall variability of memory color estimations [6]. At the same time, there is some evidence that observers prefer illuminations that produce the colors of color-diagnostic objects as similar as possible to the observer’s memory colors [10].

Effects of Memory Colors

Through memory colors, object and color identification interact in several ways. These interaction effects will be summarized below; but see [2, p. 15] for more references.

Object Recognition

Firstly, there is ample evidence that memory colors promote object and scene recognition. In particular, color-diagnostic objects and scenes are identified more easily when seen in their typical colors (for a review see [11]).

Priming and Cueing Effects

Secondly, color-diagnostic objects automatically elicit the respective memory color and vice versa. In particular, so-called object Stroop tasks have shown that colors are identified more rapidly when seen on the respective color-diagnostic object (e.g., [12]). In turn, priming paradigms revealed that colors facilitate the identification of the labels for color-diagnostic objects (e.g., [13]). Finally, mentioning color-diagnostic objects verbally directs gaze automatically towards their memory colors. Similar to cueing, this implies an orientation of overt attention towards memory colors (e.g., [5]).

Color Memory

Moreover, there is also some evidence that memory colors influence color memory. When memorizing the actual color of color-diagnostic objects, this memorized color seems to be shifted towards the object's memory color [14].

Color Categorization

Memory colors also serve as references in color naming. There is evidence that the category membership of ambiguous colors close to the category boundary may be changed by showing the colors on different color-diagnostic objects. For example, a color that is halfway between orange and yellow will be classified rather as orange (or yellow) depending on whether it has been seen on a carrot (or a banana, respectively) (e.g., [15, 16]).

Color Constancy

Finally, memory colors support color constancy. When the colors in a scene change due to a change of the illumination, knowledge about the original color helps to estimate the color change that is due to the illumination. However, according to available evidence the contribution of memory colors to color constancy is small compared to the contribution of low-level mechanisms such as adaptation (e.g., [17]; but see [18]).

The Memory Color Effect

According to the classical idea of a “memory color effect” memory colors directly influence the appearance of the actual color of the objects. Familiarity with the object-color association implies an expectation about the color of a concrete exemplar of an object, and this expectation shapes the perception of its actual color. In this way, memory color effects indicate that color appearance depends on the object, on which the color is shown, as well as on the experience and the prior knowledge of the observer. For example, an observer may know that bananas are yellow. Does this knowledge influence how this observer sees the actual color of a banana? Or can she or he perceive and appreciate the actual color independently of their prior knowledge?

History

Although the precise idea about memory color effects was introduced by Hering [1], the idea that prior experiences with objects in the environment influences the way an observer perceives their colors has already been formulated by Hermann von Helmholtz in 1867. During the twentieth century several studies pursued Hering's original idea that memory colors affect the perception of the objects' actual colors. Most of these investigations have shown that observers overestimate the saturation of the object colors when the hue corresponds to the typical color of the objects. However, there were also other studies that could not replicate this effect. Moreover, these earlier studies could not unambiguously show that memory colors really affect perception rather than biasing memory retrieval and judgment (for further references, see [2]).

Recent Developments

More recent studies used a broad range of methods and techniques to approach this question. They could confirm that observers overestimate the proportion of the objects' typical hue, when judging the colors of color-diagnostic objects. For example, observers would judge a banana to be more yellow than it actually is (e.g., [15]; for further references, see [2]).

One of these approaches could even show that a color-diagnostic object may automatically bring about the impression of its typical color [19]. In this case, a banana should still appear to be yellow even if it is completely grey. In this approach, participants had to adjust the colors of the objects so that they look grey to them. Now, if the object alone induces the impression of its typical color then the observers have to counteract this impression in order to see the object so that it looks subjectively grey to them. And indeed, the participants of these studies shifted the color adjustment towards the color opposite to the typical color. For example, they adjusted the banana slightly towards blue because blue is the opponent color to yellow. They did this, even though they had only to match the color of the object to the color of the background. This implies that they really saw

the banana as yellow or at least slightly yellowish when it was actually grey.

State of the Art

Follow-up studies that also used the achromatic adjustment method showed that images with less perceptual information, such as outline shapes, yield weaker memory color effects. This explains why some of the classical studies that used outline shapes yielded inconsistent findings. Moreover, these follow-up studies showed that the memory color effect is robust to dramatic changes of the illumination [6].

All of these studies used fruits and vegetables as stimuli. The colors of these objects do not cover the whole color space and the chromatic distributions of these objects are all based on natural surfaces. For example, there are no fruits and vegetables that occur in a very colorful blue or purple. Another follow-up study used artificial, man-made objects, such as a Smurf cartoon character, a Nivea lotion tin, and a German mailbox. The results of this study revealed memory color effects for these artificial objects that had all kinds of chromatic distributions. This study also showed that the memory color effect occurred most strongly along the daylight axes, along which observers are least certain about their color judgment [2].

Outlook: The Role of Memory Colors

In everyday life, people use colors to identify objects in their environment. However, the link between colors and the surface properties of objects is not as simple and direct as everyday life experience might suggest. For example, the light reaching the retina intermingles contributions from illumination and surface properties. Memory colors reinforce the association between colors and particular objects, namely, color-diagnostic objects. In this way, memory colors may increase the reliability of object identification and hence support the functional role of color as an object attribute. For future research, the question arises of whether memory colors play a constitutive role in color appearance and

communication, leading to a functional adaptation of the observer's color identification to their physical and social environment.

Cross-References

- [Chromatic Contrast Sensitivity](#)
- [Color Vision, Opponent Theory](#)
- [Fechner's Colors and Behnam's Top](#)
- [Stroop Effect](#)
- [Subjective Colors](#)

References

1. Hering, E.: Outlines of a theory of the light sense. Harvard University Press, Cambridge, MA, Translated by Hurvich, L. and Jameson, D. (1878/1964)
2. Witzel, C., Valkova, H., Hansen, T., Gegenfurtner, K. R.: Object knowledge modulates colour appearance. *Iperception* **2**(1), 13–49 (2011)
3. Gleason, T.R., Fiske, K.E., Chan, R.K.: The verbal nature of representations of the canonical colors of objects. *Cogn. Dev.* **19**(1), 1–14 (2004)
4. Kimura, A., Wada, Y., Yang, J., Otsuka, Y., Dan, I., Masuda, T., Kanazawa, S., Yamaguchi, M.K.: Infants' recognition of objects using canonical color. *J. Exp. Child Psychol.* **105**(3), 256–263 (2010)
5. Huettig, F., Altmann, G.T.: Looking at anything that is green when hearing “frog”: how object surface colour and stored object colour knowledge influence language-mediated overt attention. *Q. J. Exp. Psychol.* **64**(1), 122–145 (2011)
6. Olkkonen, M., Hansen, T., Gegenfurtner, K.R.: Color appearance of familiar objects: effects of object shape, texture, and illumination changes. *J. Vis.* **8**(5), 1–16 (2008)
7. Smet, K.A.G., Ryckaert, W.R., Pointer, M.R., Deconinck, G., Hanselaer, P.: Colour appearance rating of familiar real objects. *Color Res. Appl.* **36**(3), 192–200 (2010)
8. Prevor, M.B., Diamond, A.: Color-object interference in young children: a stroop effect in children 3(1/2)–6(1/2) years old. *Cogn. Dev.* **20**(2), 256–278 (2005)
9. Johnson, E.K., McQueen, J.M., Huettig, F.: Toddlers' language-mediated visual search: they need not have the words for it. *Q. J. Exp. Psychol.* **64**(9), 1672–1682 (2011)
10. Smet, K.A.G., Ryckaert, W.R., Pointer, M.R., Deconinck, G., Hanselaer, P.: Memory colours and colour quality evaluation of conventional and solid-state lamps. *Opt. Express* **18**(25), 26229–26244 (2010)
11. Bramao, I., Reis, A., Petersson, K.M., Faisca, L.: The role of color information on object recognition: a review and meta-analysis. *Acta Psychol. (Amst)* **138**(1), 244–253 (2011)

12. Naor-Raz, G., Tarr, M.J., Kersten, D.: Is color an intrinsic property of object representation? *Perception* **32**(6), 667–680 (2003)
13. Nijboer, T.C., van Zandvoort, M.J., de Haan, E.H.: Seeing red primes tomato: evidence for comparable priming from colour and colour name primes to semantically related word targets. *Cogn. Process.* **7**(4), 269–274 (2006)
14. Van Gulick, A.E., Tarr, M.J.: Is object color memory categorical? *J. Vis.* **10**(7), 407 (2010)
15. Ling, Y., Allen-Clarke, L., Vurro, M., Hurlbert, A.C.: The effect of object familiarity and changing illumination on colour categorization. *Percept.* **37** ECVF 2008 Abstract Supplement, 149 (2008)
16. Mitterer, H., de Ruiter, J.P.: Recalibrating color categories using world knowledge. *Psychol. Sci.* **19**(7), 629–634 (2008)
17. Granzier, J., Gegenfurtner, K.: Effects of memory colour on colour constancy for unknown coloured objects. *Iperception* **3**(3), 190–215 (2012)
18. Kanematsu, E., Brainard, D.H.: No measured effect of a familiar contextual object on color constancy. *Color Research & Application*, doi:10.1002/col.21805 (2013)
19. Hansen, T., Olkkonen, M., Walter, S., Gegenfurtner, K.R.: Memory modulates color appearance. *Nat. Neurosci.* **9**(11), 1367–1368 (2006)

Mesoamerican Color Survey Digital Archive

Kimberly A. Jameson¹, Nathan A. Benjamin², Stephanie M. Chang², Pruthi S. Deshpande³, Sergio Gago⁵, Ian G. Harris⁴, Yang Jiao⁴ and Sean Tauber¹

¹Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA, USA

²Calit2, Computer Science, University of California, Irvine, Irvine, CA, USA

³Cognitive Sciences, University of California, Irvine, CA, USA

⁴Computer Science, University of California, Irvine, CA, USA

⁵Calit2, School of Engineering, University of California, Irvine, Irvine, CA, USA

Definition

The Mesoamerican Color Survey (MCS) collected color-naming and categorization data from approximately 900 speakers from each of

116 indigenous languages from regions in Mesoamerica or Central America. Analyses of these data were originally reported by Dr. Robert E. MacLaury, principal investigator of the survey. The MCS data exist as a public-access color categorization and naming digital archive, in conjunction with other color categorization data collected by MacLaury, and are available at <http://colcat.calit2.uci.edu/>.

Introduction

Human categorization behavior is widely studied across the behavioral sciences. It underlies many cognitive functions, including concept formation, decision making, learning, and communication. Color appearance, similar to other natural categorization domains, has distinctive features or properties that vary along continuous dimensions. Semantic color categories, their formation, their best exemplars and boundaries, and the influence of these on human behavior have been the topics of much empirical study, receiving considerable attention from anthropologist, linguists, cognitive scientists, and psychologists. The general aim of such research is to understand human color categorization and how it is cognitively and culturally represented across languages. For the case of color categories, the literature suggests that there are “universal trends” in human color representations with the category best exemplars being predictable across languages [1–7], and there are culturally specific color categorization influences [8–15], as well as human evidence [16–22] and simulated color category evolution evidence [23–27], suggesting combinations of universal, cultural, and pragmatic influences on color categorization and naming behaviors.

The Mesoamerican Color Survey (MCS) is one of the two existing databases (the other being the World Color Survey or WCS; see “► [World Color Survey](#)”) which directly investigated, on a large scale, color naming and categorization across many linguistic societies. The MCS and the WCS employ nearly identical standardized procedures for evaluating large numbers of color stimuli, languages, and informants. The

Mesoamerican Color Survey was conducted by Robert E. MacLaury during the years 1978–1981. MCS data were collected by MacLaury himself, or by research associates and colleagues in Mesoamerica, whom MacLaury trained and directed. The MCS data represent interviews with 900 speakers of some 116 Mesoamerican languages. It is estimated that more than 100 indigenous languages are spoken in Mexico and Central America. Like most languages each MCS language has a color lexicon that partitions environmental color appearance stimuli according to a pattern that is specifically relevant to a given language’s speakers. But every local MCS system of color categorization also shares characteristics with systems observed for other Mesoamerican languages and with those of languages elsewhere in the world. MacLaury published analyses of the MCS data in the context of his *Vantage Theory* modeling approach ([28–30], see ► [Vantage Theory of Color](#)).

MCS and Theories of Color Categorization

Humans can discriminate on the order of 10^6 different colors, many more colors than any individual can name reliably. Most of the 7,000 living languages have some form of a color lexicon. Various theories in the literature have sought to account for empirical findings on how each language represents color concepts and how such representations evolve, changing over time, and successively form hierarchical linguistic representations of color appearance similarity. In 1969 Brent Berlin and Paul Kay were among the first to systematically empirically examine how different linguistic societies represented color and theorize how color lexicons might similarly evolve ([1], see ► [Berlin and Kay Theory](#)). In the data they collected from 20 unrelated language families, they found (a) that there exist universal cross-linguistic constraints on color naming and (b) that basic color terminology systems tend to develop in a partially fixed order.

Berlin and Kay made two conjectures about the evolution of color lexicons: (1) There is a limited

set of basic color terms (abbreviated as BCTs) in most languages, which are distinct from other color terms that an individual might use to name colors. (2) Color lexicons evolve from simple to complex, along highly constrained paths, starting from two BCTs corresponding to warm-or-light and dark-or-cool categories in the most reduced lexicons and extending to 11 BCTs as seen in English and other languages spoken in industrialized societies. The Berlin and Kay theory is the basis for a mainstream approach to color categorization theory, along with its subsequent formulations using the World Color Survey (WCS) data ([2–4, 7], see ► [World Color Survey](#)”).

While much empirical support exists for the Berlin and Kay/WCS theoretical approach, alternative explanations have also been suggested. Researchers directly involved in the WCS project, who collaborated with Berlin and Kay, have suggested that some linguistic societies might follow alternative patterns of color lexicalization, suggesting possible deviations from a hue-based model of color category universality that was central to the Berlin and Kay and WCS research program. The suggestions raised, for example, questioned whether a universal BCTs conjecture positing “a total universal inventory of exactly 11 basic color categories exists from which the 11 or fewer basic color terms of any language are always drawn” ([1], p. 2) was the best model to describe the extensive amount of data accumulated by the Berlin and Kay and WCS efforts. Robert E. MacLaury was among those who sought to explore alternative explanations, which was the impetus for his direction of the MCS project [11, 28–30]. In the 1970s MacLaury was an anthropologist working with Brent Berlin and Paul Kay, conducting fieldwork in Oaxaca, Mexico, and later working on the World Color Survey, and the Mesoamerican Color Survey was the basis for his PhD dissertation obtained from the University of California, Berkeley in 1986. In 1997 MacLaury successfully used the MCS data as the basis for an alternative color-naming and categorization theory, called *Vantage Theory* [28]. That approach presents a model of color categorization at the lexeme level where color categories are constructed as vantages (see ► [Vantage Theory](#)

of Color”). In the context of this approach, MacLaury’s book provides an extensive inventory of the organization and semantics of color categorization in Mesoamerica.

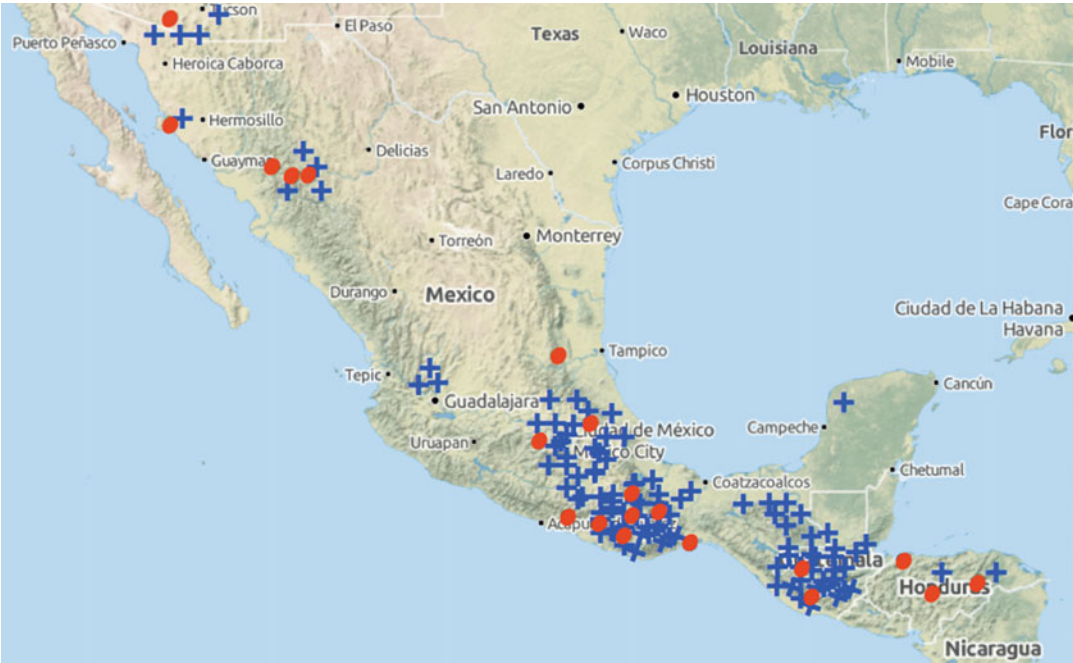
The MCS: History, Methodology, and Data

Historical Background

As noted earlier, the Mesoamerican Color Survey (MCS) includes 116 surveyed societies, each with a different Mesoamerican language or dialect, assessing approximately 900 native-language speakers from indigenous populations in Mesoamerica. The MCS employed color stimuli from the WCS stimulus palette shown elsewhere in this volume (see “► [World Color Survey](#)”). Figure 1 above visually illustrates how the MCS substantially extends the geography of areas surveyed and thus extends beyond the wealth of information provided by the widely cited and valued WCS. It

does so by replicating an estimated 20 languages found in the WCS, providing observations for 96 additional languages using the WCS’ standardized data collection procedures. Such an extensive survey of indigenous languages is likely no longer possible in 2015, due to dramatic increases in exposure over the last 25 years of native-language monolingual speakers to English-language broadcast media and entertainment.

The original MCS data was handwritten on paper, and while MacLaury published analysis and theory of MCS data (notably, [28]), thousands of pages of raw MCS data have never been publicly available and have not been subjected to the kinds of highly informative methodological approaches that have more recently been used to reveal the insights learned from the WCS data (e.g., [3–6, 20, 21, 31, 32]). MacLaury focused on MCS language groups across Northwest Mesoamerica for a particular reason: to investigate what he conceived as an alternative theory of color category systems that emphasized



Mesoamerican Color Survey Digital Archive, Fig. 1 Map depicting a portion of the geographic region surveyed by the MCS from the border of the United States to Nicaragua. Approximate locations of 116 MCS languages sampled are indicated as *blue crosses*, and 20 *red*

dots show WCS languages that duplicate some of the MCS languages surveyed. Of note are the dense samples of linguistically related MCS languages found in the areas of Oaxaca (37 languages), Guatemala (30 languages), and Mexico City (33 languages)

“brightness”-based categories in naming – as an alternative to historically prominent theories emphasizing hue-based categories in naming – and to test hypotheses and give interpretations of the data grounded in a relativistic “framing” idea, which MacLaury called Vantage Theory.

MacLaury employed the WCS methodology from 1978 up until approximately 2004. Subsequent to 2004 the archive containing the MCS was inactive, unlike that of the WCS, with its raw data inaccessible to researchers while it was preserved in private storage by the MacLaury estate. Beginning in 2010 funding to initiate preservation and digitization of the archive was obtained. The initial preservation of the MCS archive, including its estimated 70 additional worldwide languages, came in the form of funding to purchase the archive from MacLaury’s estate, supported by the University of California Pacific Rim Research Program (award to K.A. Jameson, PI), and was augmented by a 2014 National Science Foundation funding (#SMA-1416907, K.A. Jameson, PI) that in part aimed to establish a public-access digital archive of transcribed MacLaury data and construct a user-friendly Wiki to generally facilitate collaborative color categorization research.

MCS Methodologies

Using individual samples and a color stimulus array shown in the WCS stimulus palette (see Fig. 1, Kay and Cook’s ECST entry), the MCS included three independent tasks done by every native-language speaker surveyed. These consist of (1) a “naming task” that involved the free naming of 330 color samples presented in a fixed random order; (2) a “focus task,” identification of named category best exemplars or category “foci”; and (3) a “mapping task” that involved the demarcation of named category boundaries, producing a “boundary map.” The latter “mapping task” was used in Berlin and Kay’s original work, but for simplicity was omitted from the WCS. In the MCS, naming and focus data were collected for 900 individuals, and the mapping task method was reintroduced and assessed on an estimated 365 individuals. Data collection procedures used are detailed in works of MacLaury

[28, 29] and are also similarly reviewed in WCS materials [2].

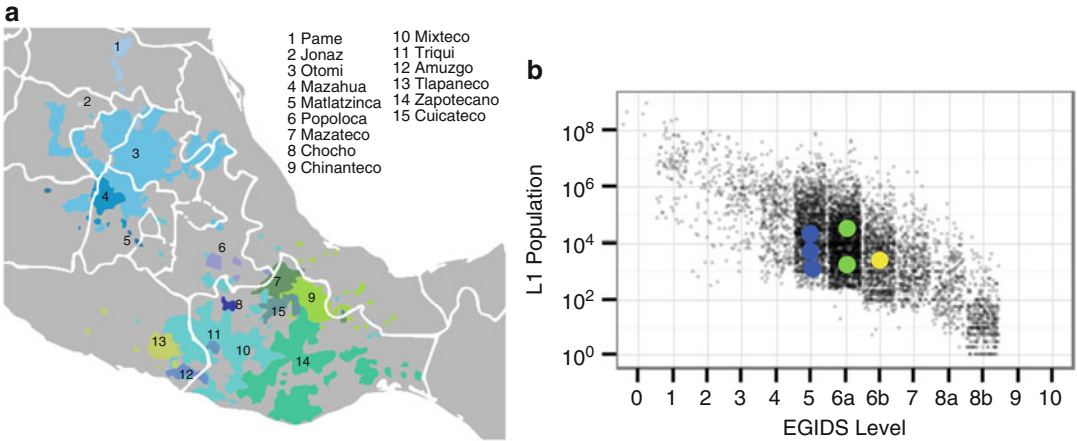
These MCS methods, along with the selection of languages it assessed, permit an unprecedented study of concept formation and linguistic representation, because of the following: (1) Many MCS languages are phylogenetically related to varying degrees. (2) Many MCS languages emphasize different perceptual dimensions (e.g., either primarily brightness or hue) and permit new analyses of dimensional salience. (3) Most MCS languages that satisfy (1) and (2) are related spatially, or geographically, and are thus known or expected candidates for linguistic borrowing, drift, and bilingual influences.

MCS Data

The research scope of the MCS data is substantial because it has features that can be exploited for addressing a range of heretofore unexplored color categorization research questions concerning how color categorization systems evolve, how semantic meaning drifts, and how meaning may be impacted by the influences of neighboring systems.

For example, Fig. 2’s representation of one language family (Otomanguean, Glottolog: otom 1299) shows one example that is well represented in the MCS, and, like that shown in panel (a), the MCS generally contains clusters of phylogenetically related languages, usually from contiguous geographic regions. Thus, for this one language family, as Fig. 2a illustrates, the MCS provides an estimated 369 surveyed participants, for which most subdivisions shown in panel (a) were assessed. Moreover, panel (b) illustrates that for some subdivisions, using Chinantec as an example, one may find several variations: That is, the MCS has six variations that cover three different levels of Chinantec language development/enderment.

In addition, the MCS also has many languages that emphasize different dimensions of color appearance space compared to the hue-based systems emphasized by languages typically investigated in the literature. For each language surveyed, the actual MCS datasets typically include detailed investigator notes or



Mesoamerican Color Survey Digital Archive, Fig. 2 Example of the typical level of representation for one language family (Otomanguean) found in the MCS. Panel (a) shows a map of Oaxaca and Mexico City areas shown above in Fig. 1. Numbered color regions represent Otomanguean family language branches listed in inset (Retrieved on 12/02/2014 from “https://commons.wikimedia.org/w/index.php?title=File:Otomanguean_Languages.png&oldid=89424714”. Image: wikipedia.org. https://commons.wikimedia.org/wiki/File:Otomanguean_Languages.png) exemplifying spatial density of related languages that is typical of the MCS. Panel (b) shows one of the family branches found in (a), namely, #9 Chinantec. Panel

(b) shows a level of language diversity and complexity typical of the MCS. In this example, six forms of Chinantec data are illustrated in the MCS dataset. Panel (b)’s six color dots locate MCS Chinantec languages within the cloud of all living languages (i.e., *small dots*) in relation to a language’s population (*vertical axis*) and its level of development or endangerment (*horizontal axis*). *Blue dots* are Developing (EGIDS 5) showing examples of Chinantec in vigorous use, *green dots* are Vigorous (EGIDS 6a) – unstandardized and in vigorous use among all generations. *Yellow dots* are In trouble (EGIDS 6b-7) (Chinantec living language data adapted and cited with permission from www.ethnologue.com [33]. Used by permission, © SIL International)

ethnographies, as well as the datasheets on the three tasks. Variations on information contained and common formats of MCS datasheets are now described and illustrated:

1. **Naming task:** Informants were asked to name 330 loose color chips in a fixed random order. While investigators conducting the survey tended to employ two datasheet formats for collecting naming data, individual variations do exist in the raw datasheets. Figure 3 shows two common raw datasheet formats used to collect naming task data. Image scans of raw MCS datasheets, and for a currently limited number of cases their transcribed digital data files, are provided on the archive website (detailed below).
Figure 3a, b show two common examples of MCS naming task datasheet formats as raw pdf image scans of the data. Panel (a) illustrates only one page (samples 1–81) surveyed from informant #1 (assessed in Guarijio language)

- showing a “list format” in which the investigator sequentially listed a participant’s responses to 330 samples evaluated in the fixed random order in which they were assessed. Panel (b) shows a second common format in which the investigator reports the participant’s responses to 330 samples, evaluated in the same fixed random order, in the appropriate row/column location on the MCS stimulus palette (see Fig. 1 in “► [World Color Survey](#)”). The palette format of the naming task data often has participant’s “focus” choices embedded in the datasheet as can be seen in panel (b)’s column 0 which has a “circled-X” symbol shown in cells **A0** and **J0**, recording the “foci” indicated by this participant for color appearances which in English language would gloss as the focus of “white” and the focus of “black,” respectively.
2. **Focus selection task:** Informants were shown the entire 330 color stimulus palette and asked to select the best example, or focus, of each

a

oktoma : Guarijito de Senora Baja
langue: *Guarijito*

no. de palabras: *Guarijito*
Inventar No. *1*

U [setoname] *Koursak*

1. <i>sixomúrame</i>	29. <i>tisa sixoname</i>	41. <i>setoname</i>	42. <i>tisa sixomúrame</i>
2. <i>tesa sixoname</i>	30. <i>tisa setoname</i>	43. <i>warosahérame</i>	43. <i>tisa sixomúrame</i>
3. <i>setoname</i>	31. <i>tisa tesaoname</i>	44. <i>ohóname</i>	44. <i>tisa sixoname</i>
4. <i>ohóname</i>	32. <i>ka'we sixoname</i>	45. <i>tesamúrame</i>	45. <i>setapórame</i>
5. <i>-sixoname</i>	33. <i>tisa sixomúrame</i>	46. <i>sixomúrame</i>	46. <i>tesa pódame</i>
6. <i>warósa</i>	34. <i>tisa ohómúrame</i>	47. <i>wetapórame</i>	47. <i>sixomúrame</i>
7. <i>ka'we sixoname</i>	35. <i>tisa setamúrame</i>	48. <i>ohómúrame</i>	48. <i>setamúrame</i>
8. <i>setomúrame</i>	36. <i>waróherame</i>	49. <i>warósa hérame</i>	49. <i>tisa wetapórame</i>
9. <i>warosahérame</i>	37. <i>setamúrame</i>	50. <i>sixomúrame</i>	50. <i>tesa pódame</i>
10. <i>ka'we sixoname</i>	38. <i>okorá hérame</i>	51. <i>ka'we ohóname</i>	51. <i>-tesa setoname</i>
11. <i>-ohómúrame</i>	39. <i>sixomúrame</i>	52. <i>tisa wetapórame</i>	52. <i>sixoname</i>
12. <i>sa'wame</i>	40. <i>setapórame</i>	53. <i>tesa múrame</i>	53. <i>setapórame</i>
13. <i>-tisa ohóname</i>	41. <i>-sixomúrame</i>	54. <i>setapórame</i>	54. <i>tisa sixoname</i>
14. <i>ka'we ohóname</i>	42. <i>ka'we sixomúrame</i>	55. <i>-sixomúrame</i>	55. <i>ohóname</i>
15. <i>seta pódame</i>	43. <i>sawa múrame</i>	56. <i>tisa sixoname</i>	56. <i>tisa warosahérame</i>
16. <i>tesa pódame</i>	44. <i>ohómúrame</i>	57. <i>tisa warósa hérame</i>	57. <i>sixoname</i>
17. <i>tisa setoname</i>	45. <i>tisa sixomúrame</i>	58. <i>ohómúrame</i>	58. <i>seta pódame</i>
18. <i>-ohóname (?)</i>	46. <i>setoname</i>	59. <i>wetapórame</i>	59. <i>ohóname (ka'we)</i>
19. <i>tisa sixomúrame</i>	47. <i>sixoname</i>	60. <i>tesa pódame</i>	60. <i>ohómúrame</i>
20. <i>ka'we sixoname</i>	48. <i>sixomúrame</i>	61. <i>-ohóname</i>	61. <i>tisa sixoname</i>

b

oktoma : Guarijito de Senora Baja - *Guarijito*
langue: *Guarijito*

	1	2	3	4
A	<i>tahsárame</i>			
B	<i>tahsárame</i>	<i>tisa wa'rósa</i>	<i>tesa pódame</i>	<i>tesa pódame</i>
C	<i>tesa pódame</i>	<i>tesa pódame</i>	<i>tisa wa'rósa</i>	<i>tisa wa'rósa</i>
D	<i>tisa wetapórame</i>	<i>tisa wetapórame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>
E	<i>tesa wetapórame</i>	<i>tisa setoname</i>	<i>tisa setoname</i>	<i>tesa múrame</i>
F	<i>tesa pódame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>
G	<i>tesa pódame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>
H	<i>tesa pódame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>
I	<i>tesa pódame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>
J	<i>tesa pódame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>	<i>tesa múrame</i>

Mesoamerican Color Survey Digital Archive, Fig. 3 MCS naming task datasheet formats shown as image scans of the raw data

different color name elicited in the naming task (see [28, 29] for method details). Again, individual variations do exist in the raw datasheets. Figure 4 shows two common raw datasheet formats used to collect focus selection task data. In panel (a) focus task data is reported in conjunction with naming task data, whereas in panel (b) focus selection task data is combined with mapping task data (mapping task is detailed below). In both formats shown in Fig. 4, focus selections are indicated by either a “circled-X” symbol (panel (a)) or by a plain “X” symbol alone (as in panel (b)). In focus task datasheets, it is common to find a dictionary of color terms elicited at the bottom of the sheet (as seen in both Fig. 4 panels). Mapping task datasheets most frequently delineate mapped category areas using hatched markings (e.g., Fig. 4b); however these data were also often recorded using color pencils (as shown in Fig. 5 and described in the next section). As with the naming task data, image scans of raw MCS datasheets, and transcribed digital data files for focus data, are provided on the archive website.

3. **Category mapping task:** For the same color category terms as in the focus selection task, informants were shown the entire 330 color stimulus palette and asked to indicate the regions of the stimulus palette in which the appearance of a given color term was represented. The informants in this way would map the color category stimulus regions glossed by each color term for which an investigator solicited a mapping (see [28, 29] for method details). This task is also referred to as the rice mapping task in the literature, because grains of rice were often provided to informants to mark the many stimuli that were denoted by a given color term. Figure 5 shows two raw datasheets, for Guarijio and Chinantec languages, illustrating, above and beyond data variations, slight variations in the ways investigators recorded the surveyed information. Again, in some cases Focus locations are recorded, whereas in other Mapping Task datasheets Foci are not recorded. In mapping task datasheets, it is common to find a

dictionary of color terms elicited at the bottom of the sheet and a pseudo-color code legend is given to interpret the regions defined using a particular color pencil as belonging to a specific color term (as seen in both Fig. 5 panels). (In such pseudo-color codes, the color of the ink used is not to be construed as resembling the color of the category data that it encodes.) As with the other MCS data, image scans of raw MCS datasheets are provided on the archive website.

The above details of MCS data, including Figs. 3, 4, and 5, convey the kind of coding variation found throughout the datasheets in the MCS archive. The variation is not wholly unexpected as a large number of the 116 languages of the MCS were collected by different investigators, at different locations, over the approximately 5 years during which the MCS was conducted. Notwithstanding this amount of idiosyncratic coding style, the MCS raw data, as well as the rest of the MacLaury archive, appear to be of rather consistent and thorough formatting concerning the data that is collected, especially considering the mass of information that exists in the estimated 20,000 pages of the archive.

The MCS and the Robert E. MacLaury Color Categorization Digital Archive (ColCat)

Recognizing the value in the MCS data, the interdisciplinary *ColCat* research group at UC Irvine sought to convert the paper copy of MacLaury’s research archive into a public-access digital database, similar to that developed for the WCS by Paul Kay and colleagues [2]. The entirety of MacLaury’s MCS data is included in this *ColCat* digital archive [34]. A major aim of the ColCat project is to develop a platform to make the archive, including the MCS data, readily available to the scientific research and teaching community. The project presently makes available organized sets of scanned image files of datasheet pages in the archive acquired from MacLaury’s estate and continues to implement an intuitive Internet-based



Mesoamerican Color Survey Digital Archive, Fig. 5 Two mapping task datasheet formats (panel (a) Guarijio and (b) Chinantec languages) presented as image scans of the raw data

user interface to serve as the front end of a collaborative research space with which researchers can readily explore and investigate their own basic research questions concerning the MCS data, as well as the color categorization data from other languages that MacLaury assessed. As mentioned earlier, neither the raw MCS data nor the MacLaury archive's additional non-Mesoamerican languages have been systematically organized for public use or previously published in an unanalyzed form. The additional 70 color categorization surveys (many with only a single informant, although others with frequently many more, and the largest with 40 informants) are valuable for their diversity in that they include native speakers from a wide range of languages including several Slavic languages, Hungarian, several Salishan languages of the Pacific Northwest United States, Zulu and several other South Africa/Zimbabwe languages, several native American languages, Germanic languages, European languages, Asian languages, and more.

The *Robert E. MacLaury Color Categorization (ColCat) Digital Archive* uses a content management system (CMS) for organizing, editing, and publishing the contents of the archive to a web-based graphical user interface (GUI) hosted at <http://colcat.calit2.uci.edu>. A partial list of features provided for the MCS through ColCat is summarized below:

1. User access level permissions and controls for the archive areas of the Wiki. User logins that define various levels of access for researchers, including those who may only want unregistered access to view database contents. Member level access for registered users who want to interact with the database. Editor level access as a trusted contributor to the database. Administrator access for systems operators who maintain database changes and review and approve contributions made by editor level users. Proper controls are implemented to maintain long-term integrity of the archive and to forestall loss and corruption of datafiles as well as accidental losses of data.
2. Facilitated search and query tools for exploring aspects of the datasets such as language type, language family, specific color terms, participant group subsets (e.g., by gender, age, etc.), survey geographic locations, and other filters that have been specified by content in the archived data.
3. Transcription quality indicators for evaluating conversion accuracy of handwritten data content into digitally addressable datasets. Because there are several ways the raw handwritten data in the archive can be converted into data addressable digital files (e.g., manual transcription, crowdsourced transcription, optical character recognition or OCR-based transcription, linguistic expert transcription, etc.), a useful feature of the ColCat Wiki includes utilities that permit assessment of each dataset's transcription quality and comparison across multiple forms of transcription of the same dataset. Transcription quality indicators appear on every surveyed language's archive page, and, where appropriate, indicate the degree of transcription completion, and computed degree accuracy, for each language's transcribed datasets.
4. To encourage users to explore the archive's original raw datasheets, the *ColCat Wiki* provides each survey's scanned image files of the archive's raw data organized in a browsable catalog/library of pdfs (like those seen in Figs. 3, 4, and 5 above). Onboard pdf files of raw datasheets permit visual inspection of investigator's ethnographic notes and embellishments; verification of digitized data content; and querying pdfs by language family, language type, location and so forth, as well as querying by combinations of such tags.
5. An onboard suite of supporting resources for color categorization research. For example, (a) on every surveyed language's archive page (where appropriate and provided for by the archive's contents), maps of the geographic regions are provided where surveys were conducted, permitting visualization of geographical relatedness among the archive's datasets, as well as allowing for search and

identification of database contents by exploring geographical locations (via a Google Map plugin) where the actual survey data originated. (b.) Bibliographic resources highlighting specific languages or language families, including lists of peer-reviewed journal publications, books, relevant media, and reference materials, provided for each of the archive's languages surveyed. (c.) Active links out to relevant Internet resources, such as *World Atlas of Language Structures Database* (<http://wals.info>), and the public-access site *Ethnologue: Languages of the World Database* (<http://www.ethnologue.com>).

6. As mentioned above a major aim of the archive project was to develop a platform to make the archive's data readily available to the scientific research and teaching community. Toward this end, the archive is designed to function as a collaborative research space, with user contributions enabled for trusted contributors, an informal blog feature to encourage a community of research exchange, and a platform for sharing and communicating results and for collective problem solving.

Additionally important was the aim of making the digital archive's data accessible to a broad audience comprised of users with varying levels of expertise in handling and analyzing large amounts of data. For this reason, a set of *ColCat data tools* was developed which allow users to search, filter, and analyze data in ways that best meet their research goals and technical abilities.

The planned data-handling tools and utilities aim to facilitate, for all user levels, the use of the archive's contents for original color categorization research studies.

The ColCat digital archive, including the MCS, contains data at three levels:

1. Digital image scans, in pdf format, of approximately 12,500 handwritten pages of color-naming data, ethnographic notes, and support documents from 116 Mesoamerican languages, plus pdf image scans of an additional

10,000 pages surveyed globally from other languages and nations

2. Raw transcription data of these surveys converted and collected through OCR, crowd sourcing, and expert transcribers, and
3. "Official" transcription conversions for the archive's color-naming and categorization data

The Wiki's onboard data tools are designed for visualizing and analyzing the computer-addressable archive data at levels 2 and 3. Visualization tools for transcription data (level 2) include quantitative indicators of completeness and quality of transcriptions for each language and visualizations of the variance between or within transcription methods for entire languages or subsets of images within a language. Visualization tools for color-naming data (level 3) will include visualizations of the aggregated color-name mappings for each language including information about the proportion of agreement among all respondents or subsets of respondents through the use of filters – i.e., based on gender, age, etc.

Data export tools will allow users to download both transcription and naming data (levels 2 and 3) in formats conducive to off-line analysis (i.e., CSV, Excel, etc.). Although information about aggregated color-name mappings (level 3) will be based on data from all subjects in a language, users may be interested in computing and/or comparing consensus color-name mappings for subpopulations based on their own criteria. To that end, downloadable tools will allow users to take the same analyses and visualizations that are used on the Wiki for complete populations and apply them to subpopulations that were exported by the user. The expectation is that a variety of user types will employ the Wiki's visualization tools to help identify subsets of data in the archive that meet specific research goals, while other researchers will be more inclined to export data in a familiar format (i.e., in an established WCS format) and apply their own off-line analyses. Regardless of the case, the ColCat data tools make the entire archive's data accessible to a wide range of researchers, students, and other scientists who

would like to explore and learn from the archive but are not well equipped to manage and process large quantities of data.

Uses of the MCS Archive

As a complementary database to the World Color Survey database, the addition of the MCS archive allows for an even greater, more extensive, assessment of behaviors of cross-cultural color cognition and color naming. For example, reconsider the Fig. 2(a) discussion above, concerning the 369 surveyed Otomanguean language family informants for which MCS data exists. These MCS informants can now be independently compared to, or even analyzed in conjunction with, the 175 previously surveyed WCS informants from the Otomanguean language family. In addition to the novel surveys it adds, it is this kind of duplication and augmentation that the MCS provides as a complementary database to the WCS that will increase the value of both archives and present new opportunities to greatly extend understanding of how the data in the MCS and the WCS inform on the questions originally asked by the surveys' directors, namely, what specifically are the universal and culturally specific patterns of color naming and categorization that can be empirically demonstrated for large samples of the world's diverse languages and how do they inform us on the larger question of general human tendencies for categorizing and naming objects perceived in the world.

Cross-References

- [Berlin and Kay Theory](#)
- [Color Categorical Perception](#)
- [Color Category Learning in Naming-game Simulations](#)
- [Color Dictionaries and Corpora](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Effect of Color Terms on Color Perception](#)
- [Multilingual/Bilingual Color Naming/ Categories](#)
- [Vantage Theory of Color](#)

References

1. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
2. Kay, P., Berlin, B., Maffi, L., Merrifield, W.: *The World Color Survey*. Center for the Study of Language and Information, Stanford (2003)
3. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9085–9089 (2003)
4. Regier, T., Kay, P., Cook, R.: Focal colors are universal after all. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8386–8391 (2005)
5. Lindsey, D.T., Brown, A.M.: Universality of color names. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16609–16613 (2006)
6. Lindsey, D.T., Brown, A.M.: World color survey color naming reveals universal motifs and their within-language diversity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19785–19790 (2009)
7. Kay, P., Regier, T.: Color naming universals: the case of Berinmo. *Cognition* **102**, 289–298 (2007)
8. Davidoff, J., Davies, I.R.L., Roberson, D.: Color categories in a stone-age tribe. *Nature* **398**, 203–204 (1999)
9. Roberson, D., Davies, I.R.L., Davidoff, J.: Color categories are not universal: replications and new evidence from a stone age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
10. Roberson, D., Hanley, J.R.: Color categories vary with language after all. *Curr. Biol.* **17**, 605–606 (2007)
11. MacLaury, R.E.: Color-category evolution and shuswap yellow-with-green. *Am. Anthropol.* **89**(1), 107–124 (1987)
12. Paramei, G.V.: Singing the Russian blues: an argument for culturally basic color terms. *Cross-Cult. Res.* **39**(1), 10–38 (2005)
13. Dedrick, D.: Color language universality and evolution: on the explanation for basic color terms. *Philos. Psychol.* **9**(4), 497–524 (1996)
14. Jameson, K.A.: Culture and cognition: what is universal about the representation of color experience? *J. Cogn. Cult.* **5**(3–4), 293–347 (2005)
15. Alvarado, N., Jameson, K.A.: Confidence judgments and color category best exemplar salience. *Cross-Cult. Res.* **39**(2), 134–158 (2005)
16. Jameson, K.A.: Why GRUE? An interpoint-distance model analysis of composite color categories. *Cross-Cult. Res.* **39**(2), 159–194 (2005)
17. Jameson, K.A.: Where in the world color survey is the support for the hering primaries as the basis for color categorization? In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, pp. 179–202. The MIT Press, Cambridge (2010)
18. Davies, I.R.L., Corbett, G.G.: A cross-cultural study of color grouping: evidence for weak linguistic relativity. *Br. J. Psychol.* **88**(3), 493–517 (1997)

19. Komarova, N.L., Jameson, K.A.: A quantitative theory of human color choices. *PLoS One* **8**(2), e55986 (2013). doi:10.1371/journal.pone.0055986
20. Bimler, D.: Are color categories innate or internalized? Hypotheses and implications. *J. Cogn. Cult.* **5**(3), 265–292 (2005)
21. Bimler, D.: From color naming to a language space: an analysis of data from the world color survey. *J. Cogn. Cult.* **7**(3), 173–199 (2007)
22. Bimler, D., Uusküla, M.: Clothed in triple blues: sorting out the Italian blues. *J. Opt. Soc. Am. A* **31**, A332–A340 (2014)
23. Narens, L., Jameson, K.A., Komarova, N.L., Tauber, S.: Language, categorization, and convention. *Adv. Complex Syst.* **15**(03n04), 1150022 (2012)
24. Jameson, K.A., Komarova, N.L.: Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *J. Opt. Soc. Am. A*, **26**(6), 1414–1423. Featured Reprint in *The Virtual J. Biomed. Opt.* **4**(8), (2009)
25. Jameson, K.A., Komarova, N.L.: Evolutionary models of color categorization. II. Realistic observer models and population heterogeneity. *J. Opt. Soc. Am. A*, **26**(6), 1424–1436. Featured Reprint in *The Virtual J. Biomed. Opt.* **4**(8), (2009)
26. Komarova, N.L., Jameson, K.A.: Population heterogeneity and color stimulus heterogeneity in agent-based color categorization. *J. Theor. Biol.* **253**, 680–700 (2008)
27. Komarova, N.L., Jameson, K.A., Narens, L.: Evolutionary models of color categorization based on discrimination. *J. Math. Psychol.* **51**, 359–382 (2007)
28. MacLaury, R.E.: *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. University of Texas Press, Austin (1997)
29. MacLaury, R.E.: *Color in mesoamerica. Vol. 1: a theory of composite categorization*. Doctoral dissertation. University of California, Berkeley. UMI University Microfilms, No. 8718073, Ann Arbor (1986)
30. MacLaury, R.E.: From brightness to hue: an explanatory model of color category evolution. *Curr. Anthropol.* **33**(2), 137–186 (1992)
31. Regier, T., Kay, P., Khetarpal, N.: Color naming and the shape of color space. *Language* **85**, 884–892 (2009)
32. Webster, M., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R., Paramei, G., Dedrick, D. (eds.) *Anthropology of Color*, pp. 29–53. John Benjamins, Amsterdam (2007)
33. Paul, L.M., Simons, G.F., Fennig, C.D. (eds.): *Ethnologue: Languages of the World*, Eighteenth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com> (2015)
34. Jameson, K.A., Gago, S., Deshpande, P.S., Benjamin, N.A., Chang, S.M., Tauber, S., Jiao, Y., Harris, I.G., Xiang, Z., Bhakta, H.R., MacLaury, R.E.: *The Robert E. MacLaury Color Categorization (ColCat) Digital Archive*. <http://colcat.calit2.uci.edu/>. The California Institute for Telecommunications and Information Technology (Calit2), UC Irvine (2015)

Metal Halide Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

Metal halide lamp

Definition

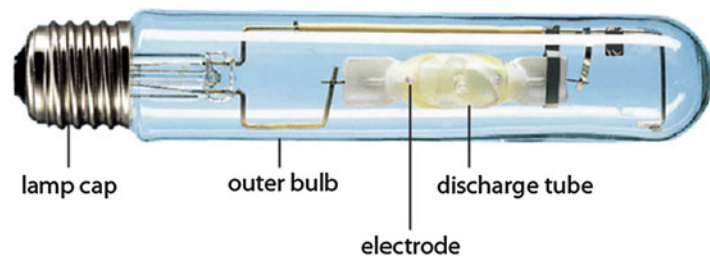
Lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a high-pressure mercury vapor with metal halide additives, that is contained in a transparent bulb.

Metal Halide Gas Discharge Lamps

Metal halide lamps are high-pressure mercury gas discharge lamps that contain metal halides in addition to the mercury [2, 3]. In the heated discharge tube, the metals of the halides take part in the discharge process and radiate their own spectrum. Compared with high-pressure mercury lamps, both color properties and efficacy are considerably improved. Thanks to the fact that no fluorescent powder is needed, the small gas discharge tube itself is the light-emitting surface. This small light-emitting surface makes the lamps extremely suitable for use in reflector and floodlight luminaires. Originally, they were solely produced in extremely high lumen packages with relatively large lamp dimensions (about the size of a 1 litre bottle). These lamps are particularly suited for use in floodlights for the lighting of stadiums. Today, compact metal halide lamps are also available in small lumen packages, which makes them very suitable for use in compact reflector luminaires for

Modified and reproduced from [1] by permission of Philips Lighting, Eindhoven, The Netherlands. © 2012 Koninklijke Philips Electronics N.V.

Metal Halide Lamp,
Fig. 1 Main parts of a
 metal halide lamp [1]



accent lighting both indoors and outdoors. These versions have become an energy-efficient alternative for halogen lamps. Special compact metal halide versions are being produced for use in film studios and theaters. The smallest metal halide lamps are used for car headlamps (these are the so-called xenon lamps, whose gas discharge tube is no larger than a match head). Metal halide lamps belong to the group of high-intensity discharge (HID) lamps, because they are available in high-lumen output (and thus high-luminous-intensity) versions.

Working Principle

Suitable metals that vaporize in the hot discharge tube so as to contribute to the discharge process cannot be added directly to the mercury in the discharge tube. This is because metals suitable for this purpose attack the wall of the gas discharge tube at the high temperatures that are needed for the gas discharge to occur. The solution to this problem is to add these metals in the form of their nonaggressive chemical compounds with a halogen (iodine, bromide, or chloride), hence the name metal halide lamps. The solid metal halide starts evaporating once, after switching on the lamp, the mercury discharge has increased the temperature in the discharge tube. When this metal halide vapor enters the area of the mercury discharge in the center of the tube with its very high temperature (around 3,000 °C), it dissociates into its separate elements: metals and halogen. There the metals in their pure, vaporized state take part in the discharge process and determine the efficacy and color characteristics of the radiation. The aggressive vaporized metal cannot reach the tube wall because at the lower wall temperature (some 1,000 °C), they

recombine again to form the harmless metal halide compound. Like all gas discharge lamps (with very few exceptions), a metal halide lamp cannot be operated without a ballast to limit the current flowing through it. It also needs, again like most gas discharge lamps, an igniter for starting the lamp.

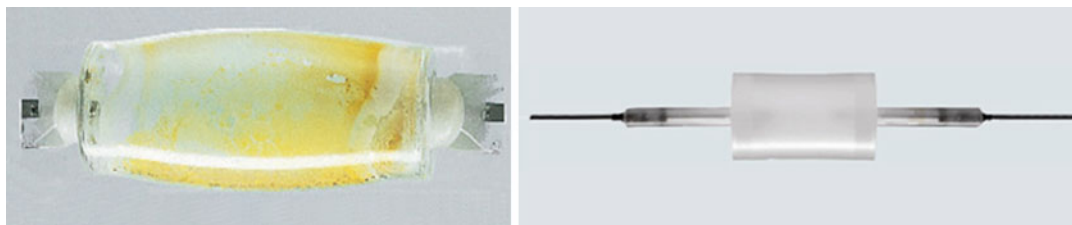
Materials and Construction

The main parts of a metal halide lamp are (Fig. 1):

- Discharge tube
- Metal halide additives
- Fill gas
- Outer bulb
- Electrodes
- Lamp cap

Discharge Tube

The discharge tube is made of either quartz or a ceramic material. Some metals of the metal halide compounds (especially sodium) have the tendency, at high temperatures, to migrate slowly through the quartz wall of the tube, with the result that there is a gradual change in the color properties of the lamp during its lifetime. The use of ceramic discharge tubes solves the problem. Such tubes are impervious to these metals, even at high operating temperatures. High-pressure sodium lamps make use of the same ceramic material. Metal halide lamps using ceramic material are referred to as ceramic metal halide discharge lamps. Figure 2 shows examples of both quartz and ceramic gas discharge tubes. Since the ceramic tube is produced by sintering minuscule aluminum-oxide particles, one cannot look



Metal Halide Lamp, Fig. 2 Gas discharge tube made of quartz (*left*) and ceramic material (*right*), respectively. The yellow material seen through the quartz tube is condensed metal halide [1]

Metal Halide Lamp, Fig. 3 *Top* tubular hard-glass, single-ended outer bulb and *bottom* quartz, tubular double-ended outer bulb



straight through the wall, although the light transmittance is more than 90 % (translucent instead of transparent material). Ceramic gas discharge tubes are operated at a higher temperature than are quartz tubes. The higher operating temperature influences the spectrum of the radiation. The result is a lower color temperature compared to that of quartz metal halide lamps and a 10 % higher efficacy.

Metal Halide Additives

In theory, some 50 different metals can be used for metal halide compounds, and different manufacturers have introduced various combinations of these metals. Examples of some of the metals used in metal halide compounds are sodium, thulium, thallium, indium, scandium, dysprosium, and tin. Most lamps use a mixture of at least three different metal halides. Each different combination results in a different spectrum, but also in different efficacy, lumen maintenance, and lifetime characteristics.

Fill Gas

Besides metal halides and mercury, an inert gas is also added to the discharge tube, like in most other discharge lamps.

Electrodes

The electrodes of metal halide lamps are of the type used in normal high-pressure mercury lamps. They consist of a tungsten rod, with a tungsten coil impregnated with emissive material, wound around it. They are of heavier construction than the normal high-pressure mercury lamps because of the higher operating temperature.

Outer Bulb

Most metal halide lamps use a hard-glass or (for very compact versions) quartz outer bulb for protection and for heat insulation of the discharge tube. Both single-ended and double-ended outer bulbs are used (Fig. 3). They may be evacuated or gas filled. In the case of quartz outer bulbs, a UV-blocking quartz is used to limit harmful UV radiation. For the same reason, some hard-glass versions use a cylindrically shaped UV-block shield around the discharge tube.

The inner wall of the outer bulb reflects a very small part of the light that consequently cannot be very well controlled by a luminaire reflector. But since the reflected amount is so small, this is normally not a problem, except where very-high-lumen-output lamps are employed. In floodlight installations, a very small amount of



Metal Halide Lamp, Fig. 4 Double-ended quartz metal halide lamp not making use of an outer bulb

Metal Halide Lamp, Fig. 5 Various lamp caps for compact single-ended metal halide lamps



uncontrollable light could give rise to disturbing light pollution. For these applications special lamps are therefore available that have no outer bulb (Fig. 4). For reasons of thermal stability and safety, those lamps without an outer bulb must be used in specially designed luminaire housings.

Lamp Caps

Metal halide lamps come with a great variety of lamp caps (Fig. 5). The single-ended lamps of higher wattage have in general E40 Edison screw caps. Some caps have special electrical insulation because of the high ignition pulses of between 600 and 5,000 V. The lower-wattage single-ended lamps have two-pin electrical connections of various types, sometimes with ceramic electrical insulation. In order to be able to exactly position the gas discharge arc in a luminaire, some lamps are provided with a so-called prefocused lamp cap (Fig. 5, right). The double-ended lamps use lamp caps of the type that are also used in double-ended halogen lamps (Fig. 3, bottom). The

lamps without an outer bulb again have different lamp caps (Fig. 4).

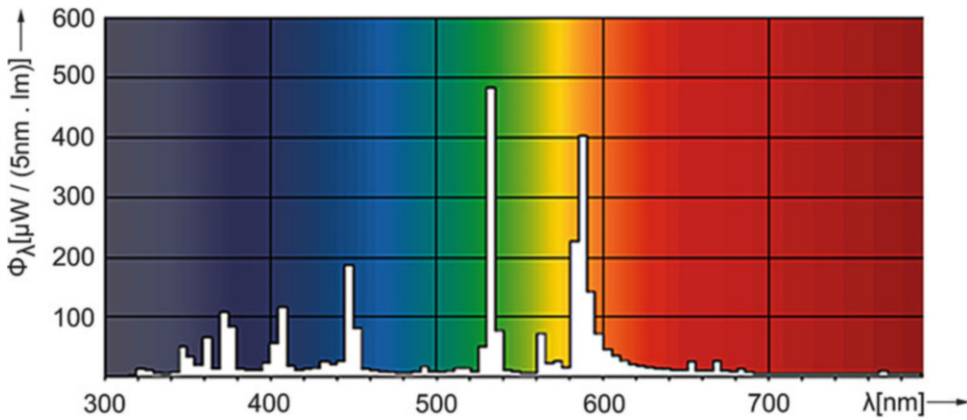
Properties

Energy Balance

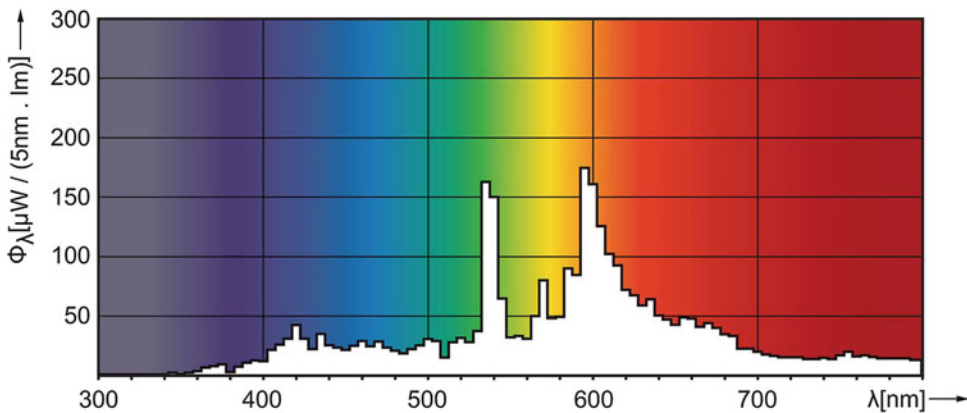
Almost 25 % of the input power of metal halide lamps is emitted in the form of visible radiation. Compare this with 15–17 % of normal high-pressure mercury lamps and 30 % of high-pressure sodium lamps.

System Luminous Efficacy

The compact versions of the metal halide lamp have lumen efficacies (depending on the metal halide mixture and gas discharge tube material used) of between 70 and 95 lm/W. When comparing this with normal halogen incandescent lamps with their maximum efficacy of 25 lm/W, it is clear that compact metal halide lamps are often very suitable to replace halogen lamps. The larger



Metal Halide Lamp, Fig. 6 Spectral energy distribution of a metal halide lamp, color type 642: T_k 4,200 K and R_a 65 [1]



Metal Halide Lamp, Fig. 7 Spectral energy distribution of a ceramic metal halide lamp, color 830: T_k 3,000 K and R_a 80 [1]

versions of metal halide lamps have efficacies from some 75 lm/W to slightly more than 105 lm/W.

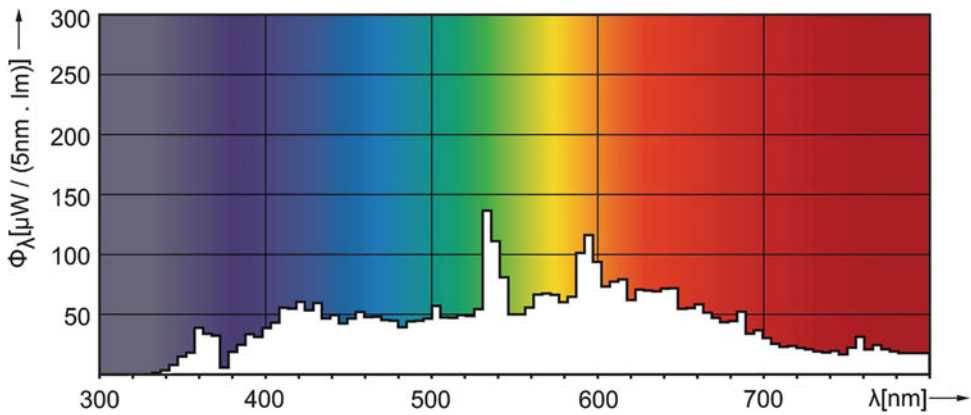
Lumen-Package Range

Compact versions are available in lumen packages from some 1,500 to 25,000 lm (corresponding wattage range 20–250 W). Larger versions range from some 20,000 lm to more than 200,000 lm (corresponding wattage range 250–2,000 W). Special (compact) types for use in film studios, theaters, and professional photography are produced in lumen packages of up to more than 1,000,000 lm (12 kW).

Color Characteristics

As has been explained in the sections above, by employing different metal halide mixtures, lamps with different spectra can be produced. Like all gas discharge lamps, the metal halide lamp has a discontinuous spectrum. Figures 6, 7, and 8 show the spectra of some typical metal halide lamps with different color temperatures T_k and color rendering index R_a . The same color designation system is used as with fluorescent lamps.

Compact metal halide lamps are normally produced in the color temperature range of 3,000 K to some 4,500 K in two color rendering varieties: R_a ca. 80 and ca. 90. The larger metal halide lamps are



Metal Halide Lamp, Fig. 8 Spectral energy distribution of a ceramic metal halide lamp, color 942: T_k 4,200 K and R_a 90 [1]

available in the color temperature range from approximately 4,000 to 6,000 K, with color rendering R_a in the range from 65 to more than 90. The special, daylight, metal halide lamp types for film studio and theater lighting have high color temperatures of between 6,000 and 8,000 K with color rendering values R_a between 65 and more than 90.

Lamp Life

The lamp life of most metal halide types is somewhat shorter than that of other gas discharge lamps. This is because the electrodes are heated to a higher temperature with a correspondingly higher evaporation rate and are gradually worn out by chemical reactions with the metal halides. It has also been shown that different metal halide lamp types employ widely different construction methods. As a consequence, lamp life varies strongly with type. The economic life of compact versions lies between some 7,000 and 14,000 h (20 % mortality). The high-lumen-package versions vary from 4,000 h (single-envelope types for stadium floodlighting) to some 10,000 h (again 20 % mortality). Those ceramic metal halide lamps specifically developed for use in road lighting, where lights may be used 4,000 h a year, have lifetimes up to 20,000 h (20 % mortality).

Lamp-Lumen Depreciation

Metal halide lamps have a higher lumen depreciation than most other discharge lamps. This is

Metal Halide Lamp, Table 1 Different metal halide lamp types

Lamp parameter	Range
Lamp size	Medium/compact/very compact
Discharge-tube material	Quartz/ceramic
Lamp shape	Single ended/double ended
Outer bulb	Outer bulb/coated outer bulb/no outer bulb
Lumen package	Large/small
Lamp circuit	Electromagnetic/HF electronic
Color temperature T_k	3,000–4,000 K/4,000–7,000 K
Color rendering index R_a	$R_a > 90$ (900 series)/ $80 < R_a < 90$ (800 series)/ $R_a < 80$
Lifetime	Standard (7,000–14,000 h)/long (up to 20,000 h)
Reflector	No reflector/built-in reflector

because of the higher degree of blackening from evaporated electrode material. Here too, the type of construction and sort of metal halides used play a role. Lumen depreciation values vary between some 20 % and 30 % (after 10,000 h). Some special types, such as those employed in road lighting, depreciate less rapidly (some 10 % after 10,000 h).

Burning Position

With the larger gas discharge tubes in particular, the burning position may affect the actual location



Metal Halide Lamp, Fig. 9 Examples of metal halide lamps used mainly in indoor lighting



Metal Halide Lamp, Fig. 10 Examples of metal halide lamps used mainly in outdoor lighting. *First three:* for road and industrial lighting. *Last four:* for sports floodlighting

of the various metals in the tube. This means that different burning positions may result in different color shifts. Also, with some types of lamp construction, the burning position can influence the lifetime of the lamp, for example, because of attack of the electrodes by some of the metals of

the halides. Many metal halide lamps therefore have restrictions as to their permitted burning position (these are, of course, specified in their accompanying documentation). Compact, single-ended types, with either quartz or ceramic tube, have universal burning positions.

Run-Up and Reignition

The metal halides in the discharge tube need time to heat up, evaporate, and dissociate into metal and halide. During this process, which takes about 2–3 min, the light output and color gradually change until the final stable condition is reached. If there is an interruption in the power supply, medium- and high-wattage lamps will take approximately 10–20 min for the pressure in the lamp to decrease enough for it to reignite. Compact ceramic lamps reignite much faster: after some 3–5 min. Immediate reignition is only possible by applying a very high-voltage pulse (typically 60 kV). For this purpose, special, hot restrike, lamps are produced with special heavily insulated electrical contacts for applying this high-voltage pulse.

Dimming

The dimming of metal halide lamps is difficult because with the resulting decrease in temperature, some of the metal halides condense, changing the constitution of the metal halides actually participating in the discharge. This in turn changes the color properties of the light. By employing a specially shaped burner, electronically driven lamp versions have been developed that can be dimmed to some 50 % without suffering from this problem.

Mains-Voltage Variations

Mains-voltage variations affect lumen output, lifetime, and light color. A mains voltage that deviates 10 % from its nominal value will result in perceivable color shifts. Thanks to the higher operating temperature of ceramic lamps, the effect on color change here is considerably smaller. Both up and down variations in the mains shorten lamp life.

Product Range

Metal halide lamps are available in a wide variety of types for many different applications, including floodlighting, road lighting, accent lighting (both interior and exterior), car head lighting, and film-studio lighting. The more important lamp

parameters, together with the corresponding range for which different versions are made, are listed in Table 1. Figure 9 shows some examples of lamp types used mainly in indoor lighting, while Fig. 10 shows those types especially used in outdoor lighting.

Cross-References

► [High-Pressure Mercury Lamp](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: *Lighting Hardware: Lamps, Gear, Luminaires, Controls*. Course book, Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: *Lamps and Lighting*, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: *IES Handbook*, 10th edn. Illuminating Engineering Society of North America, New York (2011)

Metamerism

Peter van der Burgt

Philips Lighting, Valkenswaard, The Netherlands

Synonyms

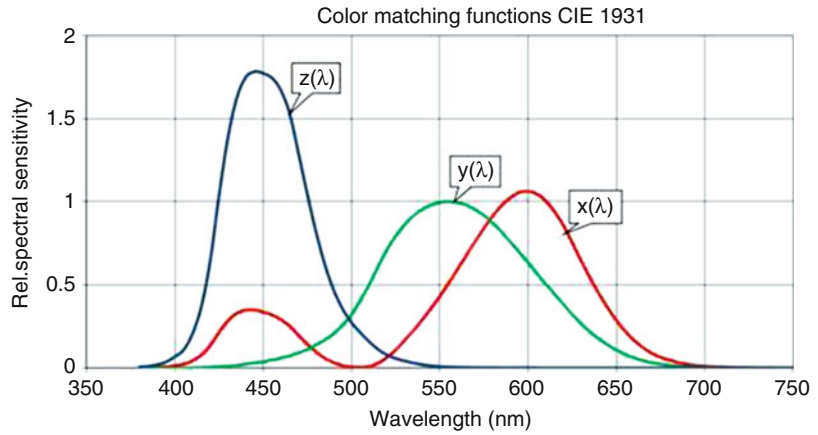
[Nil](#)

Definition

Metamerism is the matching of apparent color of objects with lights that have different *spectral power distributions*. Colors that match are called metamers.

Color of the Light

The light coming from any light source (sunlight, incandescent lamp) can be described by the

Metamerism,**Fig. 1** Color-matching functions CIE

so-called spectral power distribution, often denoted as SPD. The SPD gives the amount of power radiated by the light source as a function of wavelength in the visible range of the electromagnetic spectrum being 380–740 nm. Every SPD will generate a color impression in the human visual system. To describe the color of the light, a system has been developed by the CIE (Commission Internationale de l'Eclairage). A set of three color-matching functions called $x(\lambda)$, $y(\lambda)$, and $z(\lambda)$ has been defined, and these are known collectively as the CIE standard observer [1]. These color-matching functions are depicted below in Fig. 1.

By using the color-matching functions, the effect of a *spectral power distribution* $E(\lambda)$ can be expressed by the following equations:

$$X = \sum E(\lambda) \cdot x(\lambda) \quad (1)$$

$$Y = \sum E(\lambda) \cdot y(\lambda) \quad (2)$$

$$Z = \sum E(\lambda) \cdot z(\lambda) \quad (3)$$

The values X , Y , Z are called the tristimulus values of the color of the light with spectral power distribution $E(\lambda)$.

The tristimulus values for a light can be converted to the chromaticity or color point x, y by the following equations:

$$x = X/P \text{ with } P = X + Y + Z \quad (4)$$

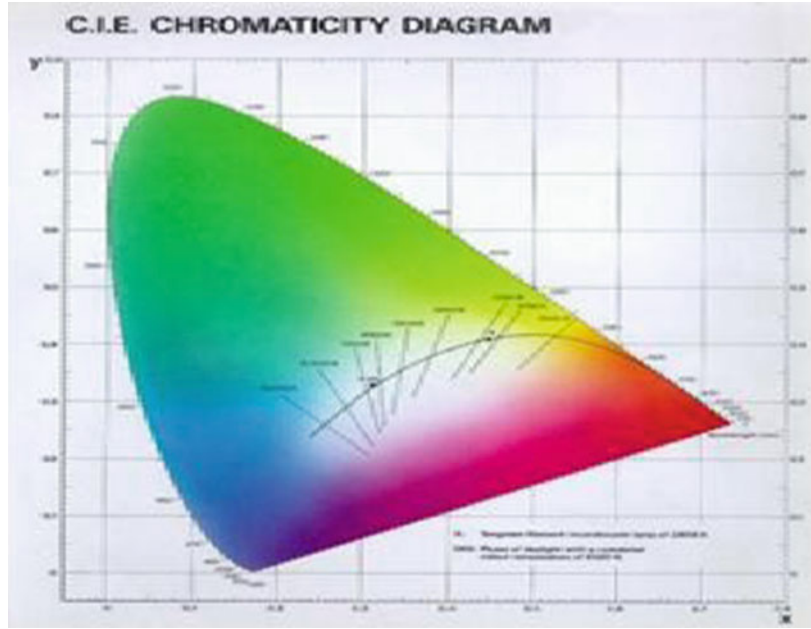
$$y = Y/P \quad (5)$$

These equations lead to the well-known CIE chromaticity diagram which is shown in Fig. 2. With the equations above, the color points for monochromatic lights can be calculated. This leads to the spectral locus, the outer curved boundary in the diagram. The straight edge on the lower part of the gamut is called the line of purples. These colors, although they are on the border of the gamut, have no counterpart in monochromatic light. Less saturated light colors appear in the interior of the figure. The diagram shows in the colored area all the chromaticities that are visible to the average person, and this colored region is called the gamut of human vision. One important property of this diagram is that lights that have the same chromaticity will give the same color impression to an observer. The chromaticity however does not predict the precise color impression of a light. For example, an incandescent lamp burning in the middle of the day may give a yellowish impression. The same lamp in the night will yield a white impression. So the precise color impression will depend on the adaptation of the observer. But at the same adaptation, the same chromaticity will lead to the same color impression.

Metamerism of Light Colors

Equations 1, 2, and 3 show that it is possible that more lights having different SPDs can

Metamerism, Fig. 2 CIE chromaticity diagram



nevertheless have the same values for the tristimulus values X, Y, Z and therefore the same chromaticity. Such lights are called metamer; they have different spectral power distributions but give the same color impression. It should be noted that this metameric character applies within the boundaries of the CIE XYZ system.

Metamerism of Object Colors

The spectral composition of a color stimulus of an object color can be described by:

$$E_R = \beta(\lambda) * E(\lambda) \quad (6)$$

with

$E(\lambda)$: spectral power distribution of the light that illuminates

$\beta(\lambda)$: spectral reflectance of the object (fluorescent objects will be neglected here)

The tristimulus values of this color stimulus are:

$$X = \sum \beta(\lambda) * E(\lambda) \cdot x(\lambda) \quad (7)$$

$$Y = \sum \beta(\lambda) * E(\lambda) \cdot y(\lambda) \quad (8)$$

$$Z = \sum \beta(\lambda) * E(\lambda) \cdot z(\lambda) \quad (9)$$

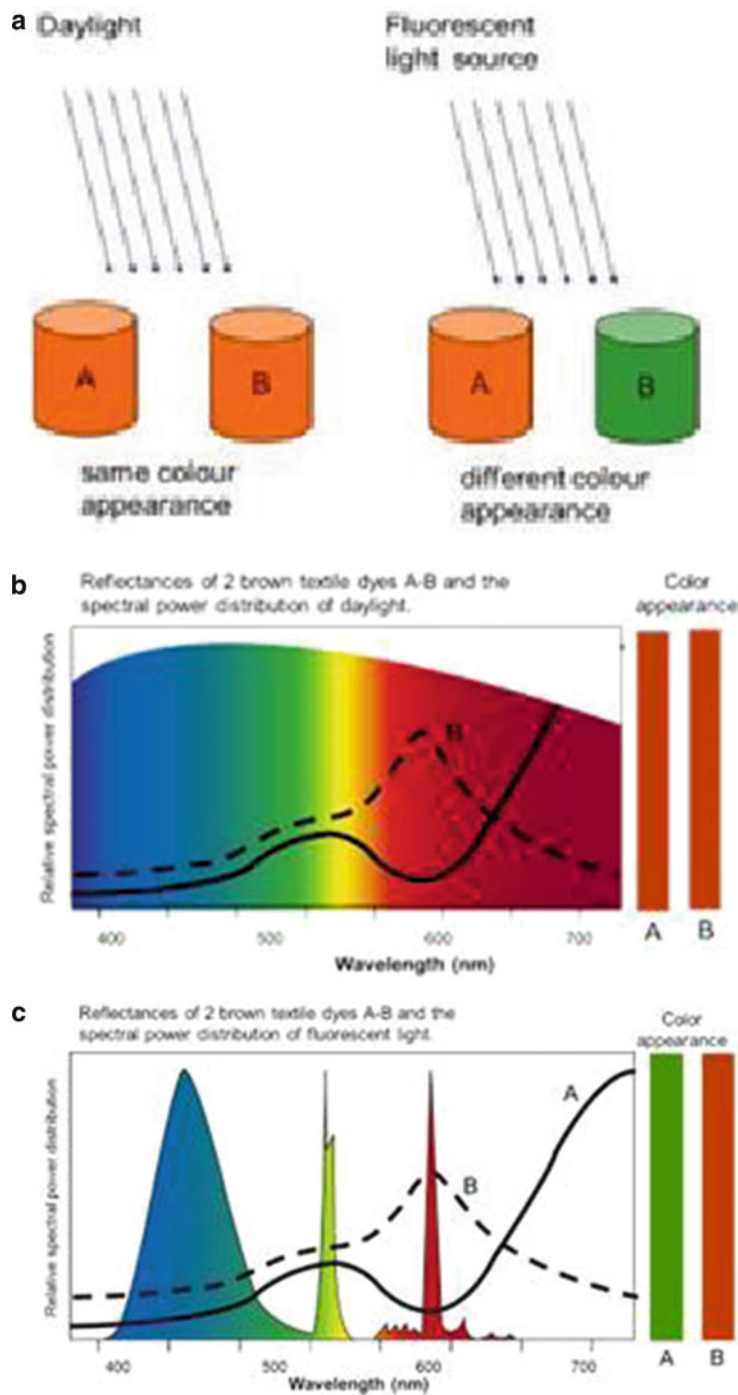
Suppose the amount of light on the object is decreased with a factor 10, so $E(\lambda) = 0.1E(\lambda)$. From experience, it is known that the human visual system adapts in such a way that the color impressions stay the same.

Now suppose that the spectral reflectance is reduced with a factor 10, so $\beta(\lambda) = 0.1 \beta(\lambda)$. Again from experience, it is known that the color impression will change: a white will turn into a gray, an orange color will turn brownish, etc.

How to account for these two observations? This can be done by defining the tristimulus values of object colors by the following equations:

$$X = 100 * \sum \beta(\lambda) * E(\lambda) \cdot x(\lambda) / \sum E(\lambda) \cdot y(\lambda) \quad (10)$$

Metamerism,
Fig. 3 (a–c) Change of
light source leads to change
in color appearance



$$Y = 100 * \sum \beta(\lambda) * E(\lambda). y(\lambda) / \sum E(\lambda). y(\lambda) \quad (11)$$

$$Z = 100 * \sum \beta(\lambda) * E(\lambda). z(\lambda) / \sum E(\lambda). y(\lambda) \quad (12)$$

By defining the values like this, they have become invariant for multiplication of $E(\lambda)$ with a constant factor, so invariant for the light level.

At the same time, multiplication of $\beta(\lambda)$ with a constant factor will lead to accordingly changes in the tristimulus values of the object color.

Another way of doing this is by normalizing $E(\lambda)$ in such a way that

$$\sum E(\lambda) \cdot y(\lambda) = 100 \quad (13)$$

Normalizing $E(\lambda)$ like this enables the use of Eqs. 7, 8, and 9 to calculate the tristimulus values.

For light sources, the light color can be defined with only 2 coordinates, the chromaticity x, y .

For color stimuli of object colors, however, 3 coordinates are necessary. These can be X, Y, Z or x, y, Y or any other three-dimensional system.

Object colors will be metameric if the three tristimulus values of the colors will be equal.

Suppose two object colors have spectral reflectances $\beta_1(\lambda)$ and $\beta_2(\lambda)$. These object colors will be metameric under a light $E(\lambda)$, normalized like in Eq. 13, if

$$\begin{aligned} X_1 &= \sum \beta_1(\lambda) * E(\lambda) \cdot x(\lambda) \\ &= \sum \beta_2(\lambda) * E(\lambda) \cdot x(\lambda) = X_2 \end{aligned}$$

$$\begin{aligned} Y_1 &= \sum \beta_1(\lambda) * E(\lambda) \cdot y(\lambda) \\ &= \sum \beta_2(\lambda) * E(\lambda) \cdot y(\lambda) = Y_2 \end{aligned}$$

$$\begin{aligned} Z_1 &= \sum \beta_1(\lambda) * E(\lambda) \cdot z(\lambda) \\ &= \sum \beta_2(\lambda) * E(\lambda) \cdot z(\lambda) = Z_2 \end{aligned}$$

These object colors are metameric under the light with spectral power distribution $E(\lambda)$. It is clear that the metameric character depends on the light source. In general, with another light with a different SPD, the relations $X_1 = X_2$, $Y_1 = Y_2$, and $Z_1 = Z_2$ will no longer hold. This means that colors metameric under one light source, i.e., giving the same color impression, can look rather

different under a light source with a different SPD. This is illustrated in Fig. 3a–c. Two brown textiles with different reflectances will have the same color appearance in daylight, whereas in a specific fluorescent light, the color appearance will be different. In Fig. 3b, c, the reflectances of the colors are depicted together with the spectral power distributions of daylight and fluorescent light.

In the paint industry, metameric object colors are very well known. In this industry, a frequently occurring problem is to reproduce a color with other pigments than the ones used for the original color. In such cases, usually samples are sought that are metameric under daylight as well as incandescent light.

Cross-References

- [CIE Chromaticity Coordinates \(xyY\)](#)
- [CIE Chromaticity Diagrams, CIE purity, CIE Dominant Wavelength](#)
- [CIE Color-Rendering Index](#)
- [CIE Special Metamerism Index: Change in Observer](#)

References

1. ISO 11664-1:2007(E)/CIE S 014-1/E:2006: Joint ISO/CIE Standard: Colorimetry Part 1. CIE Standard Colorimetric Observers

Mie Solution

- [Rayleigh and Mie Scattering](#)

Mie Theory

- [Rayleigh and Mie Scattering](#)

Minimally Distinct Border

Michael H. Brill

Datacolor, Lawrenceville, NJ, USA

Definition

In an exactly abutting pair of spatially uniform colored regions, if one of the regions is varied in radiance until the border between the regions is least conspicuous to a human observer, that border becomes a minimally distinct border (MDB) and is seen as having maximal blur and minimal contrast. The term MDB is used both for the border itself and for the psychophysical method used to derive it.

Historical Beginnings

The minimally distinct border was first reported by Boynton and Kaiser [1] in connection with psychophysical experiments. From the beginning, it was realized that the condition of MDB occurs when the two abutting regions have the same luminance. Hence, the condition of MDB is best understood in terms of the properties of luminance. (See next section.)

However, the history of the MDB is not just about luminance [2, 3]. The MDB was first used in an attempt to quantify complete color differences regardless of their size. Previously, whereas small color differences could be quantified by discrimination thresholds, large color differences were assessed by magnitude estimation. Compared to magnitude estimation, the assessment of strength of a visual border seemed to offer greater precision. In such experiments, subjects were asked to define the achromatic contrast that they deemed as equivalent to the chromatic contrast of a given border. When the radiances were titrated so as to produce a condition of MDB – hence equality of luminance – one could then look at the strength of the border as a function of the other color

coordinates. Historically, MDB experiments tended to compare a monochromatic light with a reference white, so it was thought that the MDB in such cases quantified saturation or departure from the white. As technology developed to present more light at shorter wavelengths (470 nm and lower), it was observed that the MDB could quantify only a “tritanopic saturation,” because no short-wave-sensitive (blue-sensitive) photoreceptors contributed to MDB. In fact, the MDB was observed to vanish completely for lights that were the same in all but their short-wavelength content – hence the name “tritanopic,” which derives from the nomenclature of color blindness. To quote Boynton [2]: “The blue-sensitive cones [...] seem to be free of any serious spatial or temporal responsibilities in vision.”

Luminance and MDB

Luminance of a light was initially thought to be linear in the light’s radiance and to transform perceptually to its brightness. Following this definition, two spectrally different lights were declared to have the same luminance if the radiance of one light was adjusted so its brightness matched the brightness of the other light. Early in the twentieth century, a more precise definition for luminance was devised: Two lights are equally luminous if, when the lights are presented alternately, the perceived flicker vanishes at the lowest frequency of alternation. The practice of finding equally luminous lights using this definition is called flicker photometry.

To be in harmony with both these definitions, in 1924, the Commission Internationale de l’Eclairage (CIE) developed a hybrid definition based on experiments of both sorts: flicker fusion of two alternating lights and direct brightness comparison of two lights at neighboring wavelengths [4]. As defined by the CIE, the luminance of any light is proportional to the integral of its spectral power distribution weighted by a function $V(\lambda)$ (called the luminosity function) derived from these experiments.

In the next 30 years, it was discovered that $V(\lambda)$ could not capture all the facts about iso-luminous lights. For one thing, Deane Judd found in 1951 that $V(\lambda)$ should be modified to have a larger contribution at short wavelengths. Also, the physical size of the light stimulus (field size) was found to have an influence on luminosity, as testified by the 1964 CIE colorimetric standard that was tailored to a 10° field size rather than to the 2° field size previously used. Finally, the luminosity of lights at low (scotopic) radiance follows the spectral sensitivity of the rods rather than the cones of the visual system. The low-light limit of luminance also entailed a transition to the high-light limit through a radiance domain known as the mesopic, in which both rods and cones operate.

Clearly the 1924 CIE definition of luminosity, although ubiquitous in industrial and other practical usage, diverges from what is found in actual visual systems. Therefore, vision researchers prefer to define luminance according to an empirical criterion, such as implied by flicker photometry. Once this definition is adopted, it is found that various other visual phenomena depend only on luminance of the stimulating light. In particular, visual acuity is a function only of luminance. In fact, the evidence supports the characterization of luminance as a channel in human vision, fed by long- and medium-wavelength-sensitive cones, occasionally by rods, and never by the short-wavelength-sensitive cones. The luminance channel is tuned to stimuli of the highest spatial and temporal frequencies perceivable by the visual system.

Using the vision-science definition of luminance described above, the condition of MDB between two abutting light patches occurs when the patches have the same luminance.

Measuring MDB

An MDB experiment finds the relative luminance of a light field as follows. For a fixed reference field, the MDB radiance of a monochromatic test field at wavelength λ is measured. The ratio of the reference to test radiance at MDB is proportional to the test field's luminance as a function of wavelength.

The MDB used in this way can be an empirical definition of luminance. In fact, because MDB presents the test and reference stimulus simultaneously, it has been touted as superior to flicker photometry as a method of measuring luminance.

MDB studies have been done over a variety of geometric conditions. Typically, the experiments involve a bipartite field of two semicircles (test field and reference field). The semicircles meet at a diameter on which the subject is directed to fixate. Alternatively, if one wants to compare the strength of an MDB with the strength of an achromatic border, one can divide the field into quarters of a circle, the top two comprising the achromatic border and the bottom two comprising the chromatic MDB.

It should be noted that creating the geometric conditions for MDB is not trivial. For example, an achromatizing lens is recommended to remove chromatic aberration, which would otherwise disturb the edge between the abutting fields.

Physiological Basis

The measurement of MDB was possible in macaque retina under several experimental conditions, most notably when the border was moved back and forth in the macaque's visual field. Two kinds of retinal mechanisms, fast (phasic) and slow (tonic), were examined as possible vehicles for MDB, and evidence favored the phasic mechanism [5].

As with any psychophysical measure of luminance, additivity (i.e., Abney's law) has been found to be imperfect, but this is hardly surprising since the luminance channel in vision receives inputs from more than one cone channel, and cone response is nonlinear in light intensity. Furthermore, the low-luminance limit must engage the rods as well as the cones, so additivity cannot be a strict property of MDB.

Extension to Mesopic Photometry

Raphael and MacLeod [6] discovered that equating luminance at various mesopic radiance levels

(in which both rods and cones enter the function) is correlated strongly with achieving MDB at these levels.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [Chromatic Contrast Sensitivity](#)
- [Luminance Meter](#)
- [Pattern-Induced Flicker Colors](#)

References

1. Boynton, R.M., Kaiser, P.K.: Vision: the additivity law made to work for heterochromatic photometry with bipartite fields. *Science* **161**, 366–368 (1968)
2. Boynton, R.M.: Ten years of research with the minimally distinct border. In: Armington, J.C., Krauskopf, J., Wooten, B.R. (eds.) *Visual Psychophysics and Physiology*. Academic, New York (1978)
3. Boynton, R.M.: *Human Color Vision*. Holt, Rinehart, Winston, New York (1979)
4. Wyszecki, G., Stiles, W.S.: *Color Science*, 2nd edn. Wiley, New York (1982)
5. Kaiser, P.K., Lee, B.B., Martin, P.R., Valberg, A.: The physiological basis of the minimally distinct border demonstrated in the retinal ganglion cells in the macaque retina. *J. Physiol.* **422**, 153–183 (1990)
6. Raphael, S., MacLeod, D.I.A.: Minimally distinct borders in mesopic vision. *J. Vis.* **8**(17), article 75 (2008)

Mixed Adaptation Condition

- [CIE Guidelines for Mixed Mode Illumination: Summary and Related Work](#)

Mixed Chromatic Adaptation

- [CIE Guidelines for Mixed Mode Illumination: Summary and Related Work](#)

Modern Art and Fashion Design

- [Art and Fashion Color Design](#)

Monochromator

Yi-Chen Chuang

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

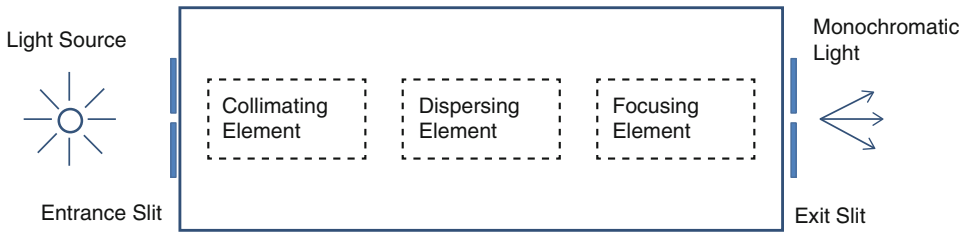
Definition

A monochromator is an optical dispersing device that is used to select a narrow band of light (i.e., optical radiation) from a wider range of wavelengths available at the input. The Greek roots “mono-” and “chroma-” refer to “single” and “color” respectively. Ideally, a monochromator should produce a single wavelength of optical radiation at its output. Although lasers produce light that is much more monochromatic than the optical monochromators discussed here, only some lasers are easily tunable but are not as simple to use.

Monochromators are included in many optical measurement instruments and systems for applications where tunable monochromatic light is required. A monochromator combined with optical detectors can be used to obtain the spectral power distribution (SPD) of light sources, reflectance or transmittance of objects, etc.

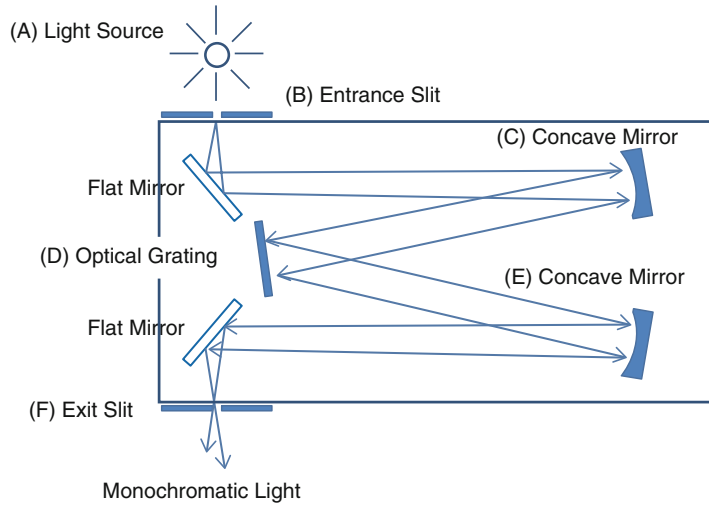
Monochromators are important for color measurement because many color-related optical characteristics are dependent on wavelength. In color science, reflected or transmitted light from a sample is usually measured, where a monochromator can be used before or after the incident light illuminates a sample. If a monochromator is used before the incident light illuminates the sample, the incident light is monochromatic light. On the other hand, if a monochromator is used after the light illuminates the sample, the monochromator is used to convert the reflected or transmitted light to monochromatic signals for analysis. It is typically used in a spectrometer (or spectroradiometer) or a spectrophotometer.

There are different types of monochromator based on its color selection mechanisms and/or designs, e.g., prism, Czerny–Turner, holographic grating, double, etc. The principles of



Monochromator, Fig. 1 Scheme of a simple monochromator

Monochromator, Fig. 2 Basic structure of a Czerny–Turner monochromator



these monochromators are introduced in the “[Overview](#)” section below.

Overview

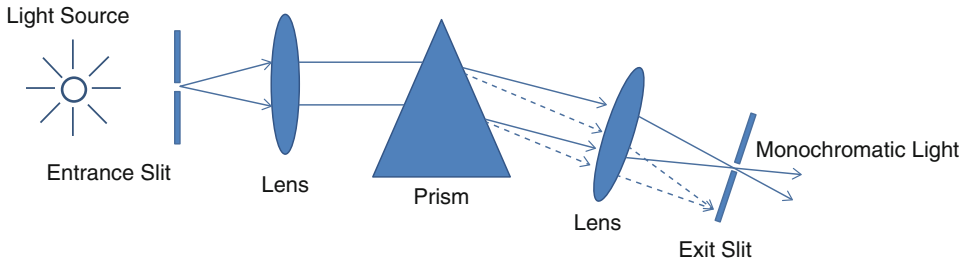
The main function of a monochromator is to separate the color components of a light. It can use either the optical dispersion phenomenon in a prism or that in a diffraction grating.

Figure 1 shows a scheme of a simple monochromator. Various types of monochromator have been developed, but a monochromator usually contains an entrance slit, an essential dispersing element, and a mechanism to direct the selected color to an exit slit for color selection purposes [1].

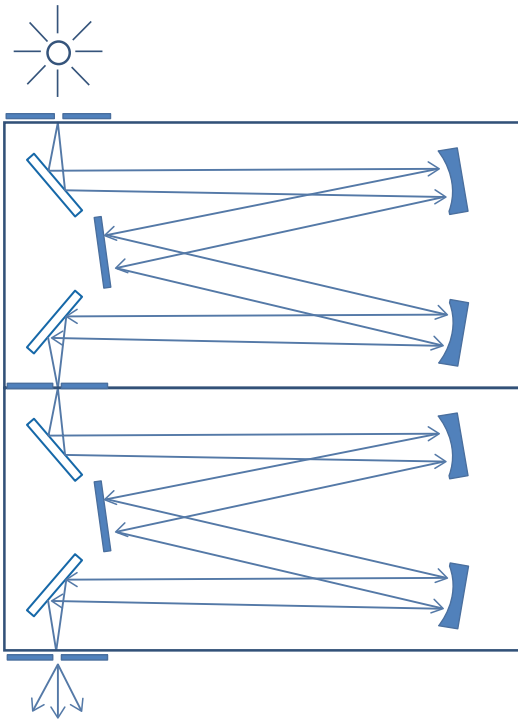
The Czerny–Turner monochromator is a type of monochromator using diffractive gratings as the dispersing element. A Czerny–Turner monochromator is sometimes referred to as a single

monochromator because the radiant flux is only diffracted once [2]. A schematic diagram of a Czerny–Turner monochromator is shown in Fig. 2. In Fig. 2, the light under measurement (A) is incident through the entrance slit (B), reflected by a concave mirror (C) to become collimated. It then illuminates the optical grating (D), disperses light to different colors, reflected and focused by the concave mirror (E) to the exit slit (F). By rotating the optical grating (D), different colors of light can reach and exit at the exit slit (F), and such colors are “selected” by the monochromator. The orientation of the optical grating (D) and the spatial location of the exit slit (F) determine which wavelengths of the light are selected. The two concave mirrors (C and E) image the entrance slit on the exit plane. In practice, the collimating and focusing elements (C and E) are usually the same mirror, used twice [2].

The double monochromator usually contains two Czerny–Turner monochromators used in a



Monochromator, Fig. 3 An example of a prism monochromator



Monochromator, Fig. 4 Double monochromator

system. Double monochromators are known for their double dispersion and thus can significantly reduce scattered light (Fig. 4) [2].

A prism monochromator uses a prism as the dispersing element (Fig. 3). It is not necessary for the collimating and focusing elements to be concave mirrors, but they can be optical lens. Usually, reflective mirrors are preferred over optical lens because mirrors conserve more optical radiation energy during the transmission process. However, because of the higher cost and small angular dispersion, prism monochromators are rarely made today [2].

A holographic grating monochromator uses holographically constructed concave gratings to simplify the optics. This eliminates some of the lenses and mirrors that would normally be required with conventional flat plane grating monochromators.

When choosing a monochromator, the following specifications should be considered: wavelength range, wavelength accuracy, measurement speeds, spectral bandwidth, spectral scattering, slit widths, grating groove density (if applicable), mechanical and thermal stability, size and weight, stray light, and dispersion [3, 4]. For example, Table 1 shows the criteria for the selection of gratings according to the above specifications. The resolution is the minimum detectable difference between two spectral peaks provided by a monochromator. Theoretically, resolution is approximately equal to slit width (mm) \times dispersion (nm/mm), where the dispersion describes how well the light spectrum is spread over the focal plane of the exit slit. However, in practice, because of aberrations in monochromators, the actual resolution is rarely equal to this. The actual resolution of a monochromator as a function of the slit width is described as the bandwidth (nm) of a monochromator.

When using a monochromator, some parameters need to be set up properly, as they affect the measurement results. For example, smaller slit widths produce smaller spectral bandwidths, which corresponds to a narrower spectral band at the exit slit (i.e., closer to the ideal single-wavelength output). However, smaller slit widths mean that less optical radiation power is transmitted through the monochromator, which results in lower output signals and decreases the signal-to-

Monochromator, Table 1 The criteria for the selection of a grating [4]

Specification of monochromators	Criterion
<i>Groove density (or groove frequency):</i> The number of grooves contained on a grating surface (lines/mm)	Groove density affects the mechanical scanning range and the dispersion properties of a system. It is an important factor in determining the resolution capabilities of a monochromator. A higher groove density results in greater dispersion and a higher resolution Select a grating that delivers the required dispersion, such as that for a charge-coupled device (CCD), array detector, or a required resolution (with an appropriate slit width) when using a monochromator
<i>Mechanical scanning range:</i> The wavelength region in which an instrument can operate	Refers to the mechanical rotation capability (not the operating or optimal range) of a grating drive system with a specific grating installed Select a grating groove density that allows operation over the desired wavelength region
<i>Blaze wavelength:</i> The angle in which the grooves are formed with regard to the normal grating, often called the blaze angle	Diffraction grating efficiency plays an important role in spectrograph throughput. Efficiency at a particular wavelength is largely a function of the blaze wavelength if the grating is ruled, or a modulation if the grating is holographic Select a blaze wavelength that encompasses the total wavelength region of the applications. More attention should be paid to covering the short wavelength of the spectrum region
<i>Quantum wavelength range:</i> The wavelength region of highest efficiency for a particular grating	Normally determined by the blaze wavelength Select a grating with maximum efficiency over the required wavelength region for the applications

noise ratio of the signals measured. The bandwidth of commonly used monochromators is from the order of 10^{-2} to 10 nm. Detailed information on monochromators can be found in manufacturers' user manuals.

It is important to evaluate the performance of a monochromator to reduce measurement uncertainties (i.e., gain more confidence regarding the accuracy of the results measured). To evaluate the performance of a monochromator, the following characteristics should be considered: wavelength accuracy, spectral bandwidth, stray light, system reproducibility, and system linearity.

Cross-References

- [Instrument: Spectrophotometer](#)
- [Spectral Power Distribution](#)
- [Spectrometer](#)
- [Spectroradiometer](#)

References

1. Webster, J.G. (ed.): The Measurement, Instrumentation, and Sensors Handbook. Springer, Heidelberg (1999)
2. Christensen, R.L., Potter, R.J.: Double monochromator systems. *Appl. Optics* **2**, 1049 (1963)
3. Kostkowski, H.J.: Reliable Spectroradiometry. Spectroradiometry Consulting, La Plata (1997)
4. Yoshizawa, T. (ed.): Handbook of Optical Metrology: Principles and Applications. CRC Press, Boca Raton (2009)

Mosaics

Gertrud Olsson
School of Architecture, KTH, Royal Institute of Technology, Stockholm, Sweden

Synonyms

[Tesselations](#); [Tesserae](#)

Definition

Arrangement of small fragments of stones, colored glasses, golden pieces, or other materials,

fixed on walls or pavements and depicting an image or making an abstract pattern. The technique of mosaic is a decorative art that reached a peak during the Byzantine period of Christian art.

Overview

The traditional mosaics consist of marbles and stones in the scale of black and white (sometimes together with red) or other materials which were available in nature. At the beginning, in the Antique age, mosaics were used only as pavements, as stones on the ground, a floor material. During the first centuries of Christianity, during the Byzantine period, the Church took over the mosaic technique. The Byzantine technique with walls done with mosaics was new in the fifth and sixth centuries. In the basilicas of Ravenna, Italy, the Byzantine mosaic masters worked with entire walls of mosaic made up of small pieces of colored glass and gold, *tesserae*, affixed to mortar. *Tesserae*, a term derived from a Greek word meaning “four-sided,” is the standard for mosaic pieces. Before being divided, the tesserae are part of a blown flat glass in the form of a “pizza plate” (in today’s terminology). *Smalti* is the technical term for the brilliant, opaque-colored crystalline material fused with glass.

The new materials that the Byzantine mosaic masters introduced were colored glass and gold. These materials offer a wider and more intense color scale than the marble mosaics. The tesserae-mosaic is a kind of color material that is mixed optically (on the retina of the eye). The material interacts in a special way with vision and light. The surface is shiny, hard, and reflective. Glass pieces are shiny and slightly irregular in pure colors and gold. The gold picks up light and reflects gold light back into the room. The entire composition involved the setting of mosaic pieces at different angles so as to reflect light as effectively as possible. Seen from a distance, the singular mosaic pieces appear to model figurative forms. In the present time, this technique is considered as somewhat comparable

to the compositions of the Neoimpressionist painters of the 1880s. Thus, the mosaic masters possessed knowledge of the eye’s ability to apprehend color mixing. They also acquired understanding of the changes colors underwent with distance and also of the interaction of light and material.

The images in the Byzantine basilicas are often constructed from contrasting colors. Red stands against green in costumes and clothing designs. A pair of red shoes usually meets green grass. When two contrasting colors meet, there is an increase in both colors. Another approach in the sixth century, used by the mosaic masters to avoid changes in color, was to apply a contour line between two colors. This contour line prevents the colors from spilling over into each other. Instead, the colors are depicted without visible distortion. The gold mosaic primarily represented the halos of the holy figures but was also placed in the background. The subject of the church mosaics was to serve the spirit, not the body. In the images, all suggestions of movement have been avoided; instead, an “eternal existence” is shown. The combined effect in the milieu of the church is a glittering, multicolored immaterial curtain. The material, the light, and the surface interact.

By studying the mosaics, it is possible to see how the mosaic masters applied the tessera pieces with skill and knowledge of color and what happens optically at a distance. The mosaic masters mixed colors optically for entire large areas of color. For instance, in order to give a lawn its various shades of green and thus get a nuanced and living surface, the mosaic pieces were placed against each other in turquoise, green, and gray-green for the purpose of combining and so forming the greenness of the grass. Gold mosaic was also used – in addition to create areas of gold color and point out important details – to mix the colors in larger fields.

When viewing a mosaic wall, the changes of the material from matte to shiny can be discerned. How the mosaic is reflected depends on its angle against the wall. Two pieces of the same inherent color placed with different tilt angles result in that each one will have a different ► [appearance](#). This

effect results from light projecting and reflecting different information to the eye. Reflection is also dependent on how the light strikes the mosaic pieces, on whether the tesserae are made of glass (shiny) or marble (often matte), on the viewing angle, and on how the viewer moves in the room. No movement of the subjects or the motifs on the walls is discernible. The hues are perceived as stable and unchangeable, and unaffected by nearby colors. But with the movement of the viewer in the room, and the changes in the eye's view, the mosaic walls are transformed from matte to shiny, and back again. Through the mosaic, the viewer participates in the process of change.

In the present time, it is possible to find the impact of late Antique mosaics in Sweden, for example. In 1919, the Swedish artist Einar Forseth made a study trip to Italy. The architect Ragnar Östberg designed the Stockholm City Hall in 1923, and Einar Forseth was appointed to create the Golden Hall. Forseth chose mosaic for his commission. The Golden Hall is thus covered with more than 18 million glass and gold mosaic pieces. The mosaics depict Swedish historical figures and interpret certain events in the Swedish history, but the technique, the imaging, and the colors are in Byzantine style with its origins in Ravenna. Besides the Stockholm City Hall, the Byzantine technique can be seen in an underground station situated in Rinkeby, a suburb of Stockholm. In his adornment from 1975, the artist Nisse Zetterberg used mosaic in gold, in the tesserae technique, applied with different angles to the background. The mosaics depict patterns, symbols, and runes inspired by the Viking age, once found in the Rinkeby excavations.

Cross-References

- ▶ [Appearance](#)
- ▶ [Color Contrast](#)
- ▶ [Neo-impressionism](#)
- ▶ [Unique Hues](#)

Motion and Color Cognition

Donald Hoffman

Department of Cognitive Sciences, University of California at Irvine, Irvine, CA, USA

Synonyms

[Achromatopsia](#); [Color from motion](#); [Dynamic color spreading](#); [Neon color spreading](#)

Definition

Motion and color cognition is a branch of cognitive neuroscience devoted to using brain imaging, psychophysical experiments, and computational modeling to understand the interactions between motion and color processing in the human visual system.

Overview

The relationships between the perceptions of color and visual motion are complex and intriguing. Evidence from patients with certain rare forms of damage to the cerebral cortex indicates that color and motion are to some extent processed separately within the visual system. This conclusion is buttressed by physiological evidence for separate neural pathways for the processing of color and motion and psychophysical evidence that the perception of motion in color displays is greatly reduced when the colors are isoluminant. However, experiments and demonstrations reveal that the perceptions of color and motion can interact. Motion can trigger the perception of color in achromatic stimuli and in uncolored regions of space. Color can determine the perceived direction of motion, and changes in color can trigger the perception of apparent motion. This article briefly reviews the evidence for separate processing of color and visual motion and for the ways they interact.

Evidence for Separate Processing of Color and Motion

Patients with the rare condition of *cerebral achromatopsia* cannot perceive colors, due to damage of cortex in the inferior temporal lobe of both cerebral hemispheres [1, 2]. Cerebral achromatopsia differs from the more common condition of color blindness, in which one or more of the normal three cone systems in the retinas are missing. A color-blind person missing the long-wavelength cone system, for instance, can still perceive many colors but has a range of color perceptions more limited than that of a normal trichromat, who has all three cone systems. Cerebral achromatopsics can have all three cones systems intact in their retinas but are still unable to perceive any colors because of damage to visual cortex. Yet many cerebral achromatopsics have no difficulty perceiving visual motion, indicating that the processing of color within the human visual system can be damaged without also damaging the processing of visual motion. (See also the Encyclopedia of Color Science and Technology entry on color naming and categorization in inherited color deficiencies and the entry on color perception and environmental impairments.)

Conversely, patients with the rare condition of *cerebral akinetopsia* cannot perceive motion, due to damage of cortex in the middle temporal area of both cerebral hemispheres [2, 3]. According to Josef Zihl and his colleagues, one such patient "...had difficulty, for example, in pouring tea or coffee into a cup because the fluid appeared to be frozen, like a glacier. In addition, she could not stop pouring at the right time since she was unable to perceive the movement in the cup (or a pot) when the fluid rose" [2, 4]. Yet this patient was able to perceive color, indicating that the processing of motion within the human visual system can be damaged without also damaging the processing of color.

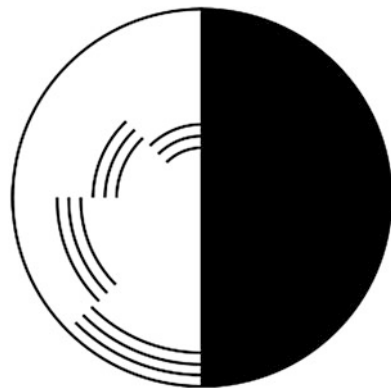
Cerebral achromatopsia and cerebral akinetopsia indicate that color and motion are to some extent processed separately within the visual system. Further evidence for their separate processing comes from neuroanatomical and

neurophysiological studies revealing that the visual system has a parvocellular pathway that processes color and detailed form information and a magnocellular pathway that processes motion, depth, and coarse form information [5]. These pathways are clearly segregated in the lateral geniculate nucleus of the thalamus, a relay station for visual information traveling from the eyes to visual cortex. One psychophysical correlate of this separate processing is the observation that perceived motion in color displays is greatly reduced when the displays are made isoluminant, an effect first discovered in 1911 by Pleikart Stumpf [6].

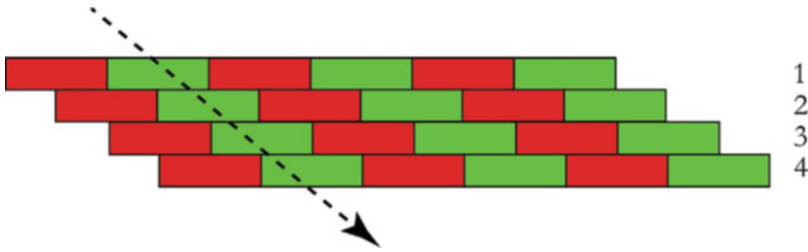
Evidence for Interactions in the Processing of Color and Motion

However, despite the anatomical, neuropsychological, and psychophysical evidence for the separate processing of motion and color, there are several intriguing visual phenomena which clearly demonstrate that the perception of motion and color can interact. An early example was discovered by the French monk Benedict Prevost in 1826, when he noticed that if he waved his hands in the dim light of the cloisters, then colors appeared near his moving fingers [7].

Charles Benham marketed a top in 1895 with a pattern similar to that shown in Fig. 1 [8]. Spinning



Motion and Color Cognition, Fig. 1 Benham's top. Spinning this achromatic pattern leads to perceived colors



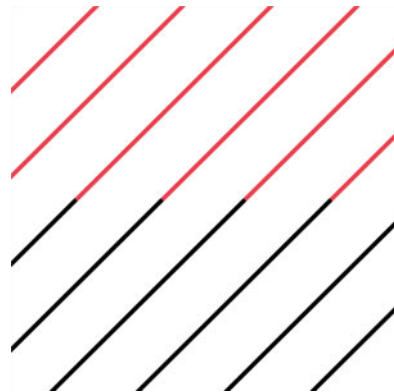
Motion and Color Cognition, Fig. 2 Color and motion correspondence. This figure depicts four frames of a movie in which the colored blocks could, in principle, be

perceived either as moving to the left or as moving to the right. Color information, even at isoluminance, biases us to see movement to the right

the top leads to the perception of the colors of an artificial spectrum, again demonstrating an interaction between color and motion. (See also the Encyclopedia of Color Science and Technology entry on illusory colors.)

Karen Dobkins and Tom Albright discovered in 1993 that color can influence the perceived direction of motion [2]. They showed subjects a movie in which a row of red and green blocks appeared to move horizontally. Four frames from such a movie are shown schematically in Fig. 2. From one frame to the next, the blocks moved exactly one half of a block width, so that the movie is compatible, in principle, with being seen as blocks moving to the left or to the right. However, if the brightness of the green blocks differs from that of the red blocks, then the visual system sees a direction of motion consistent with this, i.e., such that the brightness of each block does not appear to change as it moves. However, if all the blocks are equal in perceived brightness, the visual system sees a direction of motion consistent with the colors, i.e., such that the color of each block does not appear to change as it moves. Thus, color information by itself, apart from luminance information, can influence the perceived direction of motion. Similar evidence for the interaction of color and motion comes from experiments showing that color can affect the perceived direction of discrete apparent motion stimuli [9] and of sine-wave gratings [10, 11].

Conversely, V.S. Ramachandran discovered in 1987 that motion can “capture” color borders. If an illusory contour or a group of random dots are moved in the vicinity of a color border, then that

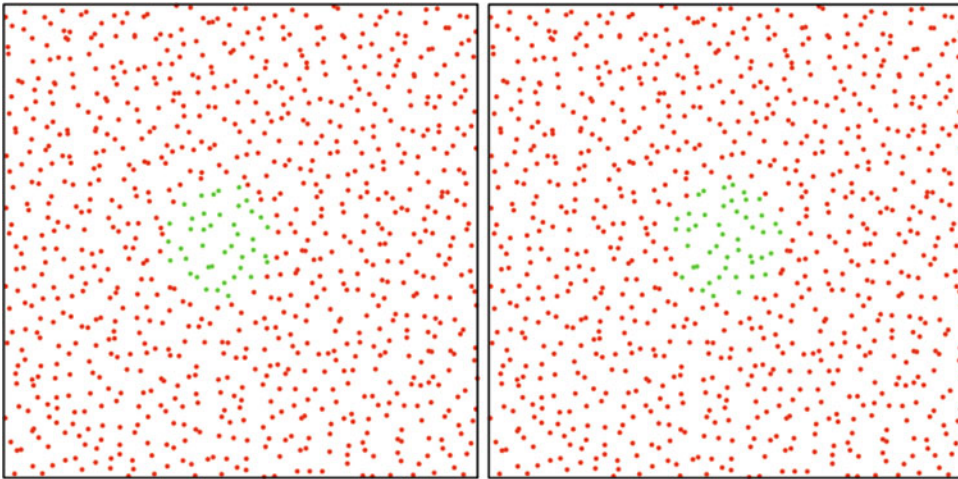


Motion and Color Cognition, Fig. 3 Wallach's display. The red neon color spreading seen in the upper half of the square becomes more pronounced when the lines are moved horizontally, but appear to be moving vertically

border appears to move in the same direction as the illusory contour or random dots [12].

Hans Wallach in 1935 created a display similar to that shown in Fig. 3 [13]. As can be seen in the figure, the oblique lines are red in the upper half of the display and black in the lower half. The red color of the lines in the upper half appears to spread faintly through the white regions, a phenomenon known as neon color spreading.

Wallach then put the oblique lines in motion, translating them sideways underneath a rectangular window, indicated in Fig. 3 by a gray frame. This leads to a bistable perception of the motion of the lines. Sometimes observers see them translating horizontally, as in fact they are. But sometimes, due to the ambiguities of motion seen through small apertures, observers see the lines translating vertically. In this latter case, observers



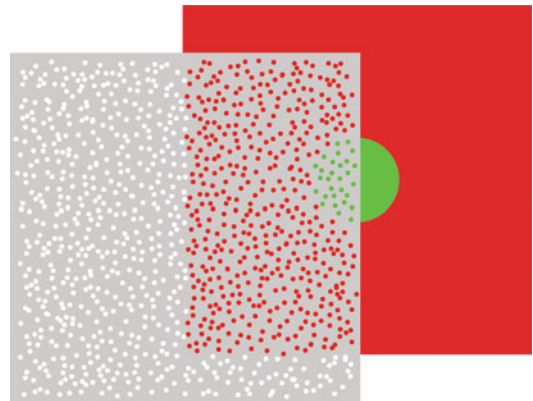
Motion and Color Cognition, Fig. 4 Two frames from a movie display of dynamic color spreading. These frames can be stereo fused to get an impression of a glowing *green* disk hovering over a field of *red* dots

report that the red neon color spreading is greatly enhanced. Thus motion can interact with color by enhancing neon color spreading.

A similar interaction between color and motion can be seen in displays of dynamic color spreading [2, 14]. Two frames from such a display are shown in Fig. 4. Each frame has a few hundred dots, which do not change position from frame to frame. Some of the dots are colored red and others, within a circular region, are colored green. The circular region in which dots are colored green moves slightly from frame to frame. This leads to the perception of a glowing green disk that moves over a field of static red dots. One can get feel for the perception of the glowing disk by using Fig. 4 as a stereo display, and cross fusing the two images. One sees a glowing green disk floating in front of the field of red dots.

Sometimes the green disk does not appear to float in front of a field of red dots. Instead, the red dots appear to be holes in a white sheet of paper, and through the holes one sees a solid green disk moving over a uniformly colored red background. This “amodal” perception of dynamic color spreading is illustrated in Fig. 5. One can get a feel for this amodal perception by again treating Fig. 4 as a stereo display, but this time fusing the two images by diverging the eyes rather than crossing them.

Displays of dynamic color spreading demonstrate that apparent motion can trigger the creation



Motion and Color Cognition, Fig. 5 Dynamic color spreading seen amodally. The *red* dots appear to be holes in a sheet of paper, and through the holes one sees a *green* disk moving over a *red* background

of illusory objects and the spreading of color over the surfaces of such illusory objects. The surface color can have a neon quality if the object is seen floating in front of the dots or a paint quality if the object is seen amodally.

Summary

Evidence from neuropsychology, neurophysiology, and psychophysics indicates that color and

motion are to some extent processed in separate pathways of the visual system, but that these visual pathways interact in a sophisticated fashion, and that together they can create the appearance of objects in motion with colored surfaces.

Cross-References

- [Behnam's Disk](#)
- [Color Phenomenology](#)
- [Color Processing, Cortical](#)
- [Color Psychology](#)
- [Subjective Colors](#)

References

1. Sacks, O.: *An Anthropologist on Mars*. Vintage Books, New York (1995)
2. Hoffman, D.D.: *Visual Intelligence: How We Create What We See*. W.W. Norton, New York (2000)
3. Zeki, S.: Cerebral akinetopsia (visual motion blindness): a review. *Brain* **114**, 811–824 (1991)
4. Zihl, J., von Cramon, D., Mai, N.: Selective disturbance of movement vision after bilateral brain damage. *Brain* **106**, 313–340 (1983)
5. Carlson, N.J.: *Physiology of Behavior*: Tenth Edition. Allyn & Bacon, New York (2009)
6. Todorovic, D.: A gem from the past: Pleikart Stumpf's (1911) anticipation of the aperture problem, Reichardt detectors, and perceived motion loss at equiluminance. *Perception* **25**, 1235–1242 (1996)
7. Cohen, J., Gordon, D.A.: The Prevost-Fechner-Benham subjective colors. *Psychol. Bull.* **46**, 97–136 (1949)
8. von Campenhausen, C., Schramme, J.: 100 years of Benham's top in colour science. *Perception* **24**(6), 695–717 (1995)
9. Papathomas, T.V., Gorea, A., Julesz, B.: Two carriers for motion perception: color and luminance. *Vis. Res.* **31**, 1883–1891 (1991)
10. Mullen, K.T., Boulton, J.C.: Absence of smooth motion perception in color vision. *Vis. Res.* **32**, 483–488 (1992)
11. Chichilnisky, E.J., Heeger, D., Wandell, B.A.: Functional segregation of color and motion perception examined in motion nulling. *Vis. Res.* **33**, 2113–2125 (1993)
12. Ramachandran, V.S.: Interaction between colour and motion in human vision. *Nature* **328**, 645–647 (1987)
13. Wallach, H.: Ueber Visuell Wahrgenommene Bewegungsrichtung. *Psychol. Forsch.* **20**, 325–380 (1935)
14. Hoffman, D.D.: The interaction of colour and motion. In: Heyer, D., Mausfeld, R. (eds.) *Colour Perception: Mind and the Physical World*. pp. 361–377. Oxford University Press, Oxford (2003).

mRGCs

- [Melanopsin Retinal Ganglion Cells](#)

Multilingual/Bilingual Color Naming/ Categories

Nancy Alvarado

Department of Psychology and Sociology,
California State Polytechnic University, Pomona,
Pomona, CA, USA

Definition

A bilingual person is a person fluent in more than one language, either from birth or acquired later in life. A multilingual person is fluent in more than two languages. Color naming is the process of assigning color terms to refer to color appearances in the world. Color categories are mental representations of the regions of color space that are designated by particular color terms or names.

Empirical Comparisons

The influence of culture on color perception and categorization is most often studied by comparing the use of single-word (monolexemic) color terms to name color categories across different languages, as in the World Color Survey [1, 2]. Color categorization and naming have been studied across a wide range of cultures to explore questions about the universality of basic color categories, relativism of color naming and/or perception, and the ways in which color perception contribute to color naming. However, such comparisons between groups of individuals speaking different languages in different cultures necessarily confound differences in individual experience and individual perception with the linguistic properties of the languages spoken and the categorization practices of the cultures themselves, resulting in a difficulty identifying the distinct contribution

of each source of influence on color-naming behavior. Comparison of the categorization and naming behavior of multilingual or bilingual individuals using each of their distinct languages provides a way of controlling for individual differences so that the influence of culture and linguistics on naming can be more readily observed.

Comparisons of multilingual/bilingual individuals performing identical color-naming or observation tasks separately in each of their languages depend on the level of fluency of the individual in each language. In some cases, a person may know two or more languages because he or she is an immigrant to a new culture, in the process of acquiring a new language while using the native language or previously acquired languages less frequently. In other cases, a person is a participant in a multicultural society in which different languages are spoken in different contexts and for different purposes but are generally spoken frequently enough to be refreshed in memory. The sequence in which languages were acquired and the relative proficiency in each language are important and should be measured in studies of multilingual or bilingual color naming. An individual in the process of assimilation to a new culture and language may experience interference from previous languages. He or she may be motivated to forget or experience pressure to forget previously acquired languages, affecting naming behavior. Studies comparing bilingual naming to monolingual naming show increased consensus of response among bilinguals, as well as a preference for use of more basic terms, perhaps reflecting a focus on what is shared across languages or perhaps a vocabulary loss affecting the less frequently used, more differentiated terms needed to describe subtleties of color samples [3, 4]. Alvarado and Jameson [5] observed the same phenomenon for emotion terms. For this reason, bilingual color categorization and naming behavior should be compared to both monolingual immigrant and nonimmigrant populations. Few studies have made such comparisons, even in the literature on bilingualism.

Code-switching occurs when bilingual or multilingual individuals spontaneously change

languages while speaking. The switching suggests that some concept might be better expressed using a particular language or that access to vocabulary is easier in that language. Study of when and how this occurs is an emerging field in linguistics [6]. In comparisons of bilingual color naming, subjects are required to switch languages to perform the same task twice but are constrained to respond in one language. It is unclear how much code-switching may be occurring mentally as subjects perform tasks. Further, the requirement to respond in a single language will have an impact on speeded tasks that must be considered in interpretation of results. In studies comparing multilingual/bilingual subjects across languages, the focus is generally not on the factors affecting voluntary choice of terms but on how a subject's perceptual choice behavior or categorization and naming of a color sample are affected by use of a specific language. Researchers have asked whether perceptual boundaries match the categories existing within a language and whether perceptual experience changes to fit a second language's categories when a new language is acquired. As yet, no studies have investigated whether an individual might prefer to name colors using the language with the richer color vocabulary, the more differentiated category structure, or perhaps the terms most relevant to performing a particular task.

Languages vary in how many color terms they contain and in where they draw color category boundaries [2]. Languages with fewer basic terms using Berlin and Kay's evolutionary-stage classification system may have broader categories with different focal hues. For example, it is common for many Southeast Asian and Asian languages to combine the green and blue categories which are distinct in English into a single category (commonly referred to as a "*Grue*" category). The Tarahumara people of northern Mexico use far fewer basic terms and form broader categories using a lightness-based naming system. This does not imply that individuals using those languages will be unable to recognize differences between green and blue color samples nor that the Tarahumara cannot perceive hues. Asian languages with a single term for the broad green/blue

range of color samples use a basic term (such as *Xanh* for Grue in Vietnamese) combined with modifiers to specify the distinct hue (thus *Xanh lá Cây* for Leaf green in Vietnamese).

In general, when bilingual individuals perform color-naming tasks in two languages, their behavior suggests a dissociation between cognitive processing based on meaning (e.g., perceptual experience) and language choices dictated by grammar or language structure. Thus, a bilingual person who is assimilating to a new culture can make meaning-based choices based on one culture while using language-related naming behaviors typical of the other culture. For example, Pavlenko and Driagina [7] found that English-speaking bilinguals followed a Russian pattern of intransitive verb use when speaking in Russian but switched to an adjectival pattern in English. In contrast, Jameson and Alvarado [4] found that Vietnamese bilinguals tended to impose both the grammar and category structure of English onto Vietnamese responses when responding in Vietnamese. Thus, bilingual subjects responding in Vietnamese tended to name the category orange (a distinct category in English but not Vietnamese) using the object gloss (the fruit name) in both English and Vietnamese, whereas those monolingual in Vietnamese (in both the United States and Vietnam) named the same color sample using a modified basic term (e.g., *Vàng Dam* or dark yellow, with the adjective following the noun). On the other hand, native Vietnamese speakers responding in English were more likely to use Vietnamese language patterns such as more modified basic terms and compound terms (e.g., yellow-brown) than native speakers of English, who were more likely to use monolexemes (tan for yellowish-brown colors) and object glosses (Bark, Baby Puke). Thus bilinguals retained the language structure but not the meaning-based categorization patterns [4, 5]. Alvarado and Jameson [6] found a similar dissociation for emotion terms. It is possible this behavior may be specific to the domain of color terms, since Pavlenko [8] has identified several studies in which a domain of terms (in her case, emotion terms) has been treated differently than other categories of concrete and abstract nouns in a language. This suggests the

need for greater caution in use of language-based properties such as concreteness/abstraction as a definition for basic color status and consideration of how the content of a domain might interact with language structure. Within-individual comparisons of naming behavior can help identify such differences in language structure as applied to a specific domain of terms, such as color names.

Color surveys that have emphasized monolexemic naming in a search for single-word basic color terms have disadvantaged those languages employing other strategies for making distinctions between colors, including use of modifiers, suffixes and prefixes to stem terms, intensifiers such as word repetition, and use of multiple-character combinations to name basic colors in Asian languages with nonphonetic writing systems. These complications of grammar with meaning complicate cross-cultural comparisons used to form theories about basic color naming but can be clarified by examining behavior of bilingual subjects because meaning-based choices are affected differently than grammar-based choices.

As noted previously [5], MacLaury [9] suggests that Berlin and Kay's evolutionary stages (in which cultures are ranked according to the number and order of emergence of basic color terms) reflect a transition from a lightness-based naming system to a hue-based system. He also suggests that cultures evolve in their naming due to a shift in emphasis from describing similarity to describing difference. Schirillo [10] suggests that evolutionary stages may also reflect a transition from use of contextualized names to use of more abstract terms (basic terms are abstract by definition). Van Brakel [11] suggests that stages reflect the influence of Western culture on indigenous naming behaviors and are thus a transition to the Western color system from a variety of viable alternatives. However, in studies of unconstrained naming, where subjects are not confined to use of basic color terms, both English speakers and those speaking other languages greatly prefer to describe colors using secondary names or modified basic terms, not basic terms alone. If cultures were evolving toward a greater emphasis on describing difference, those cultures with more

basic terms should also have more modifiers. The opposite seems to be true – use of modifiers and use of contextualized basic terms seem to vary inversely, suggesting that these are alternative strategies for describing differences in color percepts.

The behavior of multilingual/bilingual speakers may recapitulate this evolutionary process on the individual level to the extent that immigrants speaking multiple languages are in transition to English from one or more other languages. It can be argued that generalizing from individual language learning to evolution of language stages at the cultural level is justified because (a) the culture's language practices are determined by the behavior of individuals speaking that language; (b) as with individuals, cultures change in part due to contact with other cultures; and (c) the culture's evolutionary stage is measured by assessing the language behaviors of participants sampled from that culture. Otherwise, the processes by which cultures change are not hypothesized to be the same as those by which individuals acquire language abilities. In support of this idea, Jameson and Alvarado [4] found that Vietnamese bilinguals used a basic term for orange while Vietnamese monolinguals did not and that the focal samples for bilinguals responding in Vietnamese had shifted toward those of monolingual English speakers. However, for these Vietnamese bilinguals, highly specific Vietnamese object glosses and modifiers tended to fall into disuse before basic color terms. Thus bilingual speakers were unable to fully emulate the pattern of naming used by monolingual English speakers because they did not have access to a large vocabulary of object glosses in Vietnamese (probably due to disuse). A comparison of multicultural multilingual naming with immigrant multilingual naming is needed to show whether those with richer language experience show an increased tendency to use their wider vocabularies to describe additional dimensions of perceptual experience.

Alvarado and Jameson [5] suggest that cultures may not evolve from lightness to hue but rather from a single-dimensional naming system (based on lightness) to a multidimensional naming

system capturing distinctions related to hue and saturation in addition to lightness. If so, the interaction between dimensions may produce more sharply defined category boundaries than a single dimension alone. In that case, modifier use should increase as cultures attend to additional dimensions. Further, modifiers tend to evolve along with basic terms. As noted, Tarahumara terms like *Very* and *Somewhat* evolve to *Light* and *Bright*, terms which are specific to color naming. Vietnamese uses the term *Fresh* (*Tươi*) to describe saturation, a modifier not used in English but equivalent to terms such as *Vivid* or *Deep*. Areas of color space described by lightness modifiers do not vary as much as those described by modifiers capturing other dimensions. This suggests that an evolutionary-stage theory based on a broader spectrum of naming behavior, capturing more aspects of spontaneous naming behavior, might produce a different ordering of stages.

Theoretical Considerations

Comparisons of naming behavior of multilingual individuals across languages provide an approach to testing assumptions of current theories of color naming, such as the assumption that monolexemic naming using basic color terms is a valid approach to comparing languages with widely varying strategies for describing color appearances. For example, as described by Burgess et al. [12], the Tarahumara language used a postposed bound modifier (a modifier added to the end of a stem term) that specifies the relation of the currently named color to the center of the category. This system is different from Chinese, Vietnamese, and English [13], where modifiers are separate words and relational distinctions are subordinated to other distinctions. In contrast, Chinese characters are frequently constructed by compounding several characters into a single character with a new meaning, and all color terms consist of two-character pairs in which the first character specifies the color and the second character specifies that it is a noun referring to the color appearance itself (decontextualized, abstract). When the color name is used as an adjective, the second

character is the name of the object taking that color. In Japanese, a different second character is used to differentiate chromatic and achromatic color names [13]. This complicates comparison of the basicness of terms when the focus is on monolexemic naming and on the properties of the language instead of the behavior of the individuals speaking that language. Emphasis on a few basic color categories, in both study and theory, neglects the richness of color-naming behavior.

It has been argued that immigrant monolingual speakers in the United States (or elsewhere) are not a suitable population for cross-cultural investigations. When monolingual speakers in immigrant communities do not acquire a new language, there appears to be a close similarity between their responses and those of monolingual speakers in their original country of residence. This may be most true for immigrants living in large insulated neighborhoods in urban areas who can conduct daily business without needing to speak an additional language. In two studies of different domains [4, 6], monolingual Vietnamese speakers in the United States and Saigon produced responses that were closely similar, suggesting that the groups responding in the United States were a close proxy for their counterparts in Vietnam. Cultural contamination cannot be automatically presumed to invalidate every investigation. Although it might also be argued that the availability of Western media may have changed the Vietnamese lexicon, resulting in the close similarity of results among the studies, differences were noted for certain contextualized terms which parallel the differences observed between the American and Vietnamese cultures themselves. For example, differences in categorization and naming of the category orange were observed for bilingual subjects but not for monolingual immigrants living in the United States, who would presumably have been exposed to Western media. Given that such cultural differences were not eliminated by exposure to the media, we suggest that perhaps everyday social interaction and communication have a greater impact on language than passive media exposure. Findings of the persistence of regional accents in Great Britain in the face of homogenous media speech support this

conclusion. To assess this, the level of assimilation should be measured in studies of multilingual/bilingual versus monolingual immigrant speakers.

Study of the cognitive language processing of multilingual/bilingual individuals in color-naming tasks has the potential to widen our understanding of how language is used to describe perceptual experience by providing a glimpse into the cognitive processing that can be attributed to meaning distinct from the processing that is related to language, holding individual differences and cultural differences constant. However, to conduct such studies, comparisons must be broadened to compare bilingual and monolingual subjects in both the native country and the country to which such subjects have emigrated, as well as comparisons between multilingual individuals within multicultural societies in which multiple languages are regularly spoken independent of immigrant experience.

Cross-References

- [Berlin and Kay Theory](#)
- [Comparative Color Categories](#)
- [Psychological Color Space and Color Terms](#)
- [World Color Survey](#)

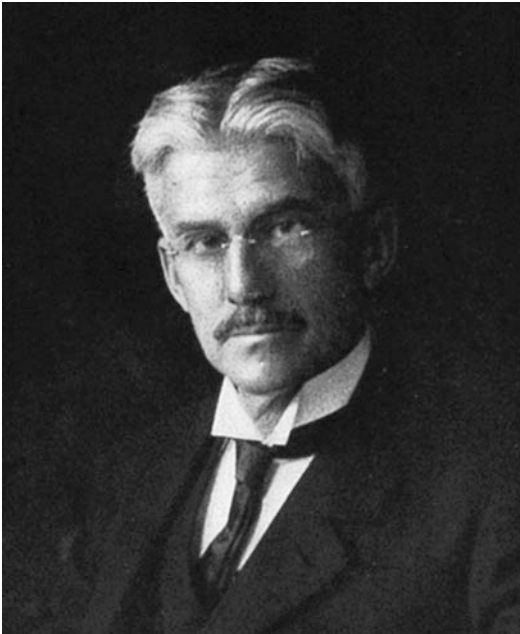
References

1. Kay, P., Berlin, B., Maffi, L., Merrifield, W., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009)
2. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
3. Jameson, K.A., Alvarado, N.: Differences in color naming and color salience in Vietnamese and English. *Color. Res. Appl.* **28**, 113–138 (2003)
4. Alvarado, N., Jameson, K.A.: The use of modifying terms in the naming and categorization of color appearances in Vietnamese and English. *J. Cogn. Cult.* **2**(1), 53–80 (2002)
5. Alvarado, N., Jameson, K.A.: Shared knowledge about emotion among Vietnamese and English bilingual and monolingual speakers. *J. Cross-Cult. Psychol.* **42**, 964–983 (2011)

6. Bullock, B.E., Toribio, A.J. (eds.): *The Cambridge Handbook of Linguistic Code-Switching*, 1st edn. Cambridge University Press, London (2009)
7. Pavlenko, A., Driagina, V.: Russian emotion vocabulary in American learners' narratives. *Mod. Lang. J.* **91**, 213–234 (2007)
8. Pavlenko, A.: Emotion and emotion-laden words in the bilingual lexicon. *Biling. Lang. Cogn.* **11**, 147–164 (2008)
9. MacLaury, R.: From brightness to hue: an explanatory model of color-category evolution. *Curr. Anthropol.* **33**, 137–163 (1992)
10. Schirillo, J.: Tutorial on the importance of color in language and culture. *Color. Res. Appl.* **26**, 179–192 (2001)
11. Van Brakel, J.: Comment on “from brightness to hue”. *Curr. Anthropol.* **33**, 169–172 (1992)
12. Burgess, D., Kempton, W., MacLaury, R.: Tarahumara color modifiers: category structure presaging evolutionary change. *Am. Ethnol.* **10**, 133–149 (1983)
13. Lin, H., Luo, M., MacDonald, L., Tarrant, A.: A cross-cultural colour-naming study. Part I: using an unconstrained method. *Color. Res. Appl.* **26**, 40–60 (2001)

Munsell, Albert Henry

Rolf G. Kuehni
Charlotte, NC, USA



Photograph from *A Color Notation* [5]

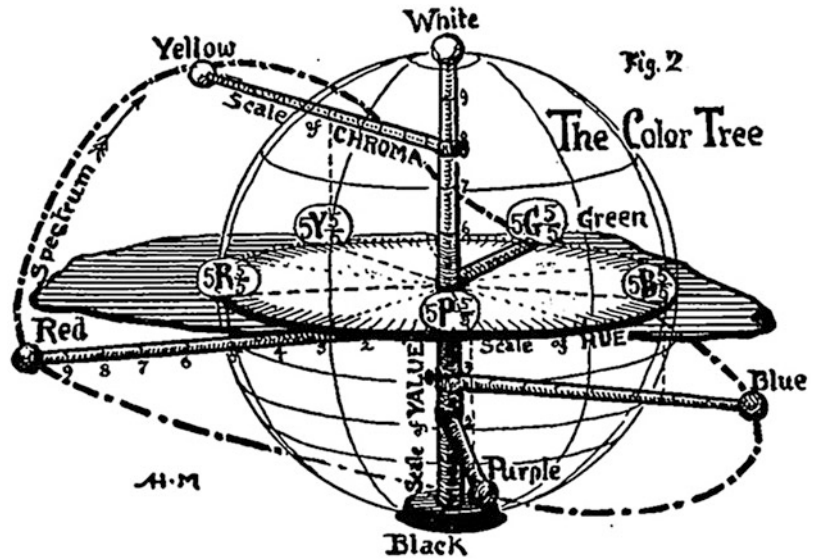
Biography

Munsell was born on January 6, 1858, in Boston, MA, where his father was in the piano business. After high school he attended the Massachusetts Normal Art School in Boston. In 1879 he studied Ogden Nicholas Rood's influential book *Modern Chromatics*. In 1881 he was named an instructor and later a lecturer at this school, positions he held for 25 years. He was awarded a scholarship that made it possible for him to study from 1885 to 1888 at the Académie Julian and the École nationale supérieure des Beaux-Arts in Paris and 1 year in Rome. After he returned he was an active painter of portraits and seascapes. In 1889 he received a patent for an adjustable artist's easel. In 1894 he married Julia Orr, the daughter of a New York financier, with whom he had a son, Albert Ector Orr Munsell, and three daughters. Munsell died on July 28, 1918, in Brookline, MA [1].

Major Accomplishments/Contributions

As a lecturer at the art school, Munsell became concerned with how to teach students about colors in a meaningful manner. In 1899 he developed a model of a balanced color sphere that, when spun rapidly, resulted in a gray appearance, thus showing a kind of color balance. He received a patent for it in 1900 [2]. Munsell met O. N. Rood, then professor of physics in New York, in 1899 showing him his color sphere, with Rood commenting positively on it. Simultaneously, Munsell concerned himself with the design of the interior of the color sphere. In April 1900 he sketched a hue circle based on the decimal system, with five primary, five secondary, and ten intermediate hues. It continues to be the basis of the modern system. Munsell, from his studies in painting, had a clear idea of lightness or value, as it was typically called in painters' circles. In 1901 he obtained a patent for a photometer he called a luminometer. It was manufactured in small numbers in the following years [3]. The basic question of using a logarithmic or a power root scale to relate the physical data of the luminometer to perceived lightness in his system was a question

Munsell, Albert Henry,
Fig. 1 Munsell's
 schematic depiction of the
 color tree



not fully resolved until after Munsell's death. In 1902 Munsell hand-plotted color intensity and lightness data for different pigments, earlier established by W. Abney, and realized that a sphere was an incomplete representation of perceived colors. As a result he began to name his color solid a "color tree." Having established a hue circle and a subjective lightness scale, he was aware that for completeness he needed to define a chromatic intensity scale. It is not established how he arrived at the term "chroma" for that purpose. In March 1902 he sketched a color solid based on the three attributes as shown in more detail in Fig. 1 (Ref. 4, 1907). Samples were to be spaced according to perceptual equal differences by parameter. Munsell began to establish such samples to fill limited hue pages. Munsell described the system in *A Color Notation*, first published in 1905 [5]. A first edition of the *Atlas of the Munsell Color System* was published in 1907 containing eight charts of painted samples of ten different hues [6]. In the 1915 edition of the atlas, the number of samples was doubled. Munsell also introduced artist's tools based on his system: Munsell crayons and Munsell water colors. As a painter Munsell was interested in color harmony, and he established nine principles based on his color solid [7].

Throughout the development of the system, he consulted with a broad group of scientists and artists and gave many presentations in the USA and in Europe. In 1917 he formed the Munsell Color Company to operate the business of producing the atlas. After his passing it was taken over by his son A. E. O. Munsell and other stockholders. At about the same time, the National Bureau of Standards began to show interest in the system and supported sample measurement and expansion of the system. An enlarged edition with 20 hues was published in 1929 [8]. In the 1940s extensive experiments were made under the auspices of the Optical Society of America resulting in the Munsell Renotations (for Munsell system colorimetric renotation data, see, e.g., Ref. [9]), colorimetric definitions of a revised version of samples of the atlas that continue to be the basis of the modern system. Munsell's is perhaps the most important color atlas system yet developed.

References

1. Munsell, A. H.: The National Cyclopaedia of American Biography 36, p. 316 (1950)
2. Munsell, A. H.: Color sphere and mount. US Patent No. 640,792 of 9 Jan 1900

3. Munsell, A.: Notebooks 1900–1918. Accessible at www.cis.rit.edu/research/mcsl2/online/munselljdiaries.php. Accessed 15 Jan 2014
4. Munsell, A.H.: Color and an Eye to Discern It. Self-Published, Boston (1907)
5. Munsell, A.H.: A Color Notation. Ellis, Boston (1905)
6. Munsell, A.H.: Atlas of the Munsell Color System. Ellis, Boston (1907)
7. Munsell, A.H., Cleland, T.M.: A Grammar of Color. Strathmore Paper, Mittineage (1921)
8. Berns, R.S., Billmeyer, F.W.: Development of the 1929 Munsell book of color. Color Res. Appl. **10**, 245–250 (1985)
9. Wyszecki, G., Stiles, W.S.: Color Science, 2d edn. Wiley, Hoboken (1982)

Further Reading

- Kuehni, R.G.: The early development of the Munsell system. Color Res. Appl. **27**, 20–27 (2002)
- Nickerson, D.: History of the Munsell color system and its scientific application. J. Opt. Soc. Am. **30**, 575–585 (1940)

Natural Color Distribution

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

Natural Colorants

► [Colorant, Natural](#)

Natural Dyes

► [Colorant, Natural](#)

Natural Scene Statistics

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

Neo-impressionism

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

Definition

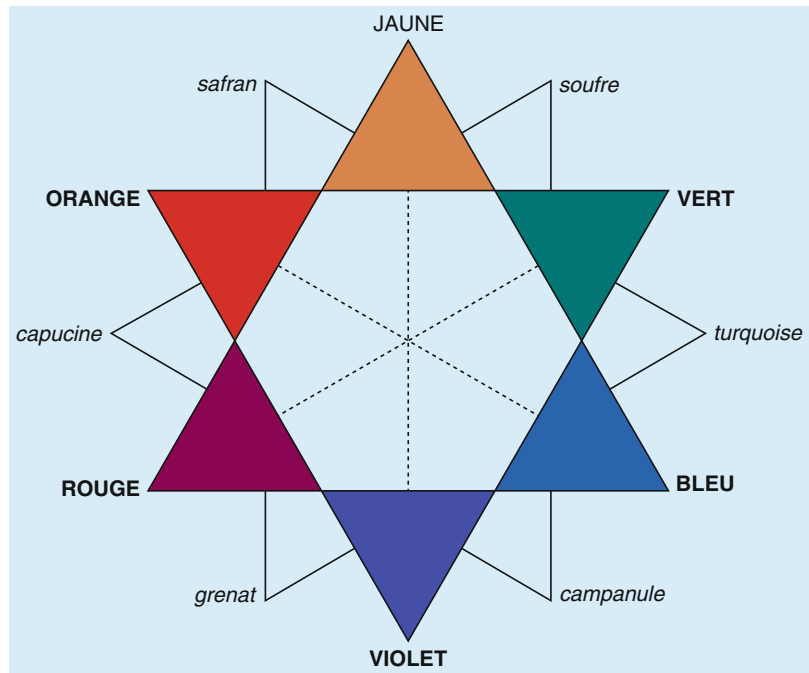
Neo-impressionism is an artistic movement coming after ► [Impressionism](#). It started as a section

of the eighth and last Impressionist Exhibition in 1886. Its founding figure was Georges Seurat (1859–1891), whose monumental painting, *A Sunday on the Grande Jatte*, reworked for its presentation in the 1886 exhibition, is a kind of standard bearer of the whole movement. Seurat died in 1891 at his early 30s. His death corresponds more or less to the end of Neo-impressionism first life, all the more since Camille Pissarro gave up the new technique around the same period. After Seurat's death, Paul Signac (1863–1935) became the leader of the movement and his chief publicist. At the beginning of the twentieth century, Neo-impressionism had a revival during the Fauve period; most of the pioneers of abstract art also had a Neo-impressionist phase in their way to abstraction. The main features of Neo-impressionism are a faith in science and color science, the use of bright colors and of a special technique (optical mixture) aimed at giving more luminosity to colors; this technique, which implies a mechanical application of the brushstroke, was also intended to suppress the skill of the hand as characteristics of the individualism dear to the Impressionists.

Overview

The term “Neo-impressionism” was coined by the critic Félix Fénéon, who was the earliest and one of the best spokesmen of the movement. The

Neo-impressionism,
Fig. 1 Color star from
 Charles Blanc, *Grammaire*
des arts du dessin, 1867



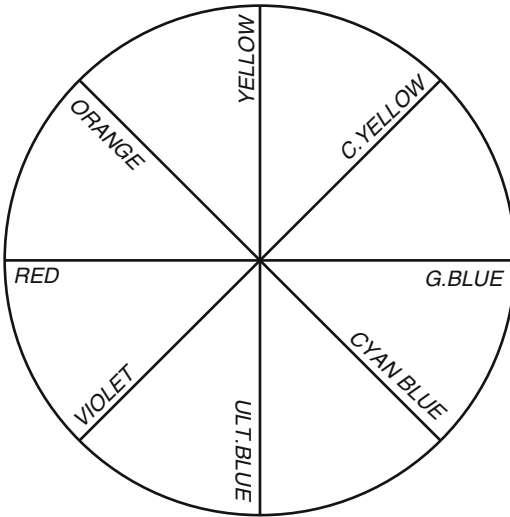
Neo-impressionist artists did not want to break with the ► [Impressionists](#) (one of them, Camille Pissarro, was himself one of the leading Impressionist painters) but continue it in a more rigorous and scientific way. However, one of the differences between the two movements is that, unlike the Impressionists, the Neo-impressionists wanted to render permanent features and not ever-changing moments. A superficial characteristic of their brushstroke is the use of small dots of color, hence the term of pointillism in order to describe their technique.

The main firsthand sources of information to understand Neo-impressionism are (a) an unpublished letter of Seurat to Maurice Beaubourg (1890) of which different drafts are known ([1], pp. 113–114), (b) a letter of Pissarro to his dealer Durand-Ruel (1886) ([1], pp. 54–55), (c) the book written by Signac in order to champion the movement (1899) [2], (d) an article written by the critic Félix Fénéon, and based on information provided by the artists (1886) ([1], pp. 108–110); this article is sometimes considered as the Neo-impressionist manifesto.

Art and Science

According to these writings, the main sources for Neo-impressionist color practice are ► [Michel-Eugène Chevreul](#) [3], basically known through Charles Blanc [4], and Ogden Rood [5]. Now these sources are contradictory, at least concerning ► [color harmony](#), as Chevreul used pairs of ► [complementary colors](#) popularized through Blanc's color star (Fig. 1), while Rood used new pairs of complementary colors based on Helmholtz's researches (Fig. 2). This explains Seurat's hesitation between the two color systems.

The Neo-impressionists used scientific sources and were very interested in what science could bring them; for this reason, they visited Chevreul on one occasion and his assistant on another, in order to understand better the “division of light” (on the importance of Chevreul for them, see [6], pp. 350–377). Their interest for color science raises the issue of whether or not they had a scientific approach to color. This is a very debated topic, as answers have been completely opposed: some specialists think Seurat and his fellow painters used a “scientific” method based on color



Neo-impressionism, Fig. 2 Rood's color wheel, 1879

science; on the contrary, others consider that the Neo-impressionist artists misunderstood the scientists' advices, so that their practice was pseudo-science. Why are there so different positions? It depends on the (implicit or explicit) conception of the relationship between art and science that frames the specialists' thinking. For those who want to bridge the gap between art and science, Seurat's practice is obviously applied science. Conversely, those who hold that science remains out of the scope of artists will emphasize the gap between scientific data and artistic practice and will insist accordingly on the artists' misunderstanding of science. In order to avoid these difficulties, the distinction between color science and color theory might prove useful ([6], pp. 12–17; [7], pp. 44–46). Their function, indeed, is different: color science is a set of experiments and principles aiming at understanding chromatic phenomena and expressing their properties, while color theory is a set of principles, doctrines, rules, and formulas directed at artists and suggesting recipes for choosing, preparing, organizing, and matching colors on the canvas. From this point of view, the difference between an artistic and a scientific approach to color can be more easily understood. For example, Chevreul had noticed that two areas of ► **complementary colors** tend to enhance each other when juxtaposed ([3],

§ 237), but he never recommended this effect to be deployed in painting, despite what is usually assumed. In fact, for him, the enhancing effect was not especially desirable. However, for painters preoccupied with luminosity and saturation, enhancing colors by the juxtaposition of complementary hues soon became an excellent recipe. Examples like this show that “misunderstanding” science is not a mistake that can be avoided by the better-informed artist, but is instead the inescapable result of a difference of aims between the painter and the color scientist.

Optical Mixture

Different origins of the so-called pointillism can be traced from the sources given by the artists. The first is Charles Blanc, who coined the expression “optical mixture,” a technique already used by Delacroix, according to him, as well as by the Impressionists. The basic idea is to mix colors not on the ► **palette**, but juxtapose them on the canvas, so that when seen at a distance, they fuse into the eye and produce a new color not present at such on the canvas. As Blanc put it:

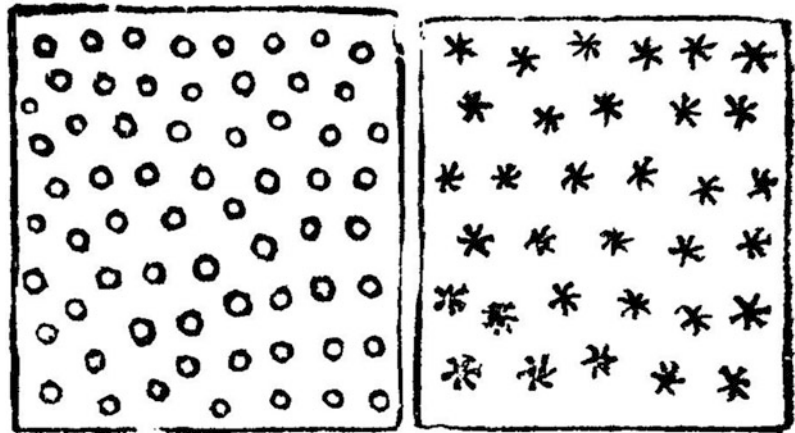
If at a distance of some steps, we look at a cashmere shawl, we generally perceive tones that are not in the fabric, but which compose themselves at the back of our eye by the effect of reciprocal reactions of one tone upon another. Two colors in juxtaposition or superposed in such or such proportions, that is to say according to the extent each shall occupy, will form a third color that our eye will perceive at a distance, without having been written by weaver or painter. This third color is a resultant that the artist foresaw and which is born of optical mixture ([4], p. 475).

In order to make his suggestions more concrete, he added a drawing of little dots (Fig. 3, left) or little stars (Fig. 3, right) on a uniform ground: “A similar phenomenon will be produced upon a yellow stuff starred with violet and upon a blue stuff sown with orange spots” ([4], p. 475).

A physicist, Ogden Rood, gave a more complete description of optical mixture in his book published in 1879 and translated into French in 1881. He first described a method given by Mile in 1838 in order to mix colors in the eye from

Neo-impressionism,

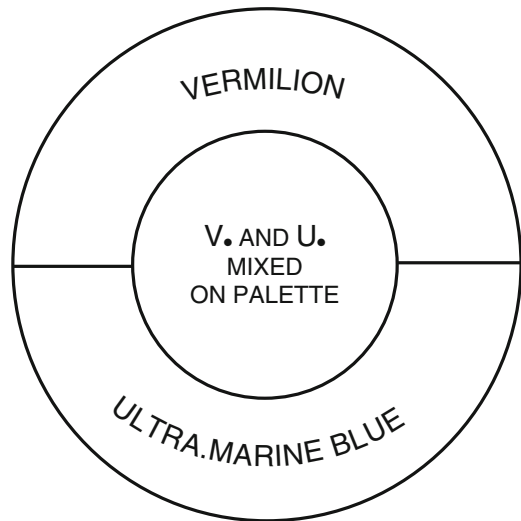
Fig. 3 Drawing from Charles Blanc's *Grammaire des arts du dessin*, 1867



small contiguous dots of colors ([5], p. 140). Rood also quoted John Ruskin, who recommended a similar method ([5], p. 140). It must be noted Monet as well as Signac praised a lot Ruskin's book *The Elements of Drawing*.

Rood himself gave painters the same advice:

There is, however, another lower degree of gradation which has a peculiar charm of its own, and is very precious in art and nature. The effect referred to take place when different colours are placed side by side in lines or dots, and when viewed at such a distance that the blending is more or less accomplished by the eye of the beholder. Under these circumstances the tints mix on the retina, and produce new colours, which are identical with those which are obtained by the method of revolving disks ([5], pp. 279–280).



Neo-impressionism, Fig. 4 Drawing of experiment from Rood's *Modern Chromatics*, 1879

Luminosity

The end of the quote explains why the Neo-impressionists were so interested by the method suggested by Rood (after Blanc and Ruskin): Rood made a fascinating experiment in order to compare the luminosity of colors mixed on the ► [palette](#) with the optical mixture of the same hues. For this purpose, he used two rotating disks. The larger disk was divided into two equal halves, painted with vermilion in the upper part and ultramarine blue in the lower part (Fig. 4).

Now, an equal number of drops of the same washes of watercolor pigments were mixed on the ► [palette](#) in order to paint the small disk with this

mixture. After rotating the two disks, Rood realized that the smaller disk looked gray and dull, while the larger one looked brighter. So, in order to obtain a similar result on both disks, he had to add black to the larger one. But the surprising result of the experiment is that Rood had to add up to 52 % of black to the larger disk in order to match the mixture made on the palette (Fig. 5).

Many artists were struck by this experiment, and Fénéon gave it as an example in order to explain the new technique and its superiority in terms of luminosity. Quoting the results of Rood's experiment, Fénéon stressed that “the luminosity

Neo-impressionism,

Fig. 5 Compared result of color mixtures, Rood's *Modern Chromatics*, 1879

Mixture on Palette.	Mixture by Rotation.
50 violet + 50 Hooker's green..	= { 21 violet + 22.5 Hooker's green + 4 vermilion + 52.5 black.
50 violet + 50 gamboge.....	= 54 violet + 20 gamboge + 26 black.
50 violet + 50 green.....	= 50 violet + 18 green + 32 black.
50 violet + 50 Prussian-blue...	= 47 violet + 49 Prussian-blue + 4 black.

of optical mixtures is always superior to that of material mixture" ([1], p. 109) and justified accordingly the new Neo-impressionist technique.

However, the theory of "optical mixture," as understood by painters, when suggesting that it should be possible to get an additive-like mixture by juxtaposing pigments so that they mix in the eye instead of physically, has been a source of enduring confusion. The point that often eluded painters is that this type of optical mixture does not add together the luminous energies of the individual colors; it merely averages them. Furthermore, there is a contradiction in the theory, when painters wished to combine the two principles: ► [complementary colors](#) and optical mixture. In principle, both have in common to heighten colors: they heighten hues in the case of juxtaposing complementary colors; they heighten luminosity in the case of optical mixture. However, painters were not aware of the fact that both principles correspond to two different laws of perception. The ► [color contrast](#) only functions if the samples of color are big enough. For when the juxtaposed zones are thin, what happens is exactly the opposite: instead of enhancing each other, that is, to exaggerate their difference, they tend to "assimilate," that is, to produce visually a dirty gray.

This leads to a puzzling conclusion: if juxtaposing dots of ► [complementary colors](#), as the Neo-impressionist painters did, does not give the paintings more luminosity, how could the great luminosity of their paintings be explained? The answer has been given by Robert Herbert, the scholar most familiar with Neo-impressionism. As he explained, if the Neo-impressionist paintings are indeed very luminous, it is because the optical mixture does not work, that is, that the dots are big enough to be still perceived at the normal distance at which we look at the paintings! And it is precisely because the dots do not mix in order to produce new

hues that they keep their luminosity ([8], p. 19). It is not the case to stress here that the painters were "wrong" in their use of theory, but only to understand that they were so fascinated by the possibility of obtaining a luminosity similar to that of color-light that they forgot scientists' warnings against the danger of confusing pigments and light colors.

This means that the numerous misunderstandings arising from the notion of optical mixture are inevitable since they are due to the very ambiguity of the concept, which suggests a strong analogy between light and the eye, as opposed to pigment mixture. However, painters work with pigments and not with light but were eager to obtain through their pigments a luminosity similar to that of additive mixture. Hence, the opposition in Signac's writings between mixtures on the ► [palette](#) he considered disgusting and optical mixture "purity."

The Neo-impressionist technique is then rather complex and cannot be limited to "pointillism." The term "pointillism" is often considered as a synonym of Neo-impressionism but is not. Signac insisted a lot on the fact that the dots are just a means and cannot be considered as a characteristic of Neo-impressionism: far more important for its technique is the "divided touch." As the painter wrote: "*Division* is a complex system of harmony, an aesthetics rather than a technique. The *dot* is only a means. To divide is to seek the power and harmony of color, through representing colored light by its pure elements and by using the optical mixture of these pure elements, separated and proportioned according to the essential laws of contrast and gradation" ([1], p. 121).

Harmony

As results from this quote, the idea of harmony in Neo-impressionism goes beyond color, due to the

Neo-impressionists' adherence to symbolism. In his letter to Beaubourg in which he explained his main aesthetic ideas, Seurat also related harmony with tones and lines ([1], pp. 113–14). The idea of color harmony must also be associated with the concept of harmony in anarchism, whose ideals were shared by several of the painters. For the anarchists, indeed, the idea of harmony makes it possible to reconcile the individualism characteristic of the movement with collective life. Similarly, a parallel can be drawn between the individuality of colors and the painting as a whole, as all the single elements used (dots) interact with each other and contribute to the general harmony of the painting [9].

Neo-impressionism and Color Development in Twentieth-Century Art

It is striking to note that most of the painters that counted for the development of color in the first half of the twentieth century passed through a Neo-impressionist period. The list is quite impressive: Matisse, Mondrian, Kandinsky, Malevich, Braque, Delaunay, Kupka, and so on. This raises the question of why Neo-impressionism was so important for the development of the practice of color in painting.

The first feature to be noted is the extraordinary homogeneity of the pictorial matter: the dots that serve to render nature are the same as used for a human body. And even in the human body, no difference is made between hair, skin, clothes, and so on. Everything is treated in the same way; besides color, this is also true for lines and spaces. This homogeneity is important for several reasons. First, it means the rejection of the skill of the hand as an expressive device. For the Neo-impressionists, what matters are the scientific bases of the technique. Furthermore, through the homogeneous construction of the surface, the touch becomes independent of the objects represented on a painting, since the same kind of dot is used on the whole surface of the canvas. From this point of view, the Neo-impressionists have emphasized special means, which are artificial devices that do not resemble nature.

Yet the fact of using artificial means has another consequence: the touch (as a means) becomes independent of the representation of nature, which was the traditional aim of painting. Finally, not only is the divided touch a technique in order to paint through color only, but it also contains a syntax, that is, a way of organizing the contiguous color dots, as well as a complex harmony system. The vocabulary (the dots), the syntax (the divided touch), and the harmony constitute a totality in itself, independently of the world of the objects the painting is supposed to represent. This is the main reason why the next generations of painters were so interested in Neo-impressionism, read avidly Signac's book, and had a Neo-impressionist phase: the dots were a way of exploring the possibility of painting through colors only, which eventually led to abstract art [10].

Summing up, it must be emphasized that a lot has been written on Neo-impressionism. On color in particular, the interested reader can also refer to G. Roque [7], J. Gage [11], and P. Smith [12], who have developed different aspects related to Seurat, and G. Roque [13], who refers to Signac and color theory.

Cross-References

- [Anchoring Theory of Lightness](#)
- [Chevreul, Michel-Eugène](#)
- [Color Contrast](#)
- [Color Harmony](#)
- [Complementary Colors](#)
- [Impressionism](#)
- [Palette](#)
- [Retinal Art](#)
- [Unique Hues](#)

References

1. Nochlin, L.: *Impressionism and Post-impressionism 1874–1904. Sources and Documents*. Prentice Hall, Englewood Cliffs (1966)
2. Signac, P.: *D'Eugène Delacroix au néo-impressionnisme*. Hermann, Paris (1st ed. 1899) (1964); excerpts in English are published in [1], pp. 116–123 (A full English translation has been published as an appendix (pp. 193–286) of Ratliff, F.: *Paul Signac and Color in Neo-impressionism*. The Rockefeller University Press, New York (1992))

3. Chevreul, M.-E.: *De la loi du contraste simultané des couleurs...* Pitois-Levrault, Paris (1839) (The latest English translation is: *The Principles of Harmony and Contrast of Colors and their Applications to the Arts* (1855). Kessinger Publishing LLC, Whitefish (2009))
4. Blanc, Ch.: *Grammaire des arts du dessin*. Renouard, Paris (1867) (The chapter on color has been translated into English in Taylor, J.C. (ed.): *Nineteenth-Century Theories of Art*, pp. 468–479. University of California Press, Berkeley (1989); the quotations in this article refer to this translation)
5. Rood, O.N.: *Student's Textbook of Color or Modern Chromatics, with Applications to Art and Industry*. D. Appleton and Company, New York (1st ed. 1879) (1899)
6. Roque, G.: *Art et science de la couleur: Chevreul et les peintres, de Delacroix à l'abstraction*, 2nd edn. Gallimard, Paris (2009)
7. Roque, G.: *Seurat and color theory*. In: Smith, P. (ed.) *Seurat Re-Viewed*, pp. 43–64. Pennsylvania State University Press, University Park (2009)
8. Herbert, R.L.: *Neo-impressionism*. Exhib. cat. Guggenheim Museum, New York (1968)
9. Roque, G.: *Harmonie des couleurs, harmonie sociale*. 48/14. *La revue du musée d'Orsay* **12**, 62–73 (2001)
10. Franz, E. (ed.): *Signac et la libération de la couleur de Matisse à Mondrian*. Exhib. cat. Musée de Grenoble, Grenoble (1977)
11. Gage, J.: *The technique of Seurat – a reappraisal*. In: Gage, J. (ed.) *Color and Meaning. Art, Science and Symbolism*, pp. 209–218. Thames and Hudson, London (1999)
12. Smith, P.: *Seurat and the Avant-Garde*. Yale University Press, New Haven/London (1997)
13. Roque, G.: *Signac et la théorie de la couleur*. In: Ferretti Bocquillon, M. (ed.) *Signac. Les couleurs de l'eau*. Exhib. cat, pp. 30–39. Musée des Impressionnismes and Gallimard, Giverny/Paris (2013)

Neon Color Spreading

► Motion and Color Cognition

Neon Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

Glow discharge lamps; Low-pressure glow discharge lamps

Definition

Lamps that produce light as a result of an electrical glow discharge, generated between two electrodes, in a low-pressure neon, argon, krypton, helium, xenon, or carbon dioxide vapor, which is contained in a transparent bulb or tube.

Note that also glow discharge lamps using no neon but one of the other vapors mentioned in the definition are referred to as neon lamps.

Neon lamps are usually not used for general lighting purposes, and therefore only the fundamentals of their working principle and of their properties are described in this encyclopedia.

Working Principle of Glow Discharge Lamps

In a glow discharge, the electrons taking part in the initial discharge are made free by external sources like thermal collisions between atoms, ultraviolet, or cosmic radiation. In a discharge bulb or tube with a gas at very low pressure and a voltage of several hundred volts applied across cold electrodes, the electrodes made free by the external sources move at high speed toward the anode and ionize some gas atoms creating positive charged ions and more free electrons. Some of the positive ions hit the cathode at high speed, thanks to the low pressure of the gas of less than 0.05 atmosphere. They release by their impact free electrons from the cathode surface. Electron production is thus not by external means like heated cathodes but by secondary emission as a result of positive ions hitting the cathode surface. Atoms can also be excited to a higher-energy state by the moving electrons and ions, and in releasing the extra energy so obtained, photons are created that are responsible for the typical soft and transparent luminous glow. The color characteristics of the glow light are dependent on the type of gas used.

The glow discharge process is only stable at small current densities in the order of milliamps. By increasing the voltage, the current is increased, and the process will change into that of an arc discharge that is typical for most discharge lamps used for lighting: low- and high-pressure sodium,

mercury, and metal halide lamps. Some arc discharge lamps make use of the glow light process to facilitate ignition of the lamp (the red neon glow light of a starting low-pressure sodium lamp). Glow light starters (used in some fluorescent lamp systems) also make use of this process.

Tubular Glow Discharge Lamps

Tubular neon lamps can be shaped into any form and produced in many different color versions. They are therefore particularly well suited for advertising signs. LED strings now offer an alternative. Tubular neon signs are also popular in decorative art.

Construction

Lead glass is usually used for the tube because it can easily be bent and fused. Sometimes red- or yellow-tinted glass is used to adapt the original color of the glow light. So will yellow-tinted glass change the color of blue glow light into green.

The length of the tube usually is less than 2 m, and its diameter varies between some 7 and 35 mm. Larger signs are constructed from various pieces of individual tubes.

The electrodes are large rings of pure iron with a diameter of 15–20 mm and a length of about 80 mm.

Pure neon as fill gas results in red light; most other colors can be produced with argon-mercury mixtures. Helium results in warm white light.

Fluorescent powders on the inner wall are often used to extend the range of colors or to improve light output.

Properties

System Luminous Efficacy

Luminous efficacy of clear glass tubes is low with some 5 lm/W. Fluorescent-covered tubes have efficacies depending on the color between some 10 lm/W for red, 30 lm/W for blue, more than 80 lm/W for green, and 75 lm/W for white. The

values vary quite a bit with the type of fluorescent powder used.

Lumen Package Range

The luminous flux of neon tubes is expressed in lumen per meter of tube length. Values vary according to type of gas, type of fluorescent powder (and thus color), lamp current, and tube diameter between 100 lm/m (clear glass neon filled) and 1600 lm/m (fluorescent-coated tube producing green light).

Color Characteristics

Applying different fill gasses, different tinted glass tubes, and different fluorescent powders enables the production of all the colors of the spectrum.

Lamp Life

Neon tubes have electrically a very long life. Their practical life is often indicated as the time after which the light output has dropped to 50 %: 10,000–20,000 h.

Run-Up and Reignition

Ignition of the cold lamp is instantaneous and hot reignition is immediate as well. This makes it possible to produce flashing neon sign installations.

Compact Glow Discharge Lamps

Neon compact glow lamps were used as indicator lamps on all kinds of electrical and electronic equipment. LED signal lamps have taken over this application. They were also, and still are, used for low-level lighting in night lamps in children's and hospital bedrooms and corridors. Also, here LEDs are an alternative. The lamp's electrodes can be shaped in many different forms so that the glow light can be used to display letters, numbers, and symbols. By making the lamp flashing a flickering neon glow, for example, in the form of a candle, flame can be produced.

Construction

Compact glow discharge lamps consist of a compact glass bulb with two electrodes closely positioned together. The nickel or iron electrodes can take any form like strip, disk, ring, cross, etc. The filling gas usually is neon giving an orange-red glow. Fluorescent coating is used for green-emitting lamps. Miniature indicator lamps in sizes from some 5 to 15 mm operate on a current of 0.5–1 mA. The operating current of the larger types in the size of an incandescent bulb goes up to 50 mA.

Properties

System Luminous Efficacy

Luminous efficacy is 0.3–1 lm/W which for indicator lamps is reasonable enough.

Lumen Package

The lumen output, less to much less than one lumen, is low but for indicator lamp applications sufficient.

Color Characteristics

Clear glass bulbs produce orange-red light. With fluorescent coating of the bulb, many more colors can be produced, especially also green which for indicator lamps is indispensable.

Lamp Life

Lamp life is 10,000–100,000 h.

Run-Up and Reignition

Cold start and reignition are both instantaneous which of course for indicator lamps is a first requirement.

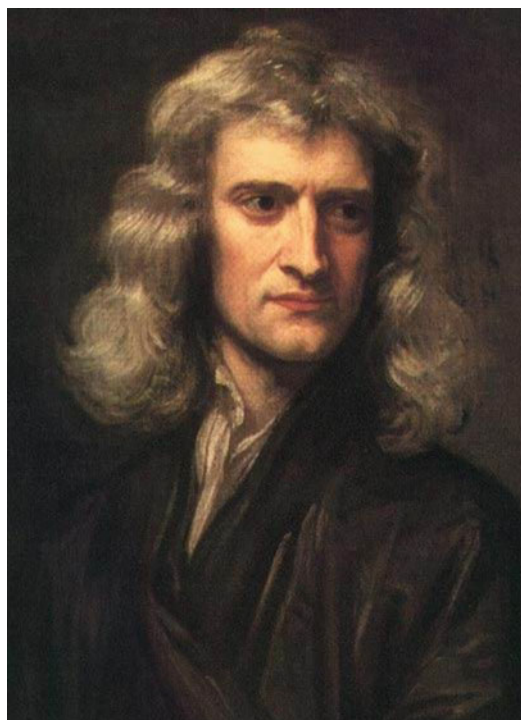
Cross-References

- ▶ [Light-Emitting Diode, LED](#)
- ▶ [Phosphors and Fluorescent Powders](#)

Newton, (Sir) Isaac

Renzo Shamey

Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA



1642/1643–1726/1727 (The difference is due to the use and report of the date in two calendars: Julian and Gregorian. According to the Gregorian calendar, the date of Newton's birth was 4 Jan, 1643 and that of his death 31 March 1727) (Portrait of Isaac Newton in 1689 (age 46) by Godfrey Kneller, Wikimedia)

Biography

Isaac Newton was an English physicist and mathematician, who made seminal contributions to several domains of science, and was considered

a leading scientist of his era and one of the most influential scientists of all time. He was born prematurely on 25 December 1642 in Lincolnshire, England, of Hannah Ayscough and Isaac Newton (father) [1]. His father passed away before he was born. His mother remarried when he was 3 years old and he was left in the care of his grandmother, a situation he resented.

He attended the King's School in Grantham where he learned Latin among other things until the age of 17. In 1661, he was admitted to Trinity College, Cambridge, and was educated in Aristotelian philosophy. However, Newton also read the works of Descartes, Galileo, and Kepler. In 1665/1666 he spent most of the time at his ancestral place in Lincolnshire because of the dangerous spread of plague in Cambridge. During that time he did most of his experimental work with glass prisms and much of his mental work that resulted in the publication of *Principia* in 1687.

In 1667 he returned to Cambridge and became a fellow of the College of the Holy and Undivided Trinity [2, 3]. In 1669, and at the age of 26, Newton became the Lucasian professor of mathematics. According to his secretary Humphrey Newton (no relation), his lectures were often poorly attended and few understood him and that sometimes he read to the walls [4]. Newton occasionally traveled to London to attend the Royal Society lectures and was named a fellow in 1672. He became its president from 1703 to 1727. Newton was given various levels of support by the Royal Society. His interpretation of the optical experiments was strongly disputed by Robert Hooke, an employee of the society since 1664, as a result of which he published his book *Opticks* only in 1704, after Hooke's death. He was supported by the society in his bitter and controversial dispute with the German polymath and philosopher Gottfried Wilhelm Leibniz over who had developed calculus first [Leibniz's notations are used today].

Newton also dwelt in politics and was a member of the House of Commons between 1689 and 1690 and then again from 1701 to 1702. In 1705 he was knighted by Queen Anne during her visit to Trinity College. He held two government offices: first he was the Warden of the Mint from

1696 to 1700 and then Master of the Mint from 1700 until his death in 1727. At the time of Newton's funeral, the French philosopher Voltaire who was in England compared Newton to Descartes and said of Newton that "he was never sensible to any passion, [and] was not subject to the common frailties of mankind, nor had any commerce with women" [5]. Newton had strong opinions on religion and wrote a number of works, not published during his lifetime, that would have then been considered heretic in that period. Newton never married and died intestate in Kensington, London, when his relatives quarreled over the division of his considerable estate. He is buried in Westminster Abbey in London, England.

Major Accomplishments/Contributions

Newton's Theory of Color

Arguably our modern understanding of light and color begins with Newton's discovery of light dispersion which he published in 1672. In the late 1660s, Newton started experimenting with the phenomenon of colors and lectured on optics [6]. At the time, it was generally thought that colors were mixtures of light and darkness. It was also believed that prisms imparted colors to light. Through observation of light refraction, Newton realized that this theory was incorrect. He demonstrated that a prism decomposes "white" light into a spectrum of colors. Newton obtained a triangular prism and began "to try therewith the celebrated Phaenomena of Colours." In his notes he states, "having darkened my chamber, and made a small hole in my window-shuts, to let in a convenient quantity of the Sun's light, I placed my Prisme at this entrance, that it might be thereby refracted to the opposite wall. It was at first a very pleasing divertisement, to view the vivid and intense colours produced thereby; but after a while applying myself to consider them more circumspectly, I became surprised to see them in an oblong form, which, according to the received laws of Refraction, I expected should have been circular" [7]. The original sketch demonstrates the dark room environment where this experiment was

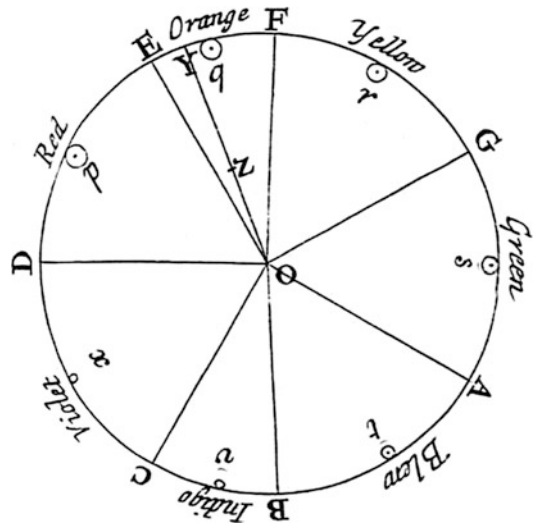
refracted by accelerating into a denser medium. To transmit forces between particles, Newton posited the existence of the ether. However, he replaced ether with occult forces based on Hermetic ideas of attraction and repulsion between particles and Newton's considerable writings on alchemy. It has been said that "Newton was not the first of the age of reason: He was the last of the magicians." Indeed, Newton's interest in alchemy cannot be isolated from his contributions to science [11] since during his time there was no clear distinction between alchemy and science.

Later physicists favored a purely wavelike explanation of light to account for the interference patterns and the general phenomenon of diffraction. Later on, Young and Fresnel combined Newton's particle theory with the wave theory and indicated that color is the visible manifestation of light's wavelength.

The Color Circle

In his observation of white light dispersion into spectral components, Newton divided the spectrum into seven named colors: violet, indigo, blue, green, yellow, orange, and red. He placed the colors in a circular fashion and called it a color circle and used it to demonstrate the results of mixtures of spectral lights, including complementarity (those that when mixed cancel each other out and result in white, gray, or black). The size of each segment differed from the other according to his calculations of its wavelength, the seven musical tones or intervals of the eight sounds and of its corresponding width in the spectrum. The choice of seven colors was out of a belief, derived from the ancient Greek, that colors, objects in the solar system, the musical notes, and the days of the week were connected. This likely explains the selection of indigo as another hue between blue and violet. Some argue, however, that in Newton's prismatic colors, "indigo" would be placed as a color that is today called blue, whereas "blue" would correspond to cyan (Fig. 2).

Newton proposed a method to determine the "fullness or intenseness" of combined colors on the circle based on the distance of the center of the combined gravity of the circles for each of the rays of light from whiteness (the distance from O to



Newton, (Sir) Isaac, Fig. 2 Colors and the associated musical notes in Newton's color wheel, shown in his book *Opticks* of 1704. The circle completes a full musical octave, from D to D [8]. Colors on opposite sides are complementary

Z in the figure shown for Color Y, which arises from the composition of all the colors in the given mixture). The color circle was, if lacking the nonspectral colors, an early representation of what became centuries later, in a modified form, the chromaticity diagram. In his system, Newton had connected violet to red in the circle, and thus, a large gamut of purples was not shown. Newton described the complementary colors and their mixture and stated "If only two of the primary colours which in the circle are opposite to one another be mixed in an equal proportion, . . . , the colour compounded of these two shall not be perfectly white, but some faint anonymous colour." [8]. Newton's concept of complementary colors was demonstrated more thoroughly in the nineteenth century by color theorists. Helmholtz established the complementary stimulus pairs, and Ogden Rood (1831–1902) emphasized that to reveal applied colors in their natural brilliance, a knowledge of the complementary hues was required [13].

A version of Newton's color circle without the indigo blue was adopted by painters to describe complementary colors. Nonetheless, this circular diagram became the model for many color systems of the eighteenth and nineteenth centuries. The

Newton, (Sir) Isaac,

Fig. 3 Newton was the Warden and then the Master of the Royal Mint in England, and his picture appeared on the one pound note, almost three centuries later [15] (With permission to reprint from the Bank of England)

**Newton, (Sir) Isaac,**

Fig. 4 An image of a German stamp commemorating Newton's 350th birthday [16]



conceptual arrangement of colors in this form also allowed the painters' primaries (red, yellow, blue) to be arranged opposite their complementary colors (e.g., red opposite green), as a way of denoting that each complementary color would enhance the other's effect through optical contrast.

In addition to his work on optics, Newton made seminal contributions to several other scientific disciplines. In his book the *Principia* or "*Mathematical Principles of Natural Philosophy*," which was published in 1687, Newton formulated the laws of motion and universal gravitation and is considered to have laid the foundations for classical mechanics [14]. He also introduced the notion of a Newtonian fluid, studied the speed of sound, and developed an empirical law of cooling among other major contributions made to scientific discovery.

Newton's image appeared on the Bank of England notes for about 10 years in the 1970s and 1980s [15]. He has also been commemorated on various stamps and coins [16] (Figs. 3 and 4).

References

1. Storr, A.: Isaac Newton. *Br. Med. J.* **291**, 1779 (1985). doi:10.1136/bmj.291.6511.1779
2. Westfall, R.S.: *Never at Rest*. Cambridge University Press (1980, 1998). ISBN 0-521-27435-4
3. Westfall, R.S.: *Isaac Newton*. Cambridge University Press (2007). ISBN 978-0-19-921355-9
4. More, L.T.: *Isaac Newton A Biography 1642-1727*. pp. 246, 381, and 389. Scribners, New York (1934)
5. Voltaire, F.M.A: Cassell & Co. In: Price, D., (ed.) *Letters on England*. p. 100. (1894). Also available in

pdf form http://livros.universia.com.br/?dl_name=Letters-on-England-de-Voltaire.pdf

6. Newton, I.: *Hydrostatics, Optics, Sound and Heat*. Cambridge University Digital Library (c. 1670–c. 1710)
7. Whittaker, E.T.: *A History of the Theories of Aether and Electricity: From the Age of Descartes to the Close of the Nineteenth Century*. Hodges, Figgis, Dublin (1910)
8. Newton, I.: *Opticks, or A Treatise of the Reflections, Refractions, Inflections & Colours of Light*. Dover Publications, New York (1952). Online version available on <https://archive.org/details/opticksoratreat00newtgoog>
9. Ball, W.W.R.: *A Short Account of the History of Mathematics*. Dover, New York (1908). ISBN 0-486-20630-0
10. Darrigol, O.: *A History of Optics from Greek Antiquity to the Nineteenth Century* (2012). ISBN-13: 978-0199644377
11. White, M.: *Isaac Newton: The Last Sorcerer*. Fourth Estate Limited (1997). ISBN 1-85702-416-8
12. Newton, I.: *Of Colours*. The Newton Project. <http://www.newtonproject.sussex.ac.uk/view/texts/normalized/NATP00004>. Retrieved 4 Apr 2015
13. Rood, O.N.: *Modern Chromatics, with Application to Art and Industry*. New York (1879). Also available online: <http://lcweb2.loc.gov/service/gdc/scd0001/2010/20100701001mo//20100701001mo.pdf>
14. Newton, I.: *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press, Berkeley (1999)
15. <http://www.royalmintmuseum.org.uk/history/people/mint-officials/isaac-newton/index.html>. Visited 9 Apr 2015
16. http://th.physik.uni-frankfurt.de/~jr/gif/stamps/stamp_newton.jpg. Visited 9 Apr 2015

Nickerson, Dorothy

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Dorothy Nickerson (August 5, 1900–April 25, 1985) was an American color scientist and technologist who made important contributions in the fields of color quality control, technical use of colorimetry, relationship between color stimuli and color perceptions, standardization of light sources, color tolerance specification, and others. Nickerson was born and raised in Boston and attended Boston University in 1919 and Johns Hopkins University in 1923. Later, she continued her education at summer courses and university extensions at Harvard University, George Washington University, and the Graduate School of the US Department of Agriculture. Her special interest was the science of color then in significant development.

In 1921, Nickerson joined the Munsell Color Company as a laboratory assistant and secretary to A. E. O. Munsell who in 1918 had taken over the firm from his father. In 1922, the firm moved to New York and in 1923 to Baltimore. In 1927, she was offered a position at the US Department of Agriculture where she remained until her retirement in 1964. When she joined, color science and technology were without international standards and at the beginning of industrial use. Nickerson was instrumental in developing the technology and its use in agricultural and industrial settings.

Nickerson became the first individual member of the Inter-Society Color Council, founded in 1931. She was a lifelong member, received the organization's Godlove Award, and had an award named after herself. She was a member of the US National Committee of the CIE and the International Association on Color where she received the first D. B. Judd AIC Award in 1975. Nickerson was a trustee of the Munsell Color Foundation since 1942, was its president from 1973 to 1975, and assisted in the transfer of the foundation to the Rochester Institute of Technology in 1983 where it helped fund the then new Munsell Color Science Laboratory. Nickerson was the author and coauthor of some 150 articles and other publications [1].

Major Accomplishments/Contributions

Color quality control of agricultural products: In the late 1920s, Nickerson worked on usage of disk color mixture to define the color quality of cotton and other agricultural products (see portrait image, ca. 1930, US Government) and the conversion of disk mixture data into the CIE colorimetric system [2].

Standardization of light sources for color assessment and color rendering: In the late 1930s, a major occupation of her was the development of defined light sources for visual assessment of color quality. Later, she was also active in the development and promotion of standard methods for the definition of color rendering of lights [3].

Munsell color system and its colorimetric definition: In 1940, a technical committee of the Optical Society of America (OSA) began a study of the Munsell color system, improvements and extensions of the system, and its definition in the CIE colorimetric system. Nickerson was an important participant in this effort. The final report of the committee was authored by S. M. Newhall, D. Nickerson, and D. B. Judd, and its results are known as the Munsell renotations, the specification of the aim colors of the current system. Nickerson prepared plots of the Munsell color stimuli in the CIE chromaticity diagram that remain in publication today [4, 5]. She also wrote an extended history of the Munsell system [6].

Color tolerance specification: In 1936, Nickerson published the first color difference formula for industrial use, based on the addition of increments of Munsell hue, chroma, and lightness scale values. In 1943, together with Newhall, she published realistic representations of a three-dimensional perceptually approximately uniform optimal object color solid. In 1944, together with her assistant K. F. Stultz, she published a colorimetric color difference formula, known as the Adams–Nickerson–Stultz formula, that in modified form eventually became the CIE $L^*a^*b^*$ color space and difference formula [7, 8].

Color charts: In the mid-1940s, Nickerson was active in methods for assessing the color of soils, an effort that found its expression in the Munsell Soil Color Chart, still in use today. In 1957, Munsell issued the Nickerson Color Fan, a color fan for horticultural purposes [9]. Working with D.B. Judd, the chair of the OSA committee that developed the OSA Uniform Color Space, Nickerson, as a member of the committee, was also a contributor to that effort for over 25 years and wrote a detailed history of the development of the system [10].

References

1. Bartleson, C. J., Luke, J. T.: Dorothy Nickerson, 1900–1985. *Optics News*. **11**, 28–29 (1985)
2. Nickerson, D.: A Method for Determining the Color of Agricultural Products, U.S. Dept. of Agriculture Technical Bulletin, Washington, DC 154, 32 p. (1929)
3. Nickerson, D.: Artificial daylighting for color grading of agricultural products. *J. Opt. Soc. Am.* **29**, 1–9 (1939)
4. Newhall, S.M., Nickerson, D., Judd, D.B.: Final report of the OSA Subcommittee on the spacing of the Munsell colors. *J. Opt. Soc. Am.* **33**, 385–418 (1943)
5. Wyszecki, G., Stiles, W.S.: *Color Science*, 2nd edn, pp. 853–861. Wiley, New York (1985)
6. (a) Nickerson, D. History of the Munsell color system and its scientific application. *J. Opt. Soc. Am.* **30**, 575–585 (1940); (b) Nickerson, D.: History of the Munsell color system. *Color Eng.* **7**, 42–51 (1969)
7. Nickerson, D.: The specification of color tolerances. *Textile Res.* **6**, 505–514 (1936)
8. Nickerson, D., Stultz, K.F.: Color tolerance specification. *J. Opt. Soc. Am.* **34**, 550–570 (1944)
9. Nickerson Color Fan, produced by the Munsell Color Company beginning in 1957, with 262 color samples in 40 hues, no longer produced
10. Nickerson, D.: History of the OSA Committee on Uniform Color Scales. *Optics News*, Winter, 8–17 (1977)

Nil

► Metamerism

Non-Photorealistic Rendering

Jorge Lopez-Moreno

GMRV Group, Universidad Rey Juan Carlos,
Móstoles, Madrid, Spain

Synonyms

Artistic rendering; Computer depiction; Expressive rendering

Definition

In computer graphics, non-photorealistic rendering (NPR) describes all the depiction techniques which do not aim to convey photographic realism. The reason why this term is defined by “what it is not” is that through the last decades, the computer graphics community has focused most of its efforts in creating images indistinguishable from reality (like a photograph or photorealistic), thus assimilating the subfield “photorealistic rendering” to the more general term “computer graphics.”

Some of the most well-known NPR techniques are 3D rendering with *cel shading* to simulate cartoon style, simulation of traditional art media (ink, pencil, watercolor, etc.) and artistic styles (abstract painting, impressionism, etc.), and technical depiction of objects in order to enhance clarity over realism: architecture blueprints, engineering cutaway, and explode views or anatomical dissections.

Origin of the Term

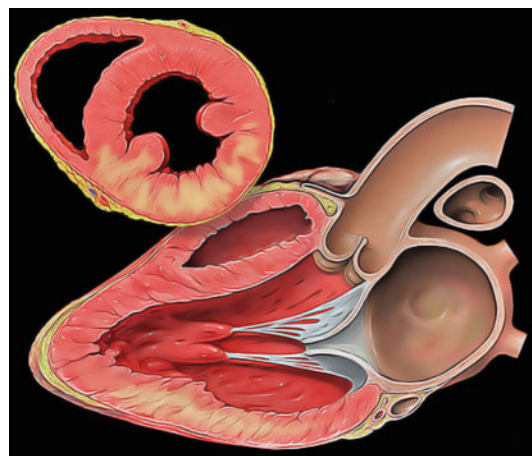
The term was coined by Salesin and Winkenbach in 1994 [1] and although its widespread use by researchers around the world, it is not free from controversy. Since the first International Symposium on Non-Photorealistic Animation and Rendering (the most well-known research conference on this subject) in the year 2000 [2], there is an ongoing discussion over the looseness of this

term, and several alternatives have been proposed such as “computer depiction” [3]. One reason is that *non-photorealistic* only describes a style of generating images, while the range of topics in NPR research encompasses both rendering aspects (multiple styles of picture generation) and interaction issues (e.g., artistic tools such as a simulated brush).

Related Fields: Art and Perception

The motivation for NPR is that realism by itself is not always desirable in order to communicate something about the scene to be depicted. For instance, painters like Van Gogh or Munch used extremely saturated and unrealistic colors and paint strokes in order to convey not only the depicted object but also a certain mood or feeling. Likewise but not for artistic reasons, stylized illustration is preferred over real images in medicine to teach anatomy, as they help in differentiating the parts of the human body which are otherwise difficult to identify in their natural state (see Fig. 1).

Given this range of applications, the NPR field has naturally embraced interdisciplinary collaboration both with the artistic community and human perception researchers: in the last decade



Non-Photorealistic Rendering, Fig. 1 Illustration of a heart inferior wall dysfunction (short and long axis views). By Patrick J. Lynch (medical illustrator) and C. Carl Jaffe (MD, cardiologist)

as applications like augmented reality move away from photorealism in a wish for conveying additional or specific information (e.g., overlaid pathology data during a surgery intervention), a strong need arises for checking that the depiction that they do provide faithfully communicates the desired information.

This interaction between art, perception, and NPR is widespread in the literature. Researchers have studied how humans paint or draw in order to derive models which predict the success of NPR algorithms [4]. For instance, by combining several depictions of the same object made by different persons, it is known that most of the strokes can be predicted thanks to three-dimensional features such as contours or curvature and thus are automated by line drawing methods like *suggestive contours* [5]. However it is also evident in these studies that even the most recent and sophisticated methods fail to predict all the elements used by artists to draw an object.

Artistic styles have also been analyzed and, when possible, simulated from the physical media (e.g., simulation of fluids for paint) to the gamut of colors or the strokes and the shapes [6]. Rendering in a specific style usually implies a combination of multiple NPR techniques: Fig. 2

shows some of the image processing methods used for rendering stylized depictions. Although most of these techniques still require a significant amount of user input (in choosing and combining multiple parameters), in recent years, some approaches for automatic abstraction or color and histogram matching have been proposed [7, 8].

Classification

The works in the field are usually classified by their input system (2D, 3D) and type of user input (assisted or automatic):

- **From 3D to 2D:** Given a 3D geometric description of the objects, their material properties and light sources, a 2D image is generated according to a non-photorealistic style. For instance, in the cel shading technique, hand-drawn cartoon style is mimicked by adding a black contour line (occluding contour) based in the culling of front faces in the geometric model and illuminating the object with a flat shading obtained by limiting the range of colors to 2 or 3 t. The continuous

N

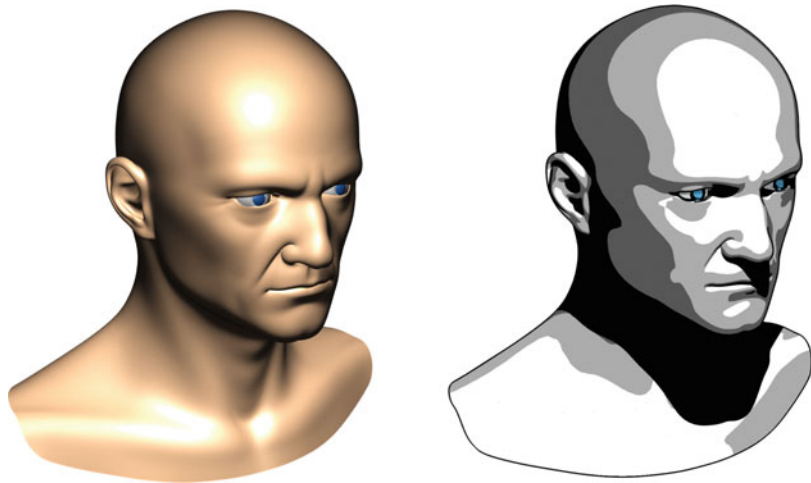


Non-Photorealistic Rendering, Fig. 2 Some examples of 2D NPR stylization. *Top row:* input images. *Bottom row:* resulting depictions, from *left to right* – oil strokes

simulation based on flow integration (LIC), edge detection with color posterization, and abstraction based in hue shift and pixel splattering

Non-Photorealistic

Rendering, Fig. 3 Toon shading render generated with available commercial software. The 3D model in the *left* was created with a subsurface scattering shading technique. The NPR version in the *right* includes shadows, contour lines, and three shades of gray



value yielded by traditional shading formulae (based on the cross product of the vertex normals and the direction of the light source) is thresholded to a limited set of values to produce this effect (Fig. 3). Nowadays, most of these rendering effects are generated as a postprocessing step in image space. Surface normals, depth, and reflectance values are computed per pixel and stored as images in different layers (a technique known as deferred shading). In this context, image processing is used to the same effect. For instance, contour lines can be generated by detecting discontinuities in the normal image with an edge filter such as Canny's detector [9].

- **From 2D to 2D:** This is usually based on image processing techniques and no user intervention. Most of the commercially available photo cameras provide at least a few of these transformations. Given a photograph as input, a purely 2D cel shading effect is obtainable by a combination of histogram thresholding (*posterize*) for color and a contour detector (Canny) for the contour lines. The simulation of brush and pencil strokes by using *line integral convolution* (LIC) [10] is another well-known example, which integrates random sampling from the luminance vector field of the original image in order to generate lines emulating the most likely strokes done by artists (see bottom left image in Fig. 2). Saliency and feature detection are also taken into

account to emphasize those strokes which have greater visual significance.

- **From 2.5D to 2D:** That is, 2D plus additional data such as incomplete 3D information (depth range or *Z-buffer*) and/or user annotations – semantic tags (*this part of the image is a cloud*), strokes (which encode information in its shape), marks (*this part of the object is on top to the other*), etc. Some recent techniques [11] include perception assumptions (such as *global convexity* and *saliency*) in order to infer depth from 2D images. This depth can further be edited (sculpted) with brush strokes or directly used in combination with traditional 3D-to-2D algorithms to render NPR images like those in Fig. 4.
- A fourth possibility is **from 2D to 3D**, that is, from a set of user strokes or images (or a combination of both) of a given scene, a 3D model is derived which allows for novel views. This conforms a field by itself called *image-based modeling and rendering* (IBMR), which lies between the areas of computer graphics and computer vision.

Examples

Since the early adoption of cel shading in the title *Jet Set Radio* (year 2002), hundreds of examples of NPR techniques have emerged in video games. Likewise, NPR is ubiquitous in animation and



Non-Photorealistic Rendering, Fig. 4 Example of perception-based NPR rendering from Ref. [11]. *Top left:* original photo. *Top right:* multitone depiction obtained

from hallucinated depth and virtual light sources. *Bottom left:* additional comic book style result with two layers of dynamic lines added. *Bottom right:* NPR color relighting

N

films. The *South Park* series is an unexpected example of 2D cartoon style derived from 3D models. A case of 2D real-life footage to 2D hand-drawn animation style is shown in the film *A Scanner Darkly*, where key frames were roto-scoped and vectorized, automatically interpolating the frames in between. However, although some automation processes were introduced, this production required a significant amount of user input.

Cross-References

- ▶ [Color Harmony](#)
- ▶ [Compositing and Chroma Keying](#)
- ▶ [Global Illumination](#)
- ▶ [Optical Art](#)
- ▶ [Sampling Problems in Computer Graphics](#)

References

1. Winkenbach, G., Salesin, D.H.: Computer-generated pen-and-ink illustration. In: ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH, pp. 91–100 New Orleans, Louisiana, USA (2008)
2. Fekete, J-D., Salesin D. (eds.): Proceedings of the 1st International Symposium on Non-photorealistic Animation and Rendering. Annecy, France. ACM (2000)
3. Durand, F.: An invitation to discuss computer depiction. In: Proceedings of the 2nd International Symposium on Non-photorealistic Animation and Rendering. Annecy, France, pp. 111–124. ACM (2002)
4. Cole, F., Golovinskiy, A., Limpaecher, A., Barros, H. S., Finkelstein, A., Funkhouser, T., Rusinkiewicz, S.: Where do people draw lines? Commun. ACM **55**, 107–115 (2012)
5. DeCarlo, D., Finkelstein, A., Rusinkiewicz, S., Santella, A.: Suggestive contours for conveying shape. ACM Trans. Graph. **22**(3), 848–855 (2003)
6. Gooch, B., Gooch, A.: Non-photorealistic Rendering. A K Peters, Natick (2001)
7. Zhao, M., Zhu, S.: Sisley the abstract painter. In: Proceedings of the 8th International Symposium on

Non-photorealistic Animation and Rendering Annecy, France, pp. 111–124. ACM (2010)

8. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. *IEEE Comput. Graph. Appl.* **21**(5), 34–41 (2001)
9. Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698 (1986)
10. Cabral, B., Leedom, L.: Imaging vector fields using line integral convolution. In: *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '93*, pp. 263–270. Anaheim (1993)
11. Lopez-Moreno, J., Jimenez, J., Hadap, S., Anjyo, K., Reinhard, E., Gutierrez, D.: Non-photorealistic, depth-based image editing. *Extended papers from NPAR 2010. Comput. Graph.* **35**(1), 99–111 (2011)

Non-Visual Lighting Effects and Their Impact on Health and Well-Being

Mariana Figueiro

Lighting Research Center, Troy, NY, USA

Synonyms

[Circadian rhythms](#); [Circadian system](#)

Definition

Biological, non-visual effects of lighting constitute a new field within lighting research and education. In addition to visual effects, light is known to affect other biological rhythms, most notably the circadian system, which generates and regulates a number of rhythms that run with a period close to 24 h. Light/dark patterns incident on the retina are the major synchronizer of circadian rhythms to the 24-h solar day. Lighting characteristics (quantity, spectrum, timing, duration, and distribution) affecting the visual and circadian systems differ. Symptoms of circadian sleep disorders, such as seasonal affective disorder, jet lag, and delayed sleep phase disorder, can be mitigated by timed light exposure.

Introduction

The neurophysiology and neuroanatomy of the human visual system is largely understood: vision is the result of complex interactions between light sources, objects and surfaces, the eye, and the brain. Lighting engineers and practitioners have developed technologies, standards, measurement devices, and applications based on this knowledge, and one of the primary goals of lighting design is to increase performance and productivity by improving vision and perception. Recent scientific discoveries, however, have added another dimension to lighting: namely, that *light is not just for vision anymore*. Specifically, scientists have discovered that daily light/dark patterns reaching the retina (back of the eye) play an important role in human health and well-being because they regulate our bodies' circadian rhythms. Circadian rhythms are every rhythm in our body that repeat approximately every 24 h. The circadian system is profoundly important for many human behaviors, as well as human well-being.

This entry provides basic knowledge about the non-visual effects of light entering the eye, focusing exclusively on the effects of light on human circadian and diurnal rhythms. Also discussed here are the lighting characteristics that affect these rhythms, as well as applications in which light/dark cycles have been experimentally shown to affect them.

Overview of the Circadian System

The Earth rotates around its axis, and, as a result, all creatures exposed to daylight on Earth experience a 24-h cycle of light and dark. Living organisms have adapted to this daily rotation of the Earth by developing biological rhythms that repeat at approximately 24-h intervals [1]. These rhythms are called circadian rhythms, from the Latin *circa* (about) and *dies* (day).

Biological Clock

Circadian rhythms are generated internally by the body, yet they are constantly aligned with the

environment by factors that are external to the body, mainly light/dark cycles. In mammals, circadian rhythms are regulated by an internal biological clock located in the suprachiasmatic nuclei (SCN) of the hypothalamus of the brain. The biological clock in humans has a natural period that is slightly greater than 24 h, and environmental cues, such as light/dark cycles, social activities, and meal times, can reset and synchronize the clock daily, ensuring that our behavioral and physiological rhythms are synchronized with the external environment.

Output Rhythms

Output rhythms are behavioral and physiological rhythms regulated by the biological clock, such as the sleep/wake cycle, alertness, core body temperature, locomotor activity, feeding and drinking behavior, and hormone production. Of special interest here are the sleep/wake cycle, alertness, core body temperature, and hormone production, especially melatonin production.

Sleep/Wake Cycle and Alertness

The sleep/wake cycle is one of the most prominent circadian rhythms. We are under the influence of two opposing systems: the sleep drive (homeostatic) and the alerting force (circadian). The sleep drive and the alerting force are distinct forces, independent from each other, though complementary, ensuring that we are asleep at night and awake during the day. The sleep drive, for example, is low when we get up, increasing steadily throughout the waking day, before diminishing rapidly within the first hours of sleep. The alerting force is regulated by the biological clock and follows a circadian rhythm, reaching a peak during the early evening and a trough during the second half of the night. The interaction between the sleep drive and the alerting force determines when we fall asleep and how well we sleep at night. Alertness is strongly correlated with the sleep/wake cycle.

Core Body Temperature

Core body temperature also follows a circadian pattern. It is high during the day, reaching a peak in the early evening, and low at night, reaching its

low point about 2 h before one naturally wakes up. Core body temperature is typically in phase with alertness and has a negative correlation with the hormone melatonin.

Hormone Production

Although there is evidence that the biological clock influences a variety of hormones, this discussion will be limited to melatonin production by the pineal gland. The pineal gland is located near the center of the brain, is about the size of a pea, and has a shape resembling that of a pinecone. It is believed that the primary function of the pineal is to convey light/dark information to the body via secretions of the hormone melatonin [2]. Melatonin is easily absorbed into the bloodstream, which makes it an ideal chemical messenger of time of day information for the entire body. Melatonin also participates in the transmission of information concerning day length or photoperiod for the organization of seasonal responses in animals (e.g., breeding).

Melatonin is produced during the subjective night (i.e., inactive period in diurnal species and active period in nocturnal species) and under conditions of darkness. As a result, melatonin levels are low during the day and high at night (in the dark). Although not yet well established, many researchers believe that melatonin can induce sleep in diurnal mammals, including humans, by acting on the SCN and reducing the alerting force, or the “wake-promoting” signal sent by the biological clock during the daytime. In addition, melatonin has an inverse relationship with core body temperature, which is closely linked to waking and sleep times. Peak melatonin levels typically occur slightly before core body temperature bottoms out. When external cues are absent, melatonin starts being endogenously produced by the body at around 9–10 p.m. (or about 2 h prior to sleeping) and stops being produced at around 7–9 a.m. Maximum melatonin levels occur between 2 and 4 a.m. [2].

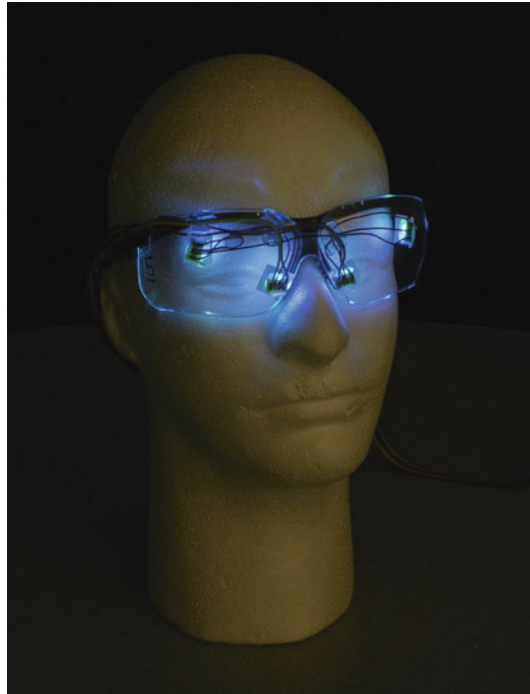
Regulation of Circadian Rhythms by Light

The 24-h light/dark cycle is the biological clock’s main synchronizer to the solar day. Light can

phase advance or phase delay human circadian rhythms, depending upon when it is applied [3]. For example, light that is applied before the minimum core body temperature, which is reached approximately 1.5–2.5 h before we naturally awaken, will delay the clock (e.g., one will wake up later the following day), and light applied after minimum core body temperature is reached will advance the clock (e.g., one will wake up earlier the following day). Although light is the main synchronizer of the biological clock to the solar day, it is not the only one. Exercise, social activities, timing of the sleep/wake cycle, and scheduled meals have also been shown to shift and synchronize the clock, although their impact on circadian rhythms seems to be weaker than the impact of light/dark cycles.

Lighting Characteristics Affecting Circadian Rhythms

The neural machinery in the mammalian retina provides light information to both the visual and circadian systems, but the two systems process optical radiation (light) differently [3]. Rods, cones, and a newly discovered photoreceptor, the intrinsically photosensitive retinal ganglion cells (ipRGCs) [4], participate in circadian phototransduction (how the retina converts light signals into neural signals for the biological clock). The quantity of polychromatic “white” light necessary to activate the circadian system is at least two orders of magnitude greater than the amount that activates the visual system. The circadian system is maximally sensitive to short-wavelength (“blue”) light, with a peak spectral sensitivity at around 460 nm (Fig. 1), while the visual system (i.e., acuity) is most sensitive to the middle-wavelength portion of the visible spectrum, with a peak at around 555 nm. Operation of the visual system does not depend significantly on the timing of light exposure and, thus, responds well to a light stimulus at any time of the day or night. On the other hand, the circadian system is dependent on the timing of light exposure, as discussed above. In addition, while the visual system responds to a light stimulus very quickly (in milliseconds), the duration of light exposure needed to affect the circadian system can take

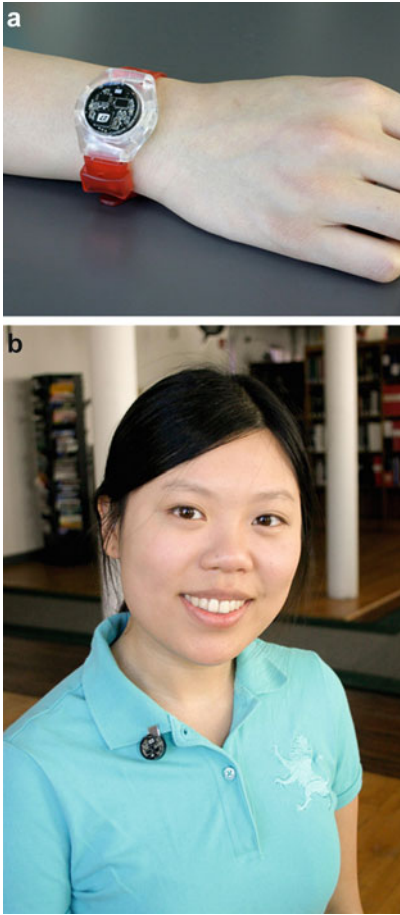


Non-Visual Lighting Effects and Their Impact on Health and Well-Being, Fig. 1 Photo of light goggles engineered by the Lighting Research Center. *Blue*, short-wavelength light has the greatest impact on the circadian system

minutes. For the visual system, spatial light distribution is critical for good visibility. It is not yet well established how light incident on different portions of the retina affects the circadian system. It is also important to note that the short-term history of light exposure affects the sensitivity of the circadian system to light; the higher the exposure to light during the day, the lower the sensitivity of the circadian system to light, as measured by nocturnal melatonin suppression and phase shifting.

Measuring Light for the Circadian System

Given that lighting characteristics affecting the visual system are different from those affecting the circadian system and that commercial light meters are calibrated to measure light for the visual system, new measurement devices that properly characterize light for the circadian system are needed. Two similar devices, the Daysimeter-S and the Daysimeter-D (also known



Non-Visual Lighting Effects and Their Impact on Health and Well-Being, Fig. 2 (a, b) Photo of the Daysimeter-D worn on the wrist and as a pin

as Dimesimeter), have been developed [5]. They are personal circadian light meters that measure light that is effective for visual and circadian responses. The Daysimeter-S has the photosensor package positioned near the plane of one cornea, and the Daysimeter-D can be worn as a pin, as a pendant, or on the wrist (Fig. 2). These photosensors are calibrated to measure “circadian” light. They also provide accurate measurements of light for vision (photopic lux). The photosensor package closely matches the eye’s acceptance of angles of light. Since circadian rhythms follow a 24-h pattern, it is necessary to measure light (and dark) over extended time periods to ascertain the stability and period of circadian rhythms. To this end, the Daysimeters

can be deployed in the field to gather light data for up to 30 days with a 30-s sampling interval.

Applications

In the following section, practical situations in which the effects of light on the circadian system could be beneficial to human health and well-being will be discussed. It is important to note that light is not currently known to cure any diseases. Further, individual differences need to be considered when designing lighting for the circadian system, so the lighting recommendations discussed below should be used as a framework only, and adjustments must be considered. It is also important to emphasize that few of the studies using light as a non-pharmacological treatment to various diseases and disorders reported the spectral power distribution of the light source, which makes it difficult to quantify the impact of light on the retina for circadian effectiveness. Consequently, generalizations from these studies are often more qualitative (e.g., bright vs. dim) than quantitative.

Tentative application tips will be made and the quantity, spectrum, duration, and timing of the light presented at the eye will be discussed. In order to simplify and standardize the tips offered here, **bright white light** will be defined as 1-h exposure of at least 600 lx measured at the eye level from a 6,500 K light source, unless otherwise stated in the text. This light level is about six times the amount of light one typically finds in an office environment without a window. “Blue” light is defined here as 1-h exposure to 40 lx at the cornea from a narrowband light source peaking at 470 nm. If duration of exposure is increased, light levels may be decreased, but it is recommended that it is never below 300 lx at the cornea for the white light and 20 lx at the cornea for the blue light. The lighting recommendations offered here are based on current knowledge from research findings, but results of future research will refine these values.

Seasonal Affective Disorder

Seasonal affective disorder (SAD) is a subtype of depression, with episodes occurring during winter

months and remitting during summer months. Symptoms of SAD include depression, hypersomnia, and weight gain due to increased carbohydrate cravings, social withdrawal, and even suicidal thoughts. It is believed that because daylight availability decreases in the winter at high latitudes, the number of people experiencing SAD increases as the latitude increases. “The winter blues” is an even more common subtype of SAD. The mechanisms of SAD are still unknown, and there are several competing hypotheses as to what causes SAD and how light can be used as a treatment. One of these is that late daybreak during winter months delays the circadian rhythms of those more susceptible to SAD; in this case, morning light is believed to be effective in treating symptoms of SAD. Another hypothesis is that the overall melatonin production of those suffering from SAD is greater during winter months than during summer months, which extends the amount of time during the 24-h day that their bodies think it is nighttime. In this case, light in the early morning or evening is recommended.

If a person is formally diagnosed with SAD by a general practitioner, insurance companies may pay for the cost of light treatment devices. A recent study [6] showed that approximately 500 lx of blue light directed at the eye ($\lambda_{\text{max}} = 470 \text{ nm}$) was able to significantly improve SAD symptoms compared to red light – used for placebo control. It has been suggested, however, that the positive impact of light on SAD symptoms is simply a result of placebo effects.

Practical Tips

- Encourage SAD patients to go for a half-hour walk outdoors in the morning (right after daybreak).
- Expose individuals to bright white light (or “blue” light) in the morning or evening.

Jet Lag

Jet lag is a temporary desynchronization between the biological clock time and the environmental time (light/dark). The symptoms include insomnia and/or hypersomnia, fatigue, poor performance, and gastrointestinal problems. Eastward travel

generally results in difficulty falling asleep, and westward travel results in difficulty maintaining sleep. Adaptation to a new time zone is usually slower after eastward travel than after westward travel. This is because: [1] those traveling east need to advance their biological clock to readjust to local time at their destination; the time that daylight is available upon arrival at the final destination will promote phase delay of the biological clock, and [2] it is easier for the timing of the biological clock to be delayed than advanced. One study [7] showed that a combination of advancing sleep schedules for 1 h per day plus morning light treatment (one half hour of 5,000 lx + half-hour of less than 60 lx at the cornea) for 3.5 h advanced the phase of the biological clock, by 1.5–1.9 h in 3 days. Although the principles for applying light treatment for reducing jet lag symptoms are known, the implementation of the light treatment may be a challenge. Airlines are starting to use colored light inside airplanes to improve mood, but it is probably very difficult to shift the biological clock to promote complete adjustment to a new time zone while one is inside the plane. Because the circadian clock is slow to shift, users need to start treatment a few days before they are scheduled to travel. However, because travelers have busy schedules, the likelihood of compliance is low. A personal light treatment device could be developed to increase the likelihood of compliance (Fig. 3).

Practical Tips

- Eastward travel: Upon arrival at your destination, avoid bright white light (including daylight) during the morning, until about noon local time, and seek exposure to bright white (or “blue”) light in the afternoon. Upon arrival, wear glasses that filter out radiation below 520 nm (“orange” glasses; Fig. 4) to avoid circadian effective light because the circadian system is maximally sensitive to short-wavelength (“blue”) light.
- Westward travel: Upon arrival at your destination, seek exposure to bright white light (or “blue” light) during the daylight hours and avoid bright white light in the evening.

Non-Visual Lighting Effects and Their Impact on Health and Well-Being, Fig. 3

Photo of a possible light box design that could aid in the adaptation of travelers' circadian systems to new time zones

**Non-Visual Lighting Effects and Their Impact on Health and Well-Being, Fig. 4**

Orange goggles are able to block short-wavelength light that maximally activates the circadian system. These goggles should be worn at times when one wants to reduce circadian effective light exposures



N

Delayed Sleep Phase Disorder

Delayed sleep phase disorder (DSPD) is a disorder of sleep timing; people suffering from DSPD typically go to bed late and wake up late (3–6 h later than typical sleeping hours). This pattern interferes with people's normal functioning because they have difficulty waking up in the morning for work, school, and social obligations, and since they go to bed late, they do not sleep for as many hours as those going to bed at more normal hours. DSPD in adolescents is common and probably associated with hormonal changes that occur at puberty. The exact causes of DSPD are not actually known, but light exposure after minimum core body temperature and dim light

during the evening have been shown to advance the phase of the biological clock of persons with DSPD [8]. Using this information, field studies [9] were conducted to investigate the impact of light exposures on dim light melatonin onset (DLMO), a primary marker for the timing of the biological clock, and on sleep duration for two populations of eighth graders. It was hypothesized for one study conducted in North Carolina that the lack of morning short-wavelength light (which was removed by wearing orange goggles; Fig. 4) would delay the timing of the students' biological clocks. For the other study conducted in New York, it was hypothesized that exposure to more evening light in spring relative to winter would

also delay the biological clocks of adolescents. In both studies, as expected, the students exhibited delayed DLMO as a result of removing short-wavelength morning light and as a result of seasonal changes in evening daylight. Also as expected, both sets of adolescents exhibited shorter sleep times; because of the delay in the timing of the biological clock, they fell asleep later but still had to get up at a fixed time in the morning. These two field studies clearly demonstrate that by controlling circadian light exposures, it is possible to practically and effectively control circadian time and thereby affect meaningful outcomes like sleep duration.

Practical Tips

- Expose individuals to bright white (or “blue”) light in the morning, after the minimum core body temperature. Note that minimum core body temperature of people with DSPD may occur as late as 9 a.m.; therefore, an understanding of the characteristics of the biological clock of those using the light treatment is needed before recommending the timing of the light exposure. Just as a tip, minimum core body temperature generally occurs 1.5–2.5 h prior to waking without an alarm clock.
- Wear glasses that filter out radiation below 520 nm (“orange” goggles; Fig. 4) in the evening and in the morning, before minimum core body temperature is reached. Because the circadian system is maximally sensitive to short-wavelength light, elimination of optical radiation below 520 nm at these specific times will reduce the chances of phase shifting the biological clock to a later time, aggravating symptoms of DSPS even further.
- Avoid the use of very bright self-luminous electronic devices (e.g., computer screens, tablets, and cell phones) during the evening hours. At a minimum, dim them down or filter their screens with “orange” filters.

Sleep in Older Adults

Sleep disturbances in older adults. Seniors living in controlled environments (assisted living and nursing homes) are perhaps the best example of a population at risk for circadian sleep disorders;

due to age-dependent reduced retinal light exposures and to fixed lighting conditions in their living environments, seniors are less likely to experience the necessary, robust 24-h, light/dark pattern needed for circadian entrainment. Prescribed 24-h light/dark patterns have been shown to influence circadian rhythms and thereby alleviate some sleep and agitation issues common among seniors, including those with Alzheimer’s disease (AD). A 24-h lighting scheme has been proposed that delivers high circadian stimulation during the daytime hours, low circadian stimulation in the evening hours, and night lights that provide perceptual cues to decrease the risk of falls at night [10], such as those shown in Fig. 5. Studies showed that exposure to bright white light (ranging from 2,500 lx to as high as 8,000 lx at the cornea) for at least 1 h in the morning for a period of at least 2 weeks was found to improve or consolidate nighttime sleep of AD patients. Greater sleep efficiency at night decreased the need to sleep during daytime hours and, in some cases, reduced agitated behavior such as pacing, aggressiveness, and speaking loudly. Research demonstrated that exposing 22 dementia patients to continuous, bright, indirect white light (average of 1,136 lx at the eye) over 4 weeks consolidated rest/activity rhythms of people with AD [11]. In a recent long-term study, it has been demonstrated that light attenuated cognitive deterioration by 5 % on the mini-mental state examination. Light also ameliorated depressive symptoms by 19 % on the Cornell Scale for Depression in Dementia and attenuated the increase in functional limitations over time by 53 % on the nurse-informant activities of daily living scale or a relative 53 % difference [12].

Practical Tips

- For older adults who want to fall asleep later in the evening, expose them to bright white (or “blue”) light in the late afternoon/early evening (before 7 p.m.), before minimum core body temperature. It is recommended that the duration of the exposure be increased to 2 h or the quantity of light be increased by a factor of two because of the optical changes to the aging eye.

Non-Visual Lighting Effects and Their Impact on Health and Well-Being, Fig. 5 Novel nightlighting system using horizontal/vertical cues that can improve postural control and stability



- Wear glasses that filter out radiation below 520 nm (“orange” goggles) in the very early morning hours (before 7–8 a.m.).
- For AD patients who do not have a consistent sleep/wake schedule, expose them to bright white (or “blue”) light at any time during the day and reduce evening light exposures. Be consistent with the timing of the light exposure, however. Because the circadian phase of AD patients is not known and because they tend to remain in continuous dim light all day, it is expected that a robust light/dark pattern will positively impact their sleep.
- Install safe nightlighting systems that provide low-level illumination and perceptual (horizontal/vertical) cues to aid in controlling postural control and stability (Fig. 5).

Entrainment for the General Population

As shown in the examples above, lighting can help reduce or mitigate symptoms associated with various circadian sleep disorders. Lighting can also aid in maintaining entrainment in the general, healthy population. Because humans

have a biological clock that runs with a period slightly greater than 24 h, we need daily morning light exposure to maintain entrainment to the solar day. In winter months, some of us can go to and come back from work or school in darkness; therefore lighting in our work environments should provide enough circadian stimulation during at least the morning hours. Although the link between circadian entrainment and productivity is yet to be established in larger, clinical studies, one study [13] to date showed that compared with a 4,000 K light source, blue-enriched white light (17,000 K) improved subjective measures of alertness, positive mood, performance, evening fatigue, irritability, concentration, and eye discomfort. Daytime sleepiness was reduced, and the quality of subjective nocturnal sleep was improved under blue-enriched white light. Although future studies should confirm and extend these results, it is suggested that a dynamic lighting system that delivers bright white light (or “blue” light) to office workers during the morning hours can help maintain entrainment and possibly improve well-being, alertness, and performance, especially in winter months, when duration of daylight availability is short [14].

The Future

This overview of the impact of light on the circadian system and its effects on our health and well-being underscores the importance of developing a new framework for lighting practices that includes not only those lighting characteristics that affect the visual system but also those that affect the circadian system. Because there are great differences between the two systems' responses to the quantity of light, its spectral composition, spatial distribution, timing, and duration, generalizations about "quality lighting" will have to be assessed by two very different sets of criteria in the future. Although the information and recommendations presented here will certainly be refined as more research is undertaken, little progress will be made in delivering "healthy lighting" to society until researchers and practitioners begin to consider, measure, calculate, and control the fundamental characteristics of light for the circadian system. Hopefully, this entry will serve as an important step toward that goal.

Finally, it is important to remember that we have no practical way to intuitively know when and what type of light we need for circadian entrainment. Like other medical monitors where we have no conscious access to what we need for good health (e.g., glucose monitors for diabetes), we need technologies to measure and track the state of our circadian system. A wireless technology connecting the Daysimeter to a smartphone is currently under development. A smartphone (or similar technology) will be able to interpret the light/dark patterns to which a person is exposed and then be able to recommend, based on models of human circadian entrainment [15], both current and future light exposure patterns for the user to maintain circadian entrainment. One day, the smartphone might be able to interface with the building lighting system itself so that children in a school, nurses in a hospital, or people at home would be exposed to the proper lighting conditions for maintaining well-being and for increasing productivity.

Acknowledgments The author would like to acknowledge the projects' sponsors (National Institute on Aging,

National Institute of Nursing Research, National Institute on Drug Abuse, National Cancer Institute, and Office of Naval Research). Mark Rea, PhD, of the Lighting Research Center is acknowledged for his technical assistance, and Nicholas Hanford of Rensselaer Polytechnic Institute is acknowledged for his editorial assistance.

Cross-References

- [Circadian Rhythms](#)
- [Light-Emitting Diode, LED](#)

References

1. Moore, R.Y.: Circadian rhythms: basic neurobiology and clinical applications. *Annu. Rev. Med.* **48**, 253–266 (1997)
2. Arendt, J.: Melatonin and the Mammalian Pineal Gland. Chapman & Hall, London (1995)
3. Rea, M.S., Figueiro, M.G., Bullough, J.D.: Circadian photobiology: an emerging framework for lighting practice and research. *Light. Res. Technol.* **34**, 177–187 (2002)
4. Berson, D.M., Dunn, F.A., Takao, M.: Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002)
5. Figueiro, M.G., Hammer, R., Bierman, A., Rea, M.S.: Comparisons of three practical field devices used to measure personal light exposures and activity levels. *Light. Res. Technol.* **45**, 421–434 (2013)
6. Glickman, G., Byrne, B., Pineda, C., Hauck, W.W., Brainard, G.C.: Light therapy for seasonal affective disorder with blue narrow-band light-emitting diodes (LEDs). *Biol. Psychiatry* **59**, 502–507 (2006)
7. Eastman, C.I., Gazda, C.J., Burgess, H.J., Crowley, S. J., Fogg, L.F.: Advancing circadian rhythms before eastward flight: a strategy to prevent or reduce jet lag. *Sleep* **28**, 33–44 (2005)
8. Rosenthal, N.E., Joseph-Vanderpool, J.R., Levendosky, A.A., Johnston, S.H., Allen, R., Kelly, K.A., Souetre, E., Schultz, P.M., Starz, K.E.: Phase-shifting effects of bright morning light as treatment for delayed sleep phase syndrome. *Sleep* **13**, 354–361 (1990)
9. Figueiro, M.G., Rea, M.S.: Lack of short-wavelength light during the school day delays dim light melatonin onset (DLMO) in middle school students. *Neuro. Endocrinol. Lett.* **31**, 92–96 (2010)
10. Figueiro, M.G.: A proposed 24 h lighting scheme for older adults. *Light. Res. Technol.* **40**, 153–160 (2008)
11. Van Someren, E.J., Kessler, A., Mirmiran, M., Swaab, D. F.: Indirect bright light improves circadian rest-activity rhythm disturbances in demented patients. *Biol. Psychiatry* **41**, 955–963 (1997)

12. Riemersma-van der Lek, R.F., Swaab, D.F., Twisk, J., Hol, E.M., Hoogendijk, W.J.G., Van Someren, E.J.W.: Effect of bright light and melatonin on cognitive and noncognitive function in elderly residents of group care facilities: a randomized controlled trial. *JAMA: J. Am. Med. Assoc.* **299**, 2642–2655 (2008)
13. Viola, A.U., James, L.M., Schlengen, L.J., Dijk, D.J.: Blue-enriched white light in the workplace improves self-reported alertness, performance and sleep quality. *Scand. J. Work Environ. Health* **34**, 297–306 (2008)
14. Van Bommel, W.J.M.: Non-visual biological effect of lighting and the practical meaning for lighting for work. *Appl. Ergon.* **37**, 461–466 (2006)
15. Kronauer, R., Forger, D., Jewett, M.: Quantifying human circadian pacemaker response to brief, extended and repeated light stimuli over the photopic range. *J. Biol. Rhythms.* **14**, 500–515 (1999)

O

Op Art

- ▶ [Optical Art](#)

Opacity – Extinction

- ▶ [Apparent Magnitude, Astronomy](#)

Opalescence

- ▶ [Iridescence \(Goniochromism\)](#)

Opponent Color Theory

- ▶ [Color Vision, Opponent Theory](#)

Opsin Genes

- ▶ [Photoreceptors, Color Vision](#)

Optic Chiasm, Chiasmal Syndrome

Jonathan Aboshiha
The UCL Institute of Ophthalmology and
Moorfields Eye Hospital, London, UK

Synonyms

[Chiasma opticum](#)

Definition

The **optic chiasm** (from the Greek for “cross-piece”) is the anatomical location where the two optic nerves, having left the eyes, join together before then separating once more as the two distinct optic tracts, thus forming an anatomical “X.”

Chiasmal syndrome is the name given to the group of symptoms and signs that occur together as a result of lesions affecting the optic chiasm.

Overview

Optic Chiasm

This optic chiasm measures 15 mm in width and 3.5 mm in height [1] and has several important structures located adjacent to it; below the optic chiasm, in a bony excavation called the sella

turcica, lies the pituitary gland, and either side of it are found the venous cavernous sinuses, which transmit the internal carotid artery and several important cranial nerves. The frontal lobe of the brain sits above the chiasm, and the floor of the third ventricle lies behind. The optic chiasm is significant because it is here that the visual signals coming from each retina decussate, and so from this point onward, there is representation from both eyes on both sides (i.e., left and right sides) of the visual pathway. The nasal optic nerve fibers, carrying signals from the nasal retinae (and hence temporal hemi-fields of vision), cross or decussate to the other side (contralateral) at the optic chiasm, while the temporal fibers, which carry signals from the temporal retinae (and hence nasal hemi-fields of vision), continue on the same side (ipsilateral). In anatomically normal individuals, approximately 50 % of the axons (i.e., those from the nasal retina) decussate across the midline to join the uncrossed fibers traveling toward the brain from the temporal retina. These fibers exit the chiasm to form the left and right optic tracts, which carry signals from the right and left hemi-fields of vision, respectively. In certain conditions, the proportion of crossing fibers at the chiasm is significantly different from the 50 % value observed in normals; in humans with albinism, a greater number of fibers from temporal retina decussate and project contralaterally [2].

Chiasmal Syndrome

Pathology in any of the key anatomical structures that lie adjacent to the optic chiasm (see above) has characteristic effects on the optic chiasm, giving rise to a unique constellation of visual signs and symptoms, collectively known as the chiasmal syndrome. Depending on the particular pathology involving these adjacent structures, and the consequent relative location of damage to the chiasm, the exact features seen in any chiasmal syndrome can help anatomically locate the site of the lesion.

A visual field defect may be the earliest sign of a lesion at the optic chiasm [3], and the classical visual field defect seen in chiasmal syndrome is the loss of the temporal hemi-fields of vision in both eyes (so-called bitemporal hemianopia). This is caused by a lesion in the body of the chiasm

which damages the crossing nasal retinal fibers. Any lesion of the chiasm (such as a tumor, vascular lesion, traumatic lesion, etc.) may cause this, and there are subtle variations in the extent of the bitemporal hemianopia that can help give clues as to the causative lesion. The exact field defect will vary according to the type of lesion and the precise individual anatomical relationship of the chiasm to the structures that surround it. The most common cause of such a field defect is a tumor of the pituitary gland [3]. The visual acuity may also be reduced, as may the color vision on testing with pseudoisochromatic plates, with a tendency to have more severe red-green deficits and milder blue-yellow losses [4].

Lesions of the anterior chiasm can cause a so-called junctional scotoma, presenting with a central field loss in one eye and a superotemporal field loss in the other – this is classically said to be due to a small bundle of inferonasal nerve fibers that briefly loop forward in the contralateral optic nerve, called Wilbrand's knee. More recently, the existence of Wilbrand's knee has been questioned as an artifactual error, relating to the fact that the cadaveric specimens on which the observation of this anatomical feature had been based had always suffered the loss of one eye before death [5], and investigations have found no clinical perimetric evidence to support the existence of this anatomical feature in human beings [6].

Lesions of the posterior chiasm may cause an uneven field loss in both eyes of the visual field on the same side, due to damage to the optic tracts (i.e., an incongruous homonymous hemianopia).

Compression of the optic chiasm may also cause atrophy of the nasal nerve fibers, with the superior and inferior nerve fibers being relatively spared – so-called “bow-tie” optic atrophy – as well as an unusual type of nystagmus (an involuntary movement of the eyes) called “seesaw” nystagmus [7]. Loss of the temporal visual fields can lead to postfixation blindness and the inability to keep the eyes both facing in the same direction, a phenomenon known as hemi-field slide [8].

The optic chiasm also lies close to a venous space (the cavernous sinus) where the nerves that control eye movement pass – lesions in this area can damage these nerves and cause further

problems with the eyes moving together, resulting in diplopia (double vision).

Anyone presenting with the above features of chiasmal syndrome should be investigated with urgent neurological imaging, in order to promptly diagnose and, if necessary, treat a potentially life-threatening cause.

Cross-References

- [Pseudoisochromatic Plates](#)
- [Visual Appearance](#)

References

1. Parravano, J.G., Toledo, A., Kucharczyk, W.: Dimensions of the optic nerves, chiasm, and tracts: MR quantitative comparison between patients with optic atrophy and normals. *J. Comput. Assist. Tomogr.* **17**(5), 688–690 (1993)
2. Hoffmann, M.B., et al.: Organization of the visual cortex in human albinism. *J. Neurosci.* **23**(26), 8921–8930 (2003)
3. Foroozan, R.: Chiasmal syndromes. *Curr. Opin. Ophthalmol.* **14**(6), 325–331 (2003)
4. Yates, J., Diamantopoulos, I., Daumann, F.: Acquired (transient and permanent) color vision disorders. In: Menu, J. (ed.) *Operational Color Vision in the Modern Aviation Environment*. North Atlantic Treaty Organization Research and Technology Organization A, Neuilly-Sur-Seine Cedex (2001). 1
5. Horton, J.C.: Wilbrand's knee of the primate optic chiasm is an artefact of monocular enucleation. *Trans. Am. Ophthalmol. Soc.* **95**, 579 (1997)
6. Lee, J.H., et al.: Wilbrand's knee: does it exist? *Surg. Neurol.* **66**(1), 11–17 (2006)
7. Unsöld, R., Ostertag, C.B.: Nystagmus in suprasellar tumors: recent advances in diagnosis and therapy. *Strabismus* **10**(2), 173–177 (2002)
8. Kirkham, T.: The ocular symptomatology of pituitary tumours. *Proc. R. Soc. Med.* **65**(6), 517 (1972)

Optical Art

Zena O'Connor

Architecture, Design and Planning, University of Sydney, Sydney, Australia

Synonyms

[Op art](#); [Retinal art](#)

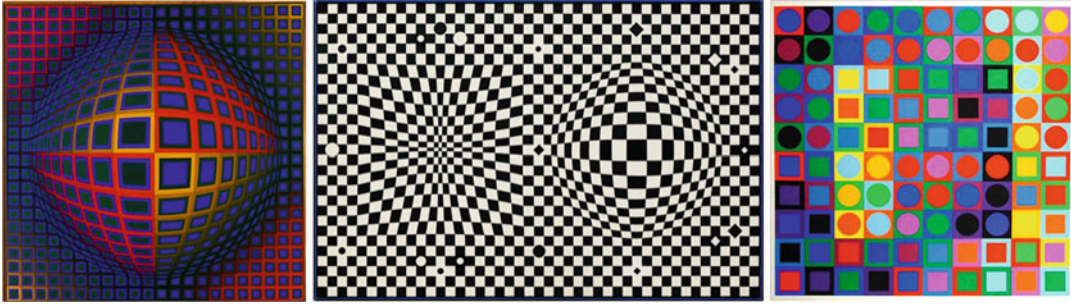
Definition

Optical art (or op art) was an art movement that emerged in the mid-twentieth century as an extension of the pop and conceptual art movements. It exploits the mechanics of human visual perception to create imagery in the viewer's mind. Op artists used line, shape, flat planes of color, and a two-dimensional format to create works that appear three dimensional and occasionally convey a sense of movement.

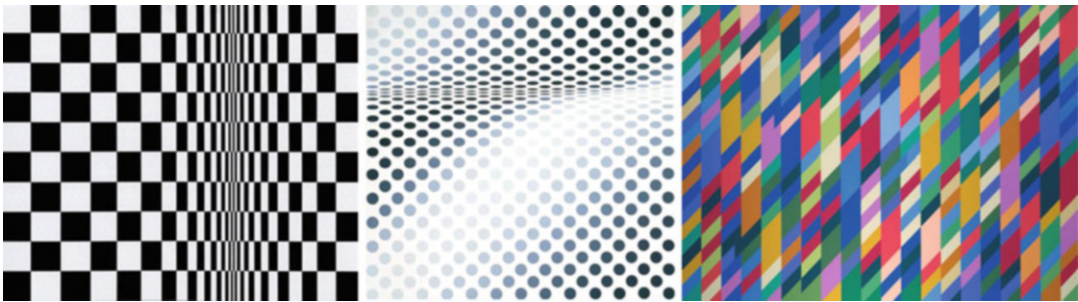
Overview

Optical or op art emerged as an extension of the pop art and conceptual art movements, both of which represented the transition from modernism to postmodernism, during the middle of the twentieth century. From a theoretical perspective, modernism rejected the accepted and established art forms but eventually became formulaic, prescriptive, and predictable, while in response, postmodernism became characterized by a more pluralistic, eclectic, diverse, and somewhat more unpredictable approach to art.

Op art “exploits the workings of perception to create a virtual reality in the viewer's mind” ([1], p. 178). Op artists aimed to create optical illusions using line, flat planes of color, and a two-dimensional format to create works that appear three dimensional and occasionally convey a sense of movement. Characterized by an emphasis on geometry, simple forms, and a strategic color ► [palette](#) (often black and white), op art also represented “geometric precision, emblematic not only of the Space Age but... (also the) technological revolution” of the 1960s ([2], p. 73). Color and light–dark contrast are dominant visual elements in op art and used strategically to create three-dimensional visual illusions from two-dimensional formats. In addition, the strategic use of color was also used to create a sense of movement across the canvas, and in this way, op art paintings represent outcomes that are more than the sum of their parts in terms of paint on canvas.



Optical Art, Fig. 1 Works by Victor Vasarely



Optical Art, Fig. 2 Works by Bridget Riley

Although she does not like the title “op art,” Bridget Riley’s works are closely associated with this movement as well as “the spirit of the mid-sixties”; and Victor Vasarely’s works are also closely aligned with the op art movement ([2], p. 72).

Victor Vasarely (1906–1997)

Op art works by Hungarian-French artist Victor Vasarely feature the illusion of three dimensionality, and this is achieved by the strategic placement of geometric shapes, color, and contrast upon the canvas. Vasarely first started to experiment with optical patterns in the 1930s and his works were included in the Museum of Modern Art’s exhibition: *The Responsive Eye*. This exhibition, which also included works by Bridget Riley, featured works that which were considered “less as objects to be examined than as generators of responses in the eye and mind of the viewer” ([3], p. 1). The works included in this exhibition featured works that used optical illusions, simultaneous contrast, and afterimages to produce “new kinds of

subjective experiences” ([3], p. 1). As a prolific painter, key works by Vasarely include *Vega Nor* (1969), *Vega III* (1957), and *Marc Positive* (1968) (Fig. 1).

Bridget Riley (Born 1931)

Works by English painter Bridget Riley feature the illusion of three dimensionality, and this is achieved by the careful placement of variations in shape as well as color or light–dark contrast. In addition, Riley’s works often feature a sense of movement, and this is achieved by the strategic placement of color around the canvas, causing the viewer’s eye to be drawn from dominant colors and contrasts around the canvas. Figure 2 features Riley’s *Movement in Squares* (1961), *Hesitate* (1964, Tate), and *Nataraja* (2000, Tate).

The art movements of op art as well as minimalism and pop art had a strong influence on popular culture as well as many areas of applied design in the late 1950s and 1960s [4–8]. Design output by Mary Quant, André Courreges, Pierre Cardin, and Terence Conran and firms such as

Marimekko are imbued with design characteristics drawn from op art and pop art. In addition, op art influenced graphic design at the time as evidenced by the International Wool Secretariat logo by Francesco Saroglia (1965).

Informative videos about op art include:

- Andrew Graham-Dixon on Bridget Riley video (6:30 min), <http://www.youtube.com/watch?v=Z1lQCTunGxg>
- Victor Vasarely video (2:54 min), <http://www.youtube.com/watch?v=uUSJ2XqBp6U>
- Op Art Avon TV advertisement, 1960s (1:09 min), <http://www.youtube.com/watch?v=ayBuJ0RtaWg>

Cross-References

- [Color Contrast](#)
- [Palette](#)
- [Texture Measurement, Modeling, and Computer Graphics](#)
- [Visual Illusions](#)

References

1. Bird, M.: One Hundred Ideas that Changed Art. Laurence King, London (2012)
2. Sandbrook, D.: White Heat: A History of Britain in the Swinging Sixties. Little, Brown, London (2006)
3. MoMA (Museum of Modern Art, New York): MoMA press release – the responsive eye. http://www.moma.org/pdfs/docs/press_archives/3439/releases/MOMA_1965_0015_14.pdf?2010 (1965). Accessed 10 Mar 2010
4. Garner, P.: Sixties Design. Taschen, Köln (1996)
5. Glancey, J.: Modern: A Portfolio of Contemporary Interior Design Styles. Mitchell Beazley, London (2000)
6. McDermott, C.: Twentieth-Century Design. Viking, London (1997)
7. Raizman, D.: History of Modern Design. Laurence King, London (2003)
8. Sparke, P.: A Century of Design: Design Pioneers of the 20th Century. Mitchell Beazley, London (1998)

Optical Radiation

- [Light, Electromagnetic Spectrum](#)

Optical Spectrum

- [Light, Electromagnetic Spectrum](#)

Optoelectronic Light Source

- [Light-Emitting Diode, LED](#)
- [Light-Emitting Diode, OLED](#)

Ostwald, Friedrich Wilhelm

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Wilhelm Ostwald was born on September 2, 1853, in Riga, the capital city of Latvia, to parents of German decent. He studied chemistry at the University of Dorpat (now Tartu, Estonia) where he received his PhD in 1878 and lectured at the Polytechnicum in Riga. In 1887, he moved to

the University of Dresden in Germany where he remained until his early retirement in 1906. He was active in many fields, including philosophy, is considered to be one of the founders of physical chemistry, and received the Nobel Prize for the discovery of chemical catalysis in 1909. Ostwald is the author of 45 textbooks and over 1000 articles. In his spare time, he was a talented painter. In 1905/1906, he spent several months in the Boston area giving courses in philosophy, physical chemistry, as well as techniques of painting at Harvard University, MIT, and the Lowell Institute, where he met Munsell and learned of the latter's early development of the Munsell color system. After he left the University of Dresden, he moved to his nearby countryseat in Grossbothen where he spent the rest of his life primarily working on color theory and developed a large color order system and a theory of color harmony. Ostwald died on April 4, 1932 Grossbothen [1].

Major Accomplishments/Contributions

In the total field of science, Ostwald saw the place of the science of color to be in psychological science [2]. He was not only fully aware of the work of Maxwell, Grassmann, and Helmholtz but was convinced that also Hering had made important contributions toward understanding of color phenomena. His immediate predecessors had focused primarily on the relationship between lights and color experience. Ostwald made significant contributions to the understanding of object colors. His main contributions in the color field are briefly mentioned below:

Color order system: Ostwald developed a color order system that combined psychological and psychophysical knowledge, resulting in *Grosser Farbenatlas* of 2500 systematically ordered samples published in 1917/1918 (Fig. 1). A reduced version of 600 samples was available in 1920. An American version of the system was published in the 1940s as *Color Harmony Manual* [3].

Distinction between unrelated and related colors: Helmholtz and others of his time investigated color primarily in terms of spectral light. Ostwald demonstrated and clarified the difference



Ostwald, Friedrich Wilhelm, Fig. 1 View of the double cone model of Ostwald's color atlas

between unrelated and related colors, the latter experienced from viewing objects in varying surroundings. He showed that colors such as gray, brown, olive, and others exist in related form only [2].

Nonlinear relationship between wavelength differences and perceived hue differences: Ostwald demonstrated the highly nonlinear relationship when establishing the hue circle of his color atlas, showing that there are two spectral regions where perceived hues change rapidly as a function of changes in wavelength, while near the beginning, middle, and end of the spectrum, the frequency of change is much reduced [4].

Farbenhalb/Vollfarben: Ostwald demonstrated that idealized complementary object colors have reflectance in half of the spectrum in one case and the other half in the other case, with the transition wavelength varying as a function of hue. He used the term *Farbenhalb* (half of the spectrum) for this situation. He used a graphic format with wavelength on the horizontal and reflectance between 0 and 1 on the vertical axis. Object colors represented by *Farbenhalb* he named *Vollfarben* (full colors), object colors of the highest saturation possible for a given hue, occupying spectral ranges with either one or two transitions in the spectrum. Full colors are the optimal object colors of a given hue at the lightness level at which they have highest saturation, varying by hue. In 1920, Schrödinger offered a mathematical proof of the

concept of optimal object colors that included Ostwald's *Vollfarben* [4].

Metamerism: In 1918, Ostwald demonstrated the existence of metameric object colors and introduced the term *metamerism* [4].

Textbooks on color: During his lifetime, Ostwald published several volumes on color. The most important ones are part of a planned five-volume work *Die Farbenlehre* (The Science of Color), of which only two volumes were published during his lifetime: Vol. I *Mathetische Farbenlehre*, 1918 (ordering of color), and Vol. II *Physikalische Farbenlehre*, 1919 (color physics).

References

1. Ostwald, W. www.wilhelm-ostwald.de. Accessed 3 Mar 2015; 3 June 2015
2. Ostwald, W.: Beiträge zur Farbenlehre. Abhandlungen der Mathematisch-Physikalischen Klasse der Königlich Sächsischen Gesellschaft der Wissenschaften **23**(3), 365–572 (1917)
3. Jacobson, E., Ostwald, W.: The Color Harmony Manual and How to Use It. Container Corp. of America, Chicago (1942)
4. Ostwald, W.: Die Farbenlehre. Physikalische Farben, vol. 2. Unesma, Leipzig (1918)

P

Paint

- [Pigment, Ceramic](#)

Painted Dress

- [Art and Fashion Color Design](#)

Paints

- [Pigment, Inorganic](#)

Pale Colors

- [Pastel Colors](#)

Palette

Roy Osborne
Color Research and Application, The Colour
Group of Great Britain, London, UK

Synonyms

[Color board](#); [Color range](#); [Color selection](#)

Definition

1. Board on which paints are set in order to mix them before applying to a painting
2. The color range or selection used in a composition, art work, or design object

Introduction

In its narrowest sense, a painter's palette is a rigid board on which a set of paints is laid out and adjacent to which is an area where its ingredients can be mixed before application to a canvas, board, or other support. In its widest sense a palette represents any range of colors an artist or designer has chosen for use in any specific project. For some tasks this may simply mean black and white with various greys mixed from them; for others it can mean an array of dozens of colors, with others mixed from them. For centuries, the palette has been an essential item in every painter's studio. Since the appearance of the Quantel Digital Paintbox (1981), many artists have replaced the paintbrush with the computer mouse, though the term "palette" is still retained to refer to the range of colors available for introduction into the visual image.

Palette, as a Board for Arranging and Mixing Paints

European palettes were traditionally made of hard wood (such as mahogany or walnut) and weighted

or shaped in order to balance comfortably on the wrist or forearm of the painter while working. From about 1400, a number of portraits show small wooden palettes with handles either laying flat on a table or propped up beside a set of brushes, commonly one for each color. The latter has corners to stop them rolling over and also imply viscosity in the paint being used. For larger works, and less viscous paint, shallow bowls of single colors and large brushes were used. After about 1500, for easel paint in either tempera or oil, a palette large enough to accommodate about a dozen colors is pierced in one corner by a thumb hole. After about 1600, large curved palettes that rest on the forearm become general and define a design that remained standard into the twentieth century, though smaller, rectangular palettes that fitted into the lids of paintboxes also continued to be used. Throughout the ages, many other materials have been employed to hold and mix paints, including shells, porcelain, glass, metal, plastic, or bone, as in the case of the scapula palettes of prehistoric times. Metal trays or containers are best for wax or encaustic painting (requiring heat), and glass jars are useful for fresco painting. Shallow, cupped palettes are suited to dilute egg tempera and watercolor paint, whereas flat boards are ideal for holding small quantities of oil, alkyd, and polymer paints.

Palette as a Color Range or Selection

The common function of a palette (whatever its size and format) is to reveal the selection of chosen colors at a glance and assist the painter in working with a degree of speed and efficiency. According to Giorgio Vasari (1568), Lorenzo di Credi (a fellow-pupil of Leonardo in the studio of Andrea del Verrocchio) “made on his palettes a great number of colour mixtures, so that they went gradually from the lightest tint to the darkest, with exaggerated and truly excessive regularity, so that sometimes he had 25 or 30 on his palette, and for each of them he kept a separate brush.” More often, color mixing was kept to a minimum, and

colors employed in the glorification of gods were kept as pure and unsullied as possible.

Historical Development of Palette

Of the many reference sources that document the historical variation and content of artists’ palettes, among the more informative are Letalle’s *Les palettes d’artistes* [1], Speed’s *The Science and Practice of Oil Painting* [2], Schmid’s *The Practice of Painting* [3], Bazzi’s *Abecedario pittorico* [4], and Birren’s *History of Color in Painting* [5].

Ancient Times

The limited range of the pigments used by the earliest painters can be ascertained from cave paintings from about 15,000 BC that have survived in northern Spain and western France. The palette employed is known to have included red ochre (hematite), yellow ochre (limonite), stibnite (grey antimony sulphide), and kaolin (white aluminium silicate). Remarkably, this four-pigment palette, known to the Greeks as the *tetrachromatikón*, not only survives in Europe but is found widely scattered throughout the globe, being a common feature in the traditional painting by the Polynesian artists of the Pacific and New Guinea, by many African tribes, and by the Australian Aborigines. Reds, greens, and yellows from vegetation may have been used but generally rapidly fade.

Ample evidence remains of the contents of the ancient Egyptian palette. In addition to fugitive dyestuffs, the Egyptians employed gypsum (calcium sulfate), white lead (lead carbonate), red lead (lead oxide), realgar (arsenic disulfide), orpiment (arsenic trisulfide), malachite green and azurite blue (two forms of copper carbonate), Egyptian blue (crystalline copper silicate), and various carbon blacks, including lamp black and bone black. Egyptian Blue has been identified in painted decorations from about 2500 BCE onward, and its manufacture combined chalk or limestone with sand and a copper mineral, such as azurite, fired in a kiln at precise temperatures.

Lapis lazuli was used in jewellery but is unlikely to have been employed as a pigment at this time.

In ancient Greece, the practice of intermixing the four earth colors was credited to the painters Polygnotus and Anaglaophon. The reliance on a palette of red ochre, yellow ochre, chalk white, and black (commonly from burnt wine lees) caused Pliny to remark, in his *Natural History* (37.12): *Quattuor coloribus solis immortalia illa opera fecere* (“Four colours alone to make their work immortal”). Apelles was acknowledged as the principal advocate of the tetrachromatikón, from which all human complexion colors could be obtained, plus the browns of hides and fur, and even the dull greens of foliage, when yellow ochre was mixed with bluish black. Additionally, Pliny writes of Apelles applying varnish to his paintings that “caused radiance in the brightness of all the colours and protected the painting from dust and dirt.” Pliny also makes a distinction between the “austere” colors (*austeri*) of the “four-earth” palette and vivid or “flowery” colors (*floridi*). Among the latter are red lead (minium), purpura lake (purpurissum), dragonsblood (cinnabaris), Armenian blue (azurite), chrysocolla, and Indigo.

Greeks and Romans used much the same palette, adapted from the Egyptians, and the only significant pigment introduced by the Romans was cinnabar (natural mercuric sulfide).

Medieval Age

After the fall of the Roman empire (about 400 CE), few surviving records refer to the constitution of artists’ palettes, owing perhaps to illiteracy or a desire to keep certain practices secret, or simply because processes were so well known that no one felt the need to write them down. Commissions and contracts might specify which pigments were to be used for particular artworks, usually with the aim of preventing the substitution of inferior colorants for precious ones similar in color, such as red lead for vermilion, or azurite for lapis lazuli, which at its finest was worth its weight in gold. The English word “guarantee” comes from the Italian term *garanza*, referring to genuine madder dyestuff. Where instructions for

selecting raw materials, preparing paints, and the correct methods of using them were recorded, it was usually in the form of secret recipe books known as formularies. One of the most informative of these is *De diversis artibus* (“On diverse arts”) by the monk Theophilus (Roger of Helmarshausen), who also describes a formula for preparing artificial vermilion and whose *Schedula diversarum artium* also describes procedures for illuminating manuscripts.

An unusually complete formulary that has survived from the early Renaissance was compiled about 1400 by the Tuscan artist Cennini Cennini. Whereas Leon Battista Alberti’s treatise on painting (1435) is largely theoretical, and refers only briefly to color mixing, Cennini’s contains over 150 short chapters covering numerous aspects of workshop practice. It describes, for example, a relatively new yellow made from lead and tin oxides heated to 800°. Cennini calls it *giallorino* and describes it as heavy and permanent and useful for painting foliage and grass. In English, it became known as massicot (and subsequently lead-tin yellow) and was sometimes glazed with Indigo for depicting foliage.

Medieval artists also learnt that palettes could be extended when raw earths were roasted or calcined. Hence, yellow ochre was transformed into a reddish pigment variously called burnt ochre, light red, or English red; raw sienna and raw umber became burnt sienna and burnt umber, the latter useful for flesh tints. During the Middle Ages, a typical workshop palette might therefore include white lead, kaolin, gypsum, massicot, orpiment, yellow ochre, burnt ochre, red lead, raw sienna, burnt sienna, raw umber, burnt umber, red ochre and red bole, cinnabar or vermilion, madder lake, brazilwood lake, green earth, sap green, verdigris, malachite green, azurite blue, flower of woad, genuine ultramarine, lamp black, bone black, and vine black.

Renaissance and Beyond

Though Hubert and Jan van Eyck were not the inventors of oil painting, they are acknowledged (e.g., by Filarete) as perfecting its use in pictorial

painting, together with Rogier van der Weyden. Antonello da Messina is often credited with introducing the medium into Italy, and by the end of the 1400s, oil painting was commonplace throughout Europe. Magnificent frescoes continued to be painted throughout the 1500s, notably by Raphael, Michaelangelo, and Annibale Carracci, but working from a handheld palette distributed with oil colors became almost universal practice from the mid-sixteenth to the mid-twentieth century, when North American abstract artists in particular began making large-scale abstract works. At the same time, commercial paints, alkyds, and acrylic paints stimulated new and bolder methods of working. Notable exceptions were the various schools of watercolor painters, who devised small, portable, and often ingenious palettes for outdoor sketching and made of either wood or lightweight metal.

Baroque and Classicism

With the decline in large commissions from the church, courts, or guilds, and the rise of the merchant class, it became more difficult for aspiring artists to become pupils and apprentices in large-scale workshops. To compensate for this, from the late 1600s onward, a series of books began to appear that offered tuition in the basic skills of assembling and applying practical palettes. Among the first was William Salmon's *Polygraphice* (1672), followed by Claude Boutet's *Traité de la mignature* (1673), modified through dozens of editions. Alternatively, a student might learn from a single master but had better chance of a more rounded education if he could enroll at one of the many new academies, though even in the 1700s, these were few and far between. Italian painting continued to be much admired, and writers often bemoaned the fact that artists like Titian had not passed their "secrets" of color-mixing down to later generations, though Lomazzo's *Trattato dell'arte de la pittura* (1584) had preserved some of the late-Renaissance practices.

The glazing of three primary colors, demonstrated in color circles, was included in Moses

Harris' *Natural System of Colours* (1766), offering a new concept of tricolor mixing, and extended in Johann Lambert's *Farbenpyramide* (1772) and James Sowerby's *New Elucidation of Colours* (1809) and culminating in Philipp Otto Runge's three-dimensional *Farben-Kugel* (1810). As the science of chemistry accelerated throughout the early nineteenth century, many new artificial pigments were added to the artists' palette, notably cobalt blue (1804), chrome yellow (1818), cadmium red and yellow (1817), Guimet's artificial ultramarine (1826), and chromium oxide green (1838).

Nineteenth Century

In excess of 50 instructional books on how to use color palettes were published during the 1800s, though few influenced the development of fine art painting as such. One exception is Ogden Rood's *Modern Chromatics* (1879), responsible (in its French edition of 1881) for clarifying Chevreul's earlier explanations of optical color mixing. By the mid-1880s, numerous artists (including Georges Seurat, Vincent van Gogh, and Camille Pissarro) were laying out their colors in spectral sequence, with white available for tinting.

Twentieth Century

Over the next 30 years, almost every progressive painter experimented with some form of pointillism, and its novel way of applying unmixed color directly from the palette to the canvas (notably in early twentieth-century French and German expressionism) led to the liberation of color as a pictorial element in its own right, and one no longer restricted to the traditional task of describing forms in the natural or built environment.

During the early years of the twentieth century, avant-garde artists became less concerned with the subtleties of matching their colors to external objects as with the relationships between the colors arrayed on the palette. Such a preoccupation was reinforced first by the publication of Albert Munsell's *A Color Notation* (1905) and

then of Wilhelm Ostwald's *Die Farbenfibel* (1916), which was highly publicized throughout European schools of art and design. One limitation of such systems, and others like them, was their emphasis on opaque colors, whereas many artists (including Paul Klee) typically experimented with translucent oil or watercolor glazes. Additionally, as if in reaction to the color explosion of the Impressionist generation, the Cubists reverted to the ancient tetrachrome palette, with Georges Braque, Pablo Picasso, and even Marcel Duchamp, restricting their palettes to white, black, red and yellow ochre, and subdued green earth. No theorists or writers of art manuals influenced their decision, yet a general reversion to classicism held sway until the mid-1950s, through various realist revivals, until vivid palettes came to the fore again in the 1960s, with the extravert and confident art movements variously labeled Pop Art, Op Art, and ► [Color Field Painting](#).

Cross-References

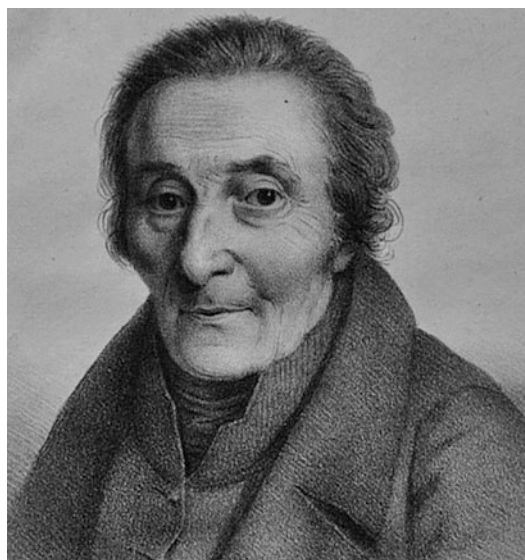
- [Ancient Color Categories](#)
- [Chevreul, Michel-Eugène](#)
- [Color Field Painting](#)
- [Colorant, Natural](#)
- [Impressionism](#)
- [Op Art](#)
- [Pigment, Inorganic](#)

References

1. Letalle, A.: *Les palettes d'artistes*. E. Sansot, Paris (1912)
2. Speed, H.: *The Science and Practice of Oil Painting*. Chapman and Hall, London (1924)
3. Schmid, F.: *The Practice of Painting*. Faber & Faber, London (1948)
4. Bazzi, M.: *Abecedario pittorico*. Longanesi, Milan (1956). Translated as *The Artists' Methods and Materials*. John Murray, London (1960)
5. Birren, F.: *History of Color in Painting*. Van Nostrand Reinhold, New York (1965)

Palmer, George

Rolf G. Kuehni
Charlotte, NC, USA



Louis Fehr: George Palmer, ca. 1820

Biography

George Palmer, also known as George Giros de Gentilly named Palmer, was an English dye chemist, color theorist, inventor, and soldier. According to his obituary, Palmer was born ca. 1746 on a ship, to English Catholic parents. Due to the eighteenth-century restrictions on activities of English Catholics, Palmer lived a double life between England and France. Nothing is known about his early years. Circa 1775, he introduced a solution of tin as a new mordant for the dyeing of wool fabrics in Louviers, France, using the last name Giros de Gentilly [1, 2]. In 1777, located in London, he published the book *Theory of Colours and Vision*, a French edition of which was published in the same year in Paris, translated by Palmer's friend Denis-Bernard Quatremère d'Isjonval, at the time active in the textile

manufacturing facility Disjonval in Sedan, France, owned by his family [3, 4]. In the same year, Palmer also invented a fawn-colored dye in London [5]. In 1781, J. H. Voigt of Gotha, Germany, editor of *Magazin für das Neueste aus der Physik und Naturgeschichte* (Journal for the Latest Physics and Natural Sciences News), describes meeting with Giros von Gentilly and the latter's conjectures about color blindness [6]. In 1785, Palmer, living in Paris, had *Lettre sur les moyens de produire, la nuit, une lumière pareille à celle du jour* (Letter Concerning the Means of Producing at Night a Light Equal to Daylight) published [7] describing the modification of oil lamplight with a blue glass mantel, a technology that became fashionable for a time. In 1786, Palmer published *Théorie de la lumière, applicable aux arts, et principalement à la peinture* (Theory of Light Applicable to the Arts, Principally to Painting) [8]. Toward the end of that decade, likely as a result of the French Revolution, Palmer became a mercenary soldier in the Corps of Engineers, at different times for Sweden, Austria, and Russia, reaching the rank of major, as described in his obituary [9]. For a time in the early nineteenth century, he lived near Leipzig in Germany where he reported on four technical inventions, one of which being a fire-extinguishing powder, a demonstration of which was reported in a local newspaper. In 1811, Palmer moved to Copenhagen into retirement and died there destitute in 1826 [10].

Major Accomplishments/Contributions

Palmer made lasting contributions to the development of color science by being the first to propose that there are three different mechanisms in the human eye that account for color vision: "The superficies of the retina is compounded of particles [light sensors] of three different kinds, analogous to the three rays of light; and each of these particles is moved by its own ray" [4]. This statement has proved true in regard to the number of different daylight sensor types, the cones, in the human eye, if not in regard to the claim of three kinds of light. Thirty-five years later, a similar

statement was made by the eminent physicist Thomas Young [11].

Voigt, in his report on Giros von Gentilly, describes him as having stated that color blindness arises if one or two of the three kinds of "particles" in the retina are inactive, a statement found to be valid [6]. In 1786, Palmer provided a hypothesis for the complementary nature of the successive contrast effect by stating that it is due to fatiguing of one or two of the light sensor types, an explanation that continues to be accepted as valid, as does his conjecture that the different kinds of sensor take different times to recover upon exposure to strong light [8].

References

1. Berthollet, C.L., Berthollet, A.B.: Elements de l'art de la teinture, 2nd edn. Firmin, Didot, Paris (1804)
2. Macquer, P. J.: Mémoire des Sieurs Maille frères et de Lafosse manufacturiers de drap a Louviers, Aug. 3, 1781. Document No. 5315, Box F12/1334B, Archives Nationales de France, Paris (1781)
3. Palmer, G.: Theory of Colours and Vision. Leacroft, London (1777)
4. Palmer, G.: Théorie du couleur et de la vision (trans. Quatremère d'Isjonval, D.-B.). Pissot, Paris (1777)
5. Palmer, G.: Letter dated June 19, 1793, Letter from Lord Hervey including two letters in French and two samples by G. Palmer about his yellow dye, Royal Society for the Encouragement of Arts, Trades, and Manufactures, London, Ref. No.: PR-MC/105/10/452 (1793)
6. Voigt, J. H.: Des Herrn Giros von Gentilly Muthmassungen ueber die Gesichtsfehler bey Untersuchung der Farben. Magazin für das Neueste aus der Physik und Naturgeschichte 1(2. St.), 57–61 (1781)
7. Palmer, G.: Lettre sur les moyens de produire, la nuit, une lumière pareille à celle du jour. Paris (1785)
8. Palmer, G.: Théorie de la lumière, applicable aux arts, et principalement à la peinture. Hardouin et Gattey, Quinquet, Paris (1786)
9. Thaarup, F.: Dagen newspaper, Copenhagen, No. 66 (March 18), No. 68 (March 21) and No. 70 (March 23) (1826)
10. Kuehni, R.G.: Which George Palmer? Color Res. Appl. 34, 5–9 (2009)
11. Young, T.: The Bakerian Lecture, On the theory of light and colours. Philos. Trans. R. Soc. Lond. 92, 12–48 (1802)

Parafovea

Masato Sakurai
Kanazawa Institute of Technology, Nonoichi,
Ishikawa, Japan

Synonyms

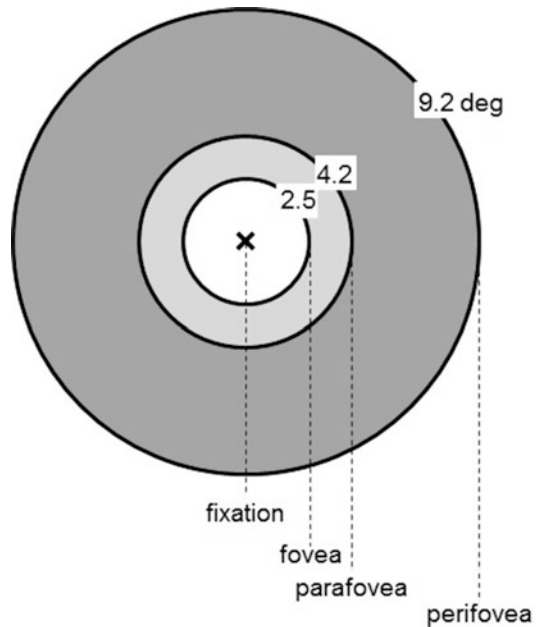
Parafoveal vision

Definition

The parafovea may refer to a position at a visual angle of approximately 4° from the normal fixation point or to the region extending to that angle from the fovea. It is compositionally distinguished from the fovea by its lower cone and ganglion cell densities and by its higher rod density. As may be expected, these differences affect its visual perception characteristics. Its spatial contrast sensitivity, visual acuity, and other visual attributes are lower than those of the fovea, but, as shown by its critical fusion frequency (CFF), its temporal frequency resolution extends to higher frequencies than that of the fovea. Its color perception is essentially the same as that of the fovea, but under low luminance, it becomes similar to tritanopic diachronic perception. These characteristics can be inferred from the relative distribution of the three cone types in the parafoveal retina. In reading, moreover, parafoveal information contributes to line of vision movement and to fine foveal information processing.

Overview

The retinal structure in the central visual field comprises three distinctive regions. As shown schematically in Fig. 1, the parafovea lies between the fovea and the perifovea. The fovea lies in the central region extending from the fixation point to a visual angle of 2.5° radial, the parafovea surrounds it and extends out to 4.2° radial, and the



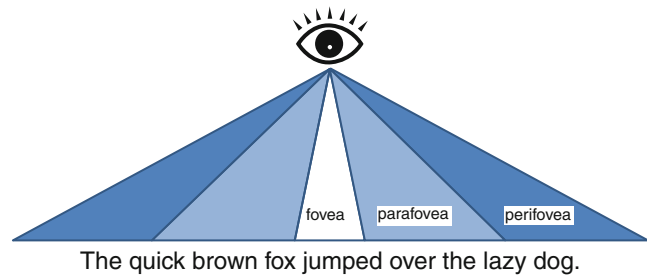
Parafovea, Fig. 1 Central visual field on monocular viewing [1, 2]

perifovea extends from there to 9.2° radial [1, 2]. These demarcations are not strictly defined, but phenomena occurring approximately 4° radial from the normal fixation point are generally referred to as parafoveal.

In the parafovea, the cone density declines from the high concentration of cones in the fovea, the rod density rises from the complete absence of rods in the fovea [3], and the density of ganglion cells, which form the retinal outlet of the optic nerve, declines together with the ganglion cell layer thickness from the high density and thickness of the foveal region [4, 5]. Although the parafovea and the fovea both form part of the macula, they differ significantly in macular pigment density [6, 7]. These structural and compositional differences suggest that they affect parafoveal visual perception.

Spatial contrast sensitivity, a fundamental characteristic of visual perception, is lower in the parafovea than in the fovea. The parafoveal contrast at an eccentricity of 4° is functionally characterized as a spatial frequency band-pass filter, like that of the fovea at 0° , but with lower overall

Parafovea, Fig. 2 Foveal, parafoveal, and perifoveal regions in reading when three characters make up 1° of visual angle [15]



sensitivity [8, 9]. In the region most sensitive to spatial frequencies of 2–8 cycles/°, the parafovea sensitivity is approximately 1/2–1/3 of the foveal sensitivity [9]. Visual acuity is also significantly lower than in the fovea. In Landolt ring visual acuity measurements, subjects with a foveal visual acuity of 2.0 exhibit a sharply lower visual acuity of 0.3 in the parafoveal region at approximately 4° [10], thus showing that a larger character size is necessary for character deciphering. It has been found that the character size threshold for deciphering increases linearly throughout the foveal region and on out to an eccentricity of 30° [11]. In temporal frequency contrast sensitivity, on the other hand, the parafoveal resolution extends to a higher frequency than the fovea, as measured by CFF (critical fusion frequency) [12].

Parafoveal color perception does not depart from that of the fovea under relatively high illuminance, but under low illuminance, it resembles tritanopic dichromatic perception (small-field foveal vision) [13]. This may be inferred from psychophysical measurements that show nearly the same L- and M-cone ratios in the fovea and the parafovea [14] but extremely low S-cone distribution in the parafovea as compared with its L- and M-cone distributions [6].

Reading experiments on parafoveal information processing have shown that efficient reading requires simultaneous parallel processing of parafoveal and foveal information [15]. Figure 2 shows the foveal, parafoveal, and perifoveal regions in reading with three characters occupying a visual angle of 1° [15]. The foveal region, which involves processing in finer detail, is narrow. It occupies only approximately seven characters, but it receives parafoveal information concerning the position of the subsequent

characters, which results in a shift in the line of sight toward them by movement of the eyeball. This has been confirmed experimentally by having subjects gaze at the fixation point and then showing them a word in the parafoveal region, which immediately elicits a response in which the subjects move the line of sight and read that word.

Three phenomena that have been observed in reading experiments concerning parafoveal information processing are known as skipping, parafoveal-on-foveal (PoF), and preview-benefit effects. The skipping effect consists of skipping every third word or so, which would presumably not be possible without identification of the word via the parafovea. The skipped words are generally short terms such as “to,” “for,” “in,” and “the,” which supports the view that a word’s lexical status contributes to the determination of whether it will be skipped.

In the PoF effect, the occurrence of a parafoveal word that is difficult to understand affects the foveal fixation and thus the field of vision. As noted above, parafoveal processing affects foveal processing, and a low-level PoF effect has been well established in experiments involving the occurrence of unfamiliar character combinations in the parafoveal region. The preview-benefit effect has been observed in experiments in which the word on the parafovea that is subsequently to be subjected to foveal processing is replaced by a related word during the saccade. If the replacement word is related to the original parafoveal word, the fixation on that word by the fovea is shorter and its processing thus faster than if the replacement word is unrelated. Recent studies have indicated that the preview-benefit effect occurs in sentences independently of the initial

sentence word, mid-sentence words, and final sentence word.

These three effects are currently the subject of intense discussion and further findings may be expected.

Cross-References

- [Chromatic Contrast Sensitivity](#)
- [Color Vision, Opponent Theory](#)
- [Cone Fundamentals](#)
- [Tritanopia](#)

References

1. Polyak, S.L.: The Retina. University of Chicago Press, Chicago (1941)
2. Osaka, N.: Psychophysical Analysis of Peripheral Vision. Kazama Shobo Press, Tokyo (1983). (in Japanese)
3. Wandell, B.A.: Foundations of Vision. Sinauer Associates, Sunderland MA, US (1995)
4. Curcio, C.A., Allen, K.A.: Topography of ganglion cells in human retina. *J. Comp. Neurol.* **300**, 5–25 (1990)
5. Iwasaki, M., Inomata, H.: Relationship between superficial capillaries and foveal structures in the human retina. *Invest. Ophthalmol. Vis. Sci.* **27**, 1698–1705 (1986)
6. Gegenfurtner, K.R., Sharpe, L.T. (eds.): Color Vision from Genes to Perception. Cambridge University Press, Cambridge, UK (1999)
7. Werner, J.S., Bieber, M.L., Scheffrin, B.E.: Senescence of foveal and parafoveal cone sensitivities and their relations to macular pigment density. *J. Opt. Soc. Am. A* **17**(11), 1918–1932 (2000)
8. Rovamo, J., Virsu, V., Nasanen, R.: Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature* **271**, 54–56 (1978)
9. Virsu, V., Rovamo, J., Laurinen, P.: Temporal contrast sensitivity and cortical magnification. *Vision Res.* **22**, 1211–1217 (1982)
10. Virsu, V., Nasanen, R., Osmoviita, K.: Cortical magnification and peripheral vision. *J. Opt. Soc. Am. A* **4**(8), 1568–1578 (1987)
11. Anstis, S.M.: A chart demonstrating variations in acuity with retinal position. *Vision Res.* **14**, 589–592 (1974)
12. Hartmann, E., Lachenmayr, B., Brettel, H.: The peripheral critical flicker frequency. *Vision Res.* **19**, 1019–1023 (1979)
13. Gilbert, M.: Colour perception in parafoveal vision. *Proc. Phys. Soc.* **63**, 83–89 (1950)
14. Nerger, J.L., Cicerone, C.M.: The ratio of L cones to M cones in the human parafoveal retina. *Vision Res.* **32**, 879–888 (1992)
15. Schotter, E.R., Angele, B., Rayner, K.: Parafoveal processing in reading. *Atten. Percept. Psychophys.* **74**, 5–35 (2012)

Parafoveal Vision

- [Parafovea](#)

Parallel Visual Pathways

- [Magno-, Parvo-, Koniocellular Pathways](#)

Pastel Colors

Malvina Arrarte-Grau
Architecture, Landscape and Color Design,
Arquitectura Paisajismo Color, Lima, Peru

Synonyms

[Pale colors](#); [Pastel crayons](#); [Pastel paintings](#); [Pastel sticks](#); [Pastel tones](#); [Soft colors](#)

Definition

In painting the term pastel comes from the French *pastille* (Delamare, 1999, p.117), meaning small loaf or lozenge, in relation to the characteristics of the stick used as a coloring medium. The noun “pastel” also denotes the artwork produced with pastel crayons. As an adjective, “pastel tones” refer to light, pale, clean colors, as those produced with pastel crayons.

The Pastel Crayon

A pastel crayon is a compact stick used as a dry coloring medium, which combines a pure powdered pigment and a neutral binding for cohesion and plasticity. A calcium carbonate component, such as chalk, gypsum, or pumice, is used in the preparation of the stick, resulting in various degrees of hardness and in gradations from the purest to the highly luminous. Hard pastels contain a higher proportion of binder, are sharper, and work for drawings and sketches. Soft pastels

usually come wrapped in paper, for they are drier and more fragile.

Pastel impressions served as the first measurement standards in colorimetry (Delamare, 1999, p.117) for their purity and similarity in color to those of the natural pigments.

The Pastel Technique

The pastel technique consists of transferring small granules of color from the crayon to the support, usually textured or abrasive for good adherence. A fixative is required for protection after its application. In pastel paintings surfaces result extremely opaque as high concentrations of pigment allow no refraction, but the effects of pastels may also be delicate and vaporous, as in sketches and drawings.

The Origin of the Pastel

It is said that pastels originated in France. Leonardo da Vinci refers to them in his *Codice Atlantico* (1495) as “a new dry substance” (Geddo C., 2008, pp. 63–87). The Venetian painter Rosalba Carriera and the French Quentin de La Tour represent the development of the technique, mainly used for portraits during the eighteenth century, with fast realistic effects. The immediate results of the pastel, appropriate for movement and landscape paintings, were the preferred art form of the impressionist Maurice Degas. The American Marie Cassat, influenced by Degas, created many of her most important works in this medium. Later, pastel tones became popular in Modern Art for their bright and luminous colors.

Pastel Colors, the Pastel Palette, and Its Application

Pastel colors refer to pale clean luminous nuances. The colors in a pastel ► [palette](#), different from the colors in a pastel crayon set, have whiteness as a common denominator. This particularity results from low chromaticness and a minimum or no content of black. In subtractive mixing pastel tones would be obtained by adding white to a base color.

The pastel ► [palette](#) evokes spring, summer, and Easter and has been traditionally used in

infant decoration for its softness and gaiety. Pastel tones are also associated to the soft and velvety textures of pastel paintings, as in wool and cotton, sugar icing, and powdered eye makeup. During the 1980s pastel tones became a trend in architecture, interior design, and product design, ranging from men’s ties to motorcycles.

References

1. Delamare F. and B. Guineau. 1999. *Les Matériaux de la Couleur*. France. Gallimard
2. Geddo C. 2008–9. Il “pastello” ritrovato: un nuovo ritratto di Leonardo? *Italian in Artes*, no. 14. <http://www.lumiere-technology.com/Artes14.pdf>. Accessed: August 24, 2015
3. <https://en.wikipedia.org/wiki/Pastel>. Accessed: August 24, 2015

Pastel Crayons

► [Pastel Colors](#)

Pastel Paintings

► [Pastel Colors](#)

Pastel Sticks

► [Pastel Colors](#)

Pastel Tones

► [Pastel Colors](#)

Pattern-Induced Flicker Colors

► [Fechner’s Colors and Behnam’s Top](#)

Pearlescence

- [Iridescence \(Goniochromism\)](#)

Perceived Reflectance

- [Anchoring Theory of Lightness](#)

Perceived Value

- [Anchoring Theory of Lightness](#)

Perception by Night

- [Automotive Lighting](#)

Perceptual Appearance

- [Appearance](#)

Perceptual Color Effects

- [Psychological Color Effects](#)

Perceptual Grouping and Color

Daniele Zavagno¹ and Olga Daneyko²

¹Department of Psychology, University of Milano-Bicocca, Milan, Italy

²Department of Neuroscience, University of Parma, Parma, Italy

Synonyms

[Gestalt grouping principles \(or laws\)](#)

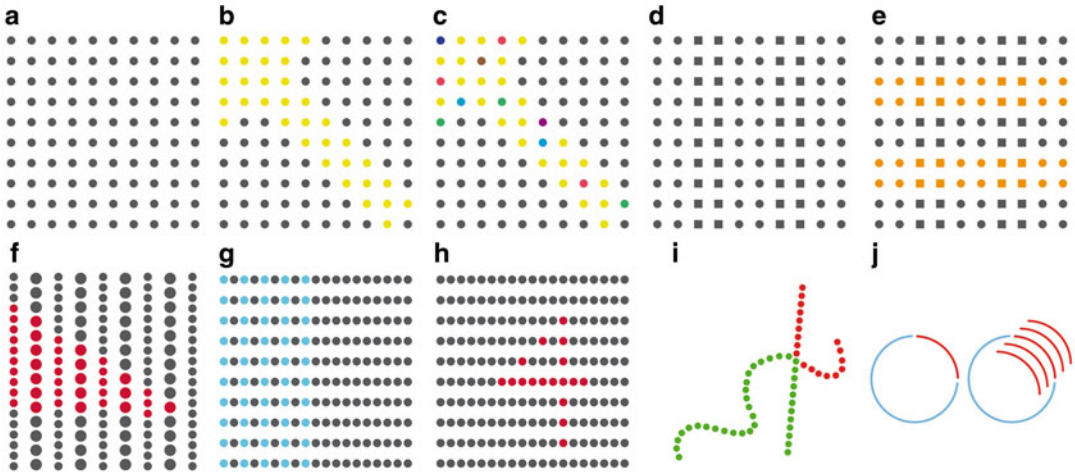
Definition

Perceptual grouping refers to principles by means of which a set of discrete elements are partitioned into groups by the visual system, thus forming higher-order perceptual units. Perceptual grouping was first studied by Max Wertheimer, one of the fathers of Gestalt psychology [1]. It is a core topic within the studies of perceptual organization, along with figure-ground segmentation [2]. As regards to color, one can consider either the role of color in perceptual grouping or the effects of perceptual grouping on the perception of color. In the first case the focus is on how color can aid and affect the perceptual organization of the visual scene. In the second case the focus is on how the color appearance of a given surface is affected by the color of other surfaces with which it is perceptually grouped.

Overview

Perceptual grouping is, along with figure-ground segmentation, a core topic in the studies of perceptual organization, i.e., of those processes that structure the sensory input into coherent units (visual objects, entities, and events), thus contributing significantly to the layout of the visual scene. From such a definition, it is clear that *grouping* and *segmentation* are two faces of the same coin, as originally postulated by the fathers of Gestalt psychology Max Wertheimer, Wolfgang Köhler [3], and Kurt Koffka [4].

The main factors that regulate perceptual grouping are the following: *proximity*, *similarity*, *common fate* (similarity in motion or change in time), *good continuation*, *closure*, *prägnanz* (or *good gestalt* or *structural coherence*), and *past experience*. These factors are also known as gestalt “principles” or “laws.” Color comes into play with the factor *similarity*. Figure 1a shows a regular square lattice of dots, and one can group the dots in many ways to form subpatterns, yet none of such patterns is specific to the lattice. By selectively introducing a color difference between the dots, the visual system can be forced to group the dots into patterns or to form larger units based



Perceptual Grouping and Color, Fig. 1 Grouping by color similarity. (a) Gray dots organized in a regular square lattice within which no other group or visual structure emerges. (b) By selectively introducing a change of color in some of the dots, one can generate specific patterns; (c) notice that the strength of the pattern is directly influenced by the color similarity of the dots: in this lattice 30 % of the dots that form the arrow are made of colors other than yellow, and the impression of an arrow-like structure is weakened. (d) This lattice is made of round and square dots; because of shape similarity, the lattice appears vertically organized in pairs of columns; (e) color similarity can overrun shape similarity, switching the perceptual organization from columns to rows. (f) This pattern would normally appear as dots organized in columns according to the factors size similarity and proximity; by introducing color, another pattern within the pattern appears. (g) Because of

proximity, the configuration normally appears organized in horizontal rows (right side); color similarity however overruns the factor proximity forcing a perceptual organization in vertical columns (left side). (h) Number 4 appears in spite of proximity and good continuation; notice however that in this configuration and the one in (f), the red patterns can be seen as if they were under the lattice, an impression that recalls transparency. (i) Color similarity does not disrupt good continuation; instead, it generates a second-order group because of which the intersection between the curvy line and the straight line becomes a salient feature of the configuration. (j) On the left, the factor closure determines the impression of a circle with a red arch; on the right, the red arch is grouped with other red ones, masking the completion effect. Notice however that in this last case, the light blue circle appears to complete amodally under the structure formed by the red arches

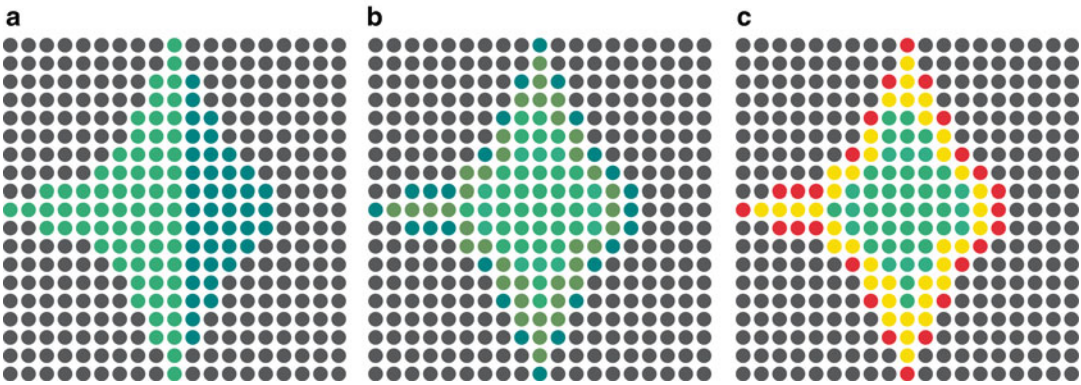
on color similarity. Figure 1b, for instance, shows a yellow arrow; notice however how the grouping effect weakens if the dots that make the arrow in Fig. 1b show instead multiple colors as in Fig. 1c. Color similarity is a relatively strong factor that can overrun other types of similarities, such as shape (Fig. 1d–e) and size (Fig. 1f). It can also overrule a relevant factor such as proximity (Fig. 1g–h); however, it cannot easily overrun good continuation (Fig. 1i). Finally, while color differences among elements do not seem to affect closure (Fig. 1j left), grouping by color similarity can be employed to mask closure (Fig. 1j right).

As a grouping factor, color similarity is not “all or none” (Fig. 2): this makes color similarity an extremely ductile tool in art, graphic design, cognitive ergonomics, etc., as it can be used to

enhance, impose, or create visual structure or to hinder or mask the perception of structure and form, such as in animal mimicry or in military camouflage.

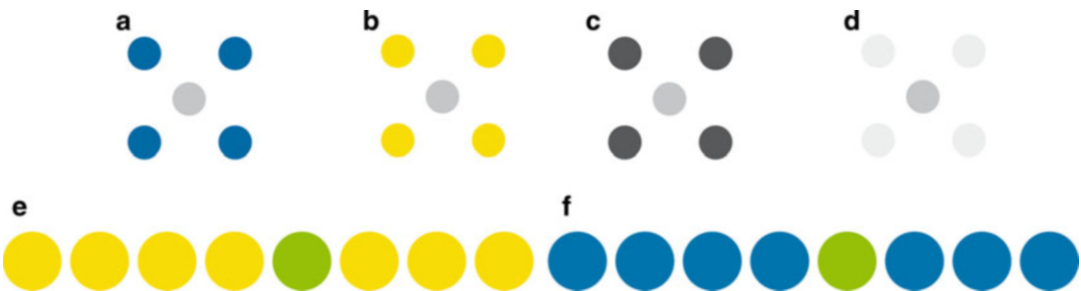
The Effects of Perceptual Grouping on Color

More problematic is the account of the effects of perceptual grouping on color. In an early paper, Fuchs [5] described situations in which perceptual grouping leads to color assimilation. Though sometimes reported by different authors (see, for instance, [6, 7]), older books dedicated to color perception [8, 9] curiously do not make mention to Fuchs’ findings on color assimilation and



Perceptual Grouping and Color, Fig. 2 Grouping by color similarity is not all or none: in (a), the *light green* and the *dark green* dots are grouped separately in substructures that are then combined together to form the impression of a global ray fish-like structure. In (b), dots are of three different shades of *green*, loosely organized with brighter greens at the center and darker greens in the periphery of the ray fish-like structure. This type of grouping is commonly observed in pictorial art and can be easily appreciated in the art movement known as Pointillism. The case

represented in (c) is somewhat similar to the one shown in Fig. 1c; however, here, the ray fish-like structure emerges more easily because of the limited number of colors used and the way they are ordered, with the *green dots* grouped in the center, the *yellow* ones surrounding the *green dots*, and the *red dots* delimiting the periphery. This kind of grouping can be exploited to show energy patterns (for instance, patterns of neural activation, intensity distributions, etc.)



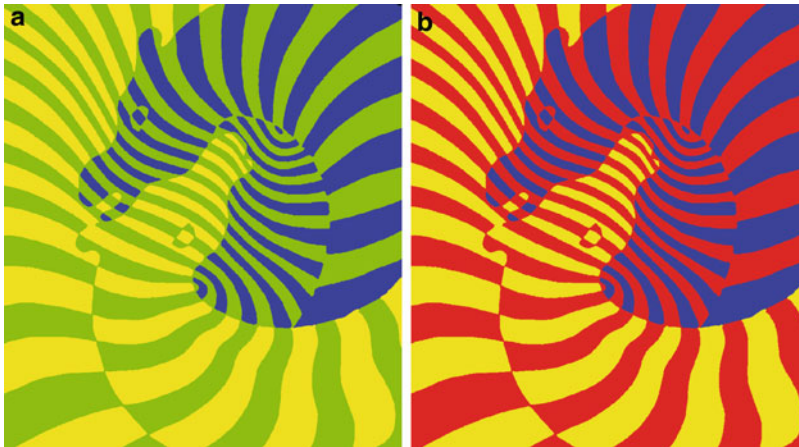
Perceptual Grouping and Color, Fig. 3 (a–b) A graphic interpretation of one of Fuchs’ displays: the central *gray dot* should appear bluish when surrounded by *blue dots* and yellowish when surrounded by *yellow dots*. (c–d) A typical lightness contrast effect due to perceptual grouping and accountable in terms of the anchoring

theory for lightness perception [6]. (e–f) Two *straight lines* of dots (proximity, good continuation), one *yellow* and the other *blue*, both including a *green dot*. Similar configurations can give rise to either assimilation or contrast

grouping. The issue may depend on the fact that Fuchs’ demonstrations do not always work on paper (or screen), maybe because they were mostly created by using light, shadows, projections, and reflections.

An attempt to graphically replicate one of Fuchs displays is made in Fig. 3a, b, in which two identical gray dots are placed one at the center of a square made of four blue dots and the other at the center of a square made of four yellow dots. According to Fuchs, the gray dot surrounded by

blue dots should appear bluish, while the gray dot surrounded by yellow ones should appear yellowish. The impression however, though rather weak, seems to go in an opposite direction: that of contrast, at least as far as lightness (achromatic surface color) is concerned. Fig. 3e, f employs proximity and good continuation to create two rows of respectively yellow and blue dots, which both incorporate an identical green dot: for those who will get an assimilation effect, the green dot grouped with the yellow dots will look somewhat



Perceptual Grouping and Color, Fig. 4 Colored zebras chromatically modified from one of Vasarely's graphic artworks. (a) The *green stripes* appear yellowish on the yellow-striped zebra, while they appear *bluish* on the blue-striped zebra. (b) The *red stripes* appear with an *orange* tinge on the yellow-striped zebra and slightly

purplish on the *blue-striped* zebra. Moreover, the assimilation effect works both ways: the *yellow* and *blue stripes* in (a) appear with a green tinge, while the physically identical *yellow* and *blue stripes* in (b) appear to be affected by the *red stripes*

more yellowish than the green dot grouped with the blue ones. It is not uncommon, however, to have observers reporting an opposite finding or the same observer reporting switches from assimilation to contrast and vice versa.

While the perceptual outcomes emerging from the observation of Fig. 3e, f are not clear, those deriving from Fig. 3a, b recall the effects of perceptual grouping on lightness: a middle gray target that is grouped with dark gray surfaces will look lighter than a middle gray target that is grouped with light gray surfaces (Fig. 3c, d) [10].

Future Directions

Was Fuchs completely wrong when he wrote about color assimilation as an effect of grouping elements into meaningful wholes? As mentioned earlier, his experimental setups were quite different from those commonly used nowadays; hence, before any conclusion can be drawn, his experimental setups need to be replicated. Moreover, Fig. 4 may be considered in support of Fuchs' claims, who insists on the idea of assimilation as a reduction of differences among the parts grouped to form a gestalten. In Fig. 4 the two

configurations made of zebra-like structures, derived from a modification of a well-known graphic artwork by Victor Vasarely, show strong effects of color assimilation in which not only the "critical" color (to use Fuchs terminology – in Fig. 4a green stripes, in Fig. 4b red ones) assimilates chromatic qualities of the stripes with which they are grouped but also the blue and yellow stripes pickup characteristics of the critical colors. For instance, in Fig. 4a the yellow stripes have a greenish tinge, while in Fig. 4b they show a slight orange tinge.

Cross-References

- [Anchoring Theory of Lightness](#)
- [Assimilation](#)
- [Color Constancy](#)

References

1. Wertheimer, M.: Untersuchungen zur Lehre von der Gestalt. II. (Investigations on the theory of shape). *Psych Forschung* **4**, 301–350 (1923)
2. Palmer, E.S.: *Vision Science. Photons to Phenomenology*. The MIT Press, Cambridge, MA (1999)

3. Köhler, W.: *Gestalt Psychology*. Liveright, New York (1929)
4. Koffka, K.: *Principles of Gestalt Psychology*. Harcourt, Brace and Company, New York (1935)
5. Fuchs, W.: Experimentelle Untersuchungen über die Änderung von Farben unter dem Einfluss von Gestalten (Experimental investigations on the alteration of color under the influence of Gestalten). *Zeitschrift für Psychologie* **92**, 249–325.
6. Metzger, W.: *Gesetze des Sehens* (The laws of seeing). Verlag Waldermar Kramer, Frankfurt am Main (1975)
7. Van Lier, R., Wagemans, J.: Perceptual grouping measured by color assimilation: regularity versus proximity. *Acta Psychol.* **97**, 37–70 (1997)
8. Katz, D.: *The World of Colour*. Kegan Paul, Trench, Trubner, London (1935)
9. Beck, J.: *Surface Color Perception*. Cornell University Press, Ithaca/London (1972)
10. Gilchrist, A.: *Seeing Black and White*. Oxford University Press, New York (2006)

Philosophy of Color

Barry Maund

Department of Philosophy, The University of Western Australia, Crawley, WA, Australia

Definition

The philosophy of any live topic is a matter of puzzles and problems and debates about their solution. Philosophical debates about color go back to the ancient Greeks, e.g., to Plato, Aristotle, the early atomist Democritus, and many others, and they are still alive. One of the most important reasons why colors are of philosophical interest is that they raise serious metaphysical issues, concerning the nature both of physical reality and of the mind. Central among these issues are questions concerning whether color is part of a mind-independent reality or not and what account can be given of experiences of color.

The Scientific Tradition

Serious thinking about color raises important philosophical puzzles and questions. One of the major

problems with color has to do with fitting what people seem to know about colors into what science, particularly physics, tells about physical bodies and their qualities. It is this problem that historically has led the major physicists, who have thought about color, to hold the view that physical objects do not actually have the colors people ordinarily and naturally take objects to possess. Oceans and skies are not blue in the way that people naively think, nor are apples red (nor green). Colors of that kind, it would seem, have no place in the physical account of the world that has developed from the sixteenth century to the twenty-first century.

Not only does the scientific mainstream tradition conflict with the common-sense understanding of color in this way, but as well the scientific tradition presents a very counterintuitive conception of color. There is, to illustrate, the celebrated remark by the eighteenth-century philosopher David Hume: “Sounds, colors, heat and cold, according to modern philosophy are not qualities in objects, but perceptions in the mind” [1]. Physicists who have subscribed to this doctrine include the luminaries: Galileo, Boyle, Descartes, Newton, Young, Maxwell, and Helmholtz. Maxwell, for example, wrote: “It seems almost a truism to say that color is a sensation; and yet Young, by honestly recognizing this elementary truth, established the first consistent theory of color” [2].

This combination of color eliminativism – the view that physical objects do not have colors, at least in a crucial sense – and color subjectivism – the view that color is a subjective quality – is not merely of historical interest. It is held by many contemporary experts and authorities on color. S. E. Palmer, a leading psychologist and cognitive scientist, writes:

People universally believe that objects look colored because they are colored, just as we experience them. The sky looks blue because it is blue, grass looks green because it is green, and blood looks red because it is red. As surprising as it may seem, these beliefs are fundamentally mistaken. Neither objects nor lights are actually ‘colored’ in anything like the way we experience them. Rather, color is a psychological property of our visual experiences when we look at objects and lights, not a physical property of those objects or lights. The colors we see are based

on physical properties of objects and lights that cause us to see them as colored, to be sure, but these physical properties are different in important ways from the colors we perceive. [3]

The examples of authorities who say similar things can be multiplied, e.g., W. D. Wright, E. H. Land, S. Zeki, and R. G. Kuehni. It should be noted, in passing, that there is considerable variation in how the subjectivist component is expressed, with colors being variously described as sensations, psychological properties of visual experiences, mental properties, representations, constructions of the brain, and properties of the brain.

The Complexities of the Scientific Tradition

The traditional scientific thinking on color is more complex, however, than has been presented so far. Central figures such as Descartes, Boyle, and Locke think that there are no colors in the physical world – no colors, as people ordinarily and naively understand them to be. But they also speak of colors as *secondary qualities*, i.e., as powers or dispositions, in bodies, to cause experiences of a certain characteristic type. It is instructive to try to understand this dual position.

It was common for such theorists to make a distinction between colors as they are in physical bodies and colors as they are in sensation. And then, with respect to the first kind of color, they made a distinction between a philosophical way of thinking and the common, everyday way of thinking of such qualities (which Newton describes as “such conceptions as vulgar people . . . would be apt to frame”). As Descartes explains, the ordinary way involves the mistake of “judging that the feature of objects that we call ‘color’ is something ‘just like the color in our sensation’” or of believing that “the self-same whiteness or greenness which I perceive through my senses is present in the body.” The philosophical way of thinking, by contrast, is illustrated in his following remark: “It is clear then that when we say we perceive colors in objects, *it is really just the same as saying* that we perceived in objects something as to whose

nature we are ignorant but which produces in us a very clear and vivid sensation, what we call the sensation of color” [4].

So, there is a way – a philosophical way – of thinking of objects as having colors. However, three things are important: (1) it is different from the ordinary way of thinking; (2) it involves a revision, or a reconstruction, of people’s ordinary thinking about color; and (3) it presupposes a notion of “sensation of color” or “color as it is in sensation.” It is crucial how this last notion is understood.

It should also be noted, however, that such theorists as Descartes, Boyle, and Locke sometimes wrote, as well, of colors in physical bodies as those features (“textures”) which grounded the associated powers/dispositions to induce the relevant experiences (sensations). Some commentators think that, in so doing, these theorists were either inconsistent or muddled or found it difficult to make up their minds. However, if Descartes, Boyle, Locke, etc. are in the business of revising or reconstructing ordinary concepts, there is no particular reason why they should not introduce several new concepts to replace the single old one, each one serving a different purpose.

Indeed, it is instructive to look at the famous passage in which Newton writes of the rays not being colored:

For as Sound in a bell, or musical String or other sounding body, is nothing but a trembling motion, and in the Air, nothing but that Motion propagated from the Object, and in the Sensorium ‘tis a sense of that Motion under the form of sound; so Colors in the Object are nothing but a Disposition to reflect this or that sort of Rays more copiously than the rest; in the Rays they are nothing there but Dispositions to propagate this or that motion into the Sensorium, and in the Sensorium they are Sensations of those Motions under the Forms of Colors. [5]

Here, he writes of the colors in the object, colors in the rays, and colors in the sensorium, by analogy with sounds in the object, sounds in the air, and sounds in the sensorium.

There seems to be a similar spirit that inspires modern theorists such as Leo M. Hurvich and D. L. MacAdam. These are color scientists who have stressed the existence of a variety of senses

of color terms. Hurvich begins the second chapter of his *Color Vision* with the question “What is color?” which he turns into a series of questions: Is color something that inheres in objects themselves? Is it related to the light falling on an object? Is it a photochemical event that occurs in the receptor layer of the eye? Is it a neural brain-excitation process? Is it a psychological event? Hurvich’s answer is “Color is all these things” [6]. MacAdam writes:

The term color is commonly used in three distinctly different senses. The chemist employs it as a generic term for dyes, pigments and similar materials. The physicist, on the other hand, uses the term to refer to certain phenomena in the field of optics. . . . Physiologists and psychologists often employ the term in still another sense. They are interested primarily in understanding the nature of the visual process and use the term, on occasion, to denote sensation in the consciousness of a human observer. [7]

One might wonder what unites these various senses of the color terms. Why, that is, should color terms be used to apply to each of these types of conditions or properties – the same terms for each? A plausible answer to this question is the following. If one is interested in the detailed causal mechanisms that underlie color perception, then there is good reason to be interested in different stages of the causal process in which objects cause experience of color: chemists and painters would be interested in the pigments, lighting engineers and physicists would be interested in the light-modifying features of objects, physiologists and psychologists would be interested in the cones and rods in the retinas, other physiologists and neuroscientists would be interested in other neural processes and pathways, and psychologists and philosophers would be interested in the experience of color and in how colors appear. What unites them all is that the theorist is interested, at each stage, in that feature of the stage that contributes to the *perception* of color, i.e., to the way colors *appear*.

Philosophical Responses

Within the philosophical community, there has been a mixed response to the view expressed in

this scientific tradition. While some philosophers share key aspects of the view, at least in broad terms and with certain modifications, there have been many who have rejected the view. There has been strong resistance from many philosophers, both to the eliminativist tendency within the scientific tradition and to the related subjectivism.

One form this resistance takes reflects the fact that each component of this traditional view is very puzzling. One response is to say that color terms – red, blue, purple, orange, yellow, green, brown, etc. – are in order; there are paradigms of colors to which the color terms apply: ripe lemons are yellow, tomatoes and rubies are red, and so on. There is no trouble, by and large, in learning these terms and teaching them in ostensive practices to children and others. Moreover, colors play a central role in so many of human social practices, such as painting, art, fashion, decoration, ceremonial, enjoying nature, etc. In the second place, it is hard to make sense of the claim that colors are properties of sensations or are psychological properties; if they are anything, they are, surely, properties of objects and light sources – of peaches and emeralds, of skies, of rainbows, of glasses of wine, of headlamps, and so on.

This second issue is of key importance. Many color scientists are agreed in thinking that one role for color terms is in application to sensations. As well as the examples already given, it is possible to find, in authoritative texts, definitions like “Color attributes are *attributes of visual sensations*, e.g., hue, saturation and brightness”; “Hue: *attribute of color perception* denoted by the terms yellow, red, blue, green and so forth”; “Brightness is the *attribute of a visual sensation* according to which a given visual stimulus appears to be more or less intense”; etc. However, it is a serious philosophical question as to what is meant by “attributes of sensation” (and likewise “colors in sensation”). There are, for example, different ways of understanding what they are meant to be:

1. Qualities *of* the sensation
2. Qualities presented *in* sensation (or experience)
3. Qualities *represented by* the sensation

Each of these views has had, and continues to have, its champions. Some philosophers have argued that the first candidate makes no sense, that it commits something like a category mistake. It is interesting that one of the first to argue this way was not a philosopher, but the famous color scientist Ewald Hering [8]. Not surprisingly, perhaps, there are other philosophers who reject this charge. One way to do so is to argue that the perceiver is not aware of the qualities as *qualities of sensations*, even though that is what they are.

One issue that complicates the debates is that psychologists have become increasingly reluctant to use the term “sensation.” For example, modern psychological texts, even ones with titles such as *Sensation and Perception*, tend not to use the term “sensation,” opting for “perception” instead. Indeed, in many of the situations in which traditional color theorists would have used the term “sensation of color,” modern theorists would use such terms as “perceived color” and “apparent color.” Interestingly, it can be found, in Newton, examples of both usages. In the famous passage in which he declares that the rays are not colored, “in them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Color,” he begins by saying: “The homogeneous light and rays which *appear red*, or rather make objects *appear* so, I call red-making; those which make objects *appear* yellow, green, blue and violet, I call yellow-making, green-making, blue-making, violet-making, and similarly for the rest.”

Accordingly, instead of using the expressions “attribute of sensation” and “sensation of color,” it would make sense to use “apparent color” and “perceived color,” as many theorists currently do, or “color as we see it,” as some philosophers suggest. So, the debates can be framed in terms of the colors objects *appear to have* or *look to have*. The philosophical issues remain, however. The point is that it is a matter of philosophical dispute as to what exactly it is for something to *look blue*, say, (i.e., to *appear blue*). It is plausible that it involves the object causing an experience in the subject, but then the question can be raised with respect to the report, say, about one’s experience of blue, whether that report is describing:

1. A quality *of* the experience
2. A quality which is *presented in* the experience
3. A quality which the experience *represents* the object as having

Not only that, but, supposing that the experience does have representational content, there is a further question as to whether it has other features as well (ones that are open to introspective access). One view is that there are sensory qualities which are presented in sensory experience and which are “projected” onto physical objects. In this way, objects appear to have qualities that they do not actually have. Since this “projection” is not a literal one, it requires explanation of course, but its defenders think that it can be provided. On this theory, the sensory qualities presented in experience also form part of the representational content of the experience. This view is sometimes combined with *color fictionalism*, the view that ordinary color language is a language about properties which objects appear to have (but do not actually have). Objects in the world are such that “it is as if they have the colors.”

Color Realism

These debates raise questions about the strength of the argument against the reality of color. Consider the passage from S. E. Palmer, quoted in the first section. It is possible to see it as presenting an argument against the reality of colors. It takes the form: when objects are seen as colored, the objects are experienced in certain ways, they are seen as having certain qualities, but the objects do not have those qualities; the colors perceived are different from any the physical objects possess. There are two claims implicit here: (i) the colors objects are perceived as having – the “apparent colors” – a certain distinctive character, and (ii) the physical sciences have shown that no qualities with that particular character play any causal part in the perception of colors. From this, it is concluded that neither objects nor lights are colored in anything like the way they are experienced.

This argument depends on one's being able to characterize the distinctive character of the "apparent colors." Here is an important point on which philosophers divide. One group holds that it is obvious that this distinctive character includes a qualitative aspect, one that is familiar to common perceivers. It is by virtue of having this qualitative character that the various colors fit together collectively, in system of relationships of similarities and differences (a psychological color space). In this system, for example, orange is between red and yellow, resembling each, whereas it is not between blue and green, nor blue and red. Another group of philosophers, however, holds that any such qualitative aspect is no part of colors themselves, where the colors are the properties one's experiences *represent* objects as having. (It may be part of the way colors are represented, rather than essential to the color itself. Think here of the difference between *heat* in a body and *the way things are experienced* as hot or cold.) The claim may be put as follows: (1) colors are certain properties, ones that objects look to have (appear to have) when one sees the objects; (2) the properties objects look to have are the qualities one's experiences *represent* the objects as having.

Everything now turns on what theory of *representation* one can call upon. Once again opinions ("carefully considered positions") differ. On one view, the qualities are ones that play the appropriate causal role in the production of the experiences of color. As a consequence, so it is argued, these qualities are ones that it requires scientific inquiry to locate. Accordingly, it is thought there are good reasons to think that surface colors of objects, for example, are spectral reflectances. On the other hand, another group of philosophers reject that theory of representation, arguing that the qualities represented are dispositional properties: dispositions to cause experiences of the right type.

Then again, there are other philosophers who defend a different version of color realism. They accept that the colors do have the relevant qualitative character, but deny that the scientists provide good argument against the reality of color. That is to say, they accept that the ordinary

conception of color is of features with a simple, qualitative character, one that is typically revealed in color perception, but they deny that the scientific argument against the reality of such qualities is cogent. Contemporary defenders of this view appeal to the claim that colors are qualities that are supervenient on light-modifying features of physical bodies. The light-modifying features cause visual experiences, but, so it is argued, colors are qualities that are revealed in experience.

There is another argument against color realism, which has a long pedigree, but which has come to prominence in recent times. One of its strongest advocates has been C. L. Hardin [9]. Any defense of color realism has to acknowledge and defend a distinction between *real color* and (merely) *apparent color* or between *real color* and *illusory color*. Any such distinction, however, needs to rely on a distinction between standard conditions and nonstandard ones and between normal, or standard, observers and nonstandard observers. Things which are *really blue*, say, are ones that will look blue to standard perceivers, in standard conditions. In other sorts of conditions, or to other kinds of perceivers, they may look to have other colors. If so, those other colors will be illusory ones, or "merely apparent" ones.

Hardin has argued that there is no way of identifying such conditions, and such perceivers, that is not arbitrary. A uniformly colored square, for example, will appear differently if the sides are, respectively, 2 cm, 10 cm, and 100 cm. And they will look to have different colors, when set against different backgrounds. The problem is even more pronounced in the case of standard observers. Competent human color perceivers divide into a number of groups that differ, for example, on which hues are instances of the unique hues: unique blue, red, green, and yellow. The differences are quite dramatic. Not only that, but there are significant differences between males and females, with respect to these abilities.

This kind of problem has led some theorists, e.g., Jonathan Cohen, to defend another form of color realism, a version of color relationalism [10]. On this theory, color is a relational property, defined with respect to relations between the object, the environmental viewing conditions,

and the kind of perceiver. On this view, there are no such properties as *blue simpliciter*, *red simpliciter*, etc.; there is only blue-for-perceiver kind A, in conditions C; blue-for-perceiver kind B, in conditions C*; and so on. It should be noted that although Cohen calls the account “realist,” it is a special type of realism. On this account, a whole host of relational properties are equally real.

There are other forms of color relationalism which have links to action-based theories of perception, as developed principally by the psychologist J. J. Gibson. A leading example is the theory defended by Evan Thompson, the *Ecological View of Colors*. On this account, colors are taken to be dependent, in part, on the perceiver and so are not intrinsic properties of a perceiver-independent world. Being colored, instead, is construed as a relational property of the environment, connecting the environment with the perceiving animal. In the case of the color of physical surfaces, “being colored corresponds to the surface spectral reflectance as visually perceived by the animal” [11].

Goethe’s Theory of Colors

This entry would not be complete without reference to the color theory of the famous German author Johann Wolfgang von Goethe. Goethe was deeply critical of Newton’s analytical approach and of his theoretical explanations of color phenomena [12]. While mainstream color science has been mostly dismissive of Goethe’s attack, there are thinkers who find a positive side to Goethe’s work.

There are two aspects to Goethe’s critique. One is that he describes experiments that, he claims, contradict Newton’s theory. The second aspect is that he provides a set of observations that Newton’s theory does not, and cannot, explain. Additionally, he claims to provide an alternative explanation for these phenomena. H. Helmholtz, the nineteenth-century physicist and color scientist, rejected the criticisms of Newton’s science as based on serious misconceptions [13]. First of all, Newton was well aware of some of the

experiments described, and, more importantly, all of these experiments can be satisfactorily handled by Newton’s theoretical approach. With respect to the other observations, it is true that Newton does not explain them, but these phenomena are not ones that his theory is designed to explain. Nevertheless, Helmholtz praised Goethe’s descriptions, of both the experiments and the other observations, as accurate and important.

The issues are perhaps best brought out by a response to Goethe’s work, by the philosopher L. Wittgenstein [14]. For a clear exposition, see an article by another philosopher, Zeno Vendler [15]. Wittgenstein points out that Goethe, despite what he thinks, does not provide a scientific theory at all; his activity has a different nature. What he is doing is better described as providing phenomenological observations, i.e., observations of how colored objects appear and of the relations between those appearances, e.g., that yellow (of the fullest saturation) is lighter than red. The resulting propositions that emerge from this sort of activity are not ones that Newton’s theory is designed to address, but they are very important, nevertheless. (Whether they can be explained by Newton’s theory supplemented by neuroscience is not so easily decided.)

Cross-References

- [Appearance](#)
- [Color Order Systems](#)
- [Color Phenomenology](#)
- [Color Vision, Opponent Theory](#)
- [Helmholtz, Hermann Ludwig von](#)
- [Hering, Karl Ewald Konstantin](#)
- [Psychologically Pure Colors](#)
- [Subjective Colors](#)
- [Unique Hues](#)

References

1. Hume, D.: Treatise of Human Nature. In: Lindsay, A.D. (ed). Dent, London(1738/1911), book III, part I, sect. 1, p. 177; book I, IV, IV, p. 216

2. Maxwell, J.C.: On color vision. In: MacAdam, D.L. (ed.) *Sources of Color Science*, p. 75. MIT Press, Cambridge, MA (1890/1970)
3. Palmer, S.E.: *Vision Science*, p. 95. MIT Press, Cambridge, MA (1999)
4. Descartes R.: Principles of philosophy. In: Cottingham, J., et al. (eds.) *Descartes: Selected Writings* (1644/1988), par. 70; see also par. pp. 68–70
5. Newton, I.: Optics. In: MacAdam, D.L. (ed.) *Sources of Color Science*, pp. 23–24. MIT Press, Cambridge, MA (1730/1970)
6. Hurvich, L.M.: *Color Vision*, p. 2. Sinauer, Sunderland (1981)
7. MacAdam, D.L.: *Color Measurement: Themes and Variations*, p. 1. Springer, Berlin (1985)
8. Hering, E.: Outline of a Theory of the Light Sense, transl. L.M. Hurvich and D. Jameson. Harvard University Press, Cambridge, Massachusetts (1920/1964)
9. Hardin, C.L.: *Color for Philosophers*. Hackett, Indianapolis (1988/1993)
10. Cohen, J.: *The Red and The Real*. Oxford University Press, Oxford (2009)
11. Thompson, E.: Chapter 5. In: *Color Vision*. Routledge, London (1995)
12. Goethe, J.W.: *Goethe's Theory of Colors*, transl. C.L. Eastlake. Cassell, London (1840/1967)
13. Helmholtz, H.: Goethe's scientific researches. In: H. Helmholtz (ed), *Popular Scientific Lectures*. pp. 1–21. Dover, New York (1881/1962)
14. Wittgenstein, L.: In: Anscombe, G.E.M. (ed.) *Remarks on Color*. Blackwell, London (1977)
15. Vendler, Z.: Goethe, Wittgenstein, and the essence of color. *Monist* **78**, 391–418 (1995)

Phosphors and Fluorescent Powders

Weidong Zhuang and Jiyou Zhong

National Engineering Research Center for Rare Earth Materials, General Research Institute for Nonferrous Metals, Beijing, China

Synonyms

[Luminescent materials](#)

Definition

Phosphor (fluorescent powder) is one kind of inorganic materials which absorbs the energy from electrons, photons, or even other forms of

microscopic particles and converts it into visible light. Phosphor (fluorescent powder) has a history of about 100 years and today plays one of the key roles in lighting, display, and imaging. So far there are varieties of phosphors finding their applications in cathode-ray tubes (CRTs), fluorescent lamps, x-ray films, plasma display panels (PDPs), and recently developed white light-emitting diodes (w-LEDs).

The Mechanism of Luminescence

Although phosphors are available in diversity, the mechanisms of luminescence in them are similar, including the following primary items:

- Transitions of rare-earth ions and transition metal ions
- Crystal field splitting and nephelauxetic effect
- Energy transfer

Transitions of Rare-Earth Ions and Transition Metal Ions

Essentially, luminescence of phosphors is a process of electron transitions, which occurs in luminescence center ions (active centers). But not all of the ions can be considered to act as optically active centers, which are usually based on rare-earth ions or transition metal ions. In this section, we will concentrate on these luminescent centers and discuss their energy levels.

Electronic Configuration, Energy Levels, and Transitions of Rare-Earth Ions

The rare-earth ions usually refer to lanthanide ions, Sc^{3+} and Y^{3+} . The electronic configurations of both trivalent and divalent rare-earth ions in the ground states are shown in Table 1. It can be seen that the 4f orbit is empty for Sc^{3+} , Y^{3+} , and La^{3+} and fully filled for Lu^{3+} , leading to no energy levels required for luminescence, while the lanthanide ions from Ce^{3+} to Yb^{3+} have partially filled 4f orbitals, so that they exhibit their characteristic energy levels, which can induce luminescence processes. Additionally, there are three divalent lanthanide ions that are stable, Sm^{2+} , Eu^{2+} , and Yb^{2+} , and their electronic

Phosphors and Fluorescent Powders, Table 1 Electronic configurations and ground state of rare-earth ions

Ions	4f electrons	Ground state	Ions	4f electrons	Ground state
Sc ³⁺	0	¹ S ₀			
Y ³⁺	0	¹ S ₀			
La ³⁺	0	¹ S ₀			
Ce ³⁺	1	² F _{5/2}			
Pr ³⁺	2	³ H ₄			
Nd ³⁺	3	⁴ I _{9/2}			
Pm ³⁺	4	⁵ I ₄			
Sm ³⁺	5	⁶ H _{5/2}	Sm ²⁺	6	⁷ F ₀
Eu ³⁺	6	⁷ F ₀	Eu ²⁺	7	⁸ S _{7/2}
Gd ³⁺	7	⁸ S _{7/2}			
Tb ³⁺	8	⁷ F ₆			
Dy ³⁺	9	⁶ H _{15/2}			
Ho ³⁺	10	⁵ I ₈			
Er ³⁺	11	⁴ I _{15/2}			
Tm ³⁺	12	³ H ₆			
Yb ³⁺	13	² F _{7/2}	Yb ²⁺	14	¹ S ₀
Lu ³⁺	14	¹ S ₀			

configurations are the same as those of Eu³⁺, Gd³⁺, and Lu³⁺, respectively. The electronic state is denoted as ^{2S+1}L_J, where S, L, and J are the spin angular momentum, orbital angular momentum, and total angular momentum, respectively.

The experimentally measured energy levels of trivalent lanthanide ions in lanthanum fluoride (LaF₃) are illustrated in *Dieke diagram* (as shown in Fig. 1). It presents the energy of electronic states ^{2S+1}L_J for trivalent ions in LaF₃ and almost any other crystals. The number of levels is determined by the symmetry of the crystal field surrounding the lanthanide ions, and the width of each level is a measure of the crystal field splitting. By taking advantage of the *Dieke diagram*, one can explain the absorption and emission spectra and assign each spectral peak of rare-earth ions in luminescent materials. That is to say, the transitions of trivalent rare-earth ions can be pointed out according to *Dieke diagram*, especially the luminescence of rare-earth ions with f-f transitions. But for the rare-earth ions, such as Ce³⁺ and Eu²⁺ with f-d transitions, it should be deserved much more attentions for their higher luminous efficiency, which will be discussed as following.

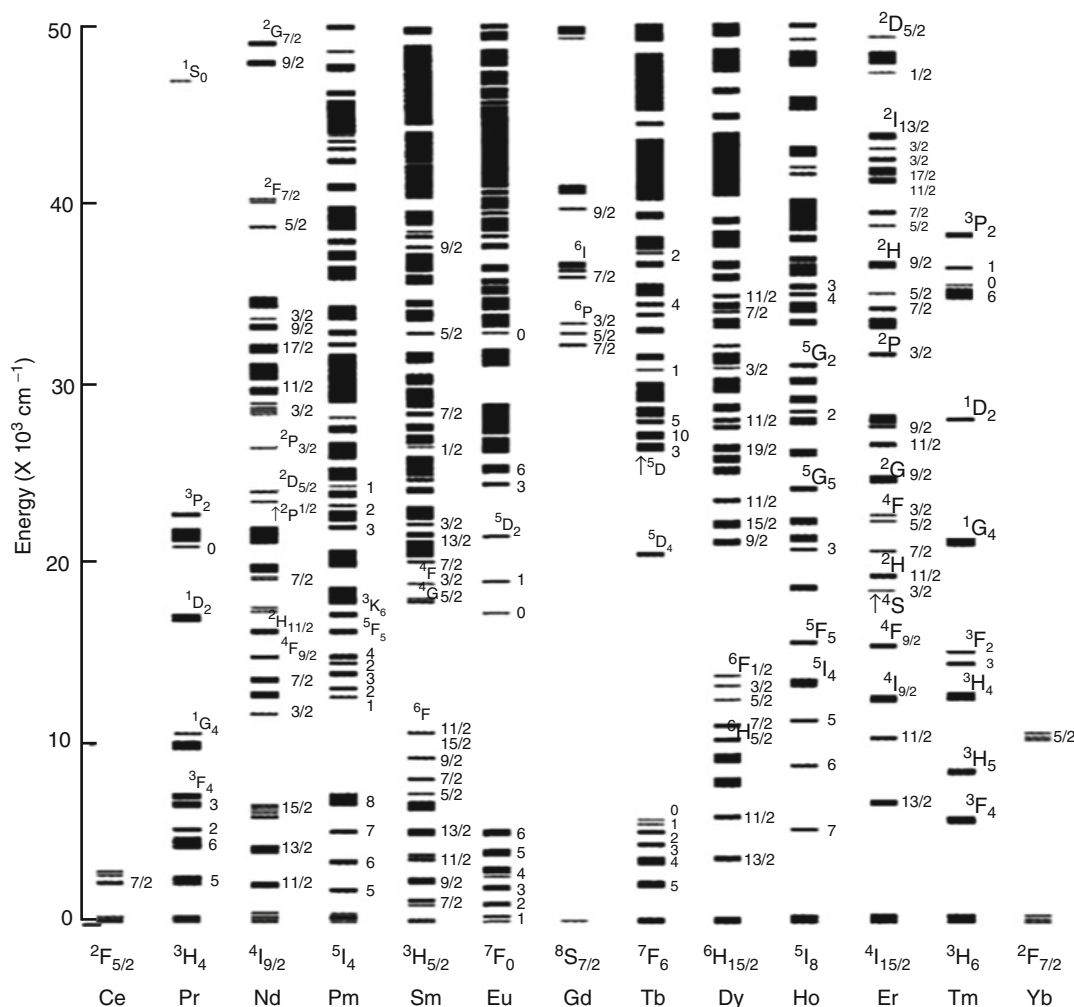
Ce³⁺: The Ce³⁺ ion has the simplest electron configuration among the rare-earth ions. The 4f¹

ground-state configuration is divided into two sublevels, ²F_{5/2} and ²F_{7/2}, and these two sublevels are separated by about 2,000 cm⁻¹ as a result of spin-orbit coupling. This is the reason for the double structure usually observed in the Ce³⁺ emission band. The 5d¹ excited state configuration is split into two to five components by the crystal field, with the splitting number depending on the crystal field symmetry. The Ce³⁺ emission is strongly affected by the host lattice through the crystal field splitting of the 5d orbital and the nephelauxetic effect and usually varies from the blue to the red spectral region.

Eu²⁺: The Eu²⁺ ion has the ground state of 4f⁷ (⁸S_{7/2}) and the excited state of 4f⁶5d¹. The luminescence is strongly dependent on the host lattice, with the emission color varying in a very broad range, from ultraviolet to red. The abundant emission colors are due to the changes in covalency (nephelauxetic effect) and crystal field splitting from one host to another. With increasing the covalency or crystal field splitting, the 5d energy level of Eu²⁺ is lowered greatly, resulting in the red shift of the absorption and emission bands.

Transitions of Transition Metal Ions

Transition metal ions are an important class of dopants in luminescent materials and solid-state



Phosphors and Fluorescent Powders, Fig. 1 Characteristic energy levels of trivalent lanthanide ions, proposed by Dieke and coworkers (Reproduced with permission from Carnall et al. [3]. Copyright 1989, AIP Publishing LLC)

laser materials, and their emission colors are very abundant. Among the transition metal ions, Mn^{2+} and Mn^{4+} are two of the most important and useful luminescent centers in phosphors, which will be discussed below.

Mn^{2+} : The emission color of Mn^{2+} varies from green to orange red, depending on the crystal field strength of the host crystal. The emission spectrum of Mn^{2+} displays a broad band, and the emission corresponds to the ${}^4\text{T}_1({}^4\text{G}) \rightarrow {}^6\text{A}_1({}^6\text{S})$ transition. In general, the emission color is green when Mn^{2+} is tetrahedrally coordinated, and it is orange red when Mn^{2+} is octahedrally

coordinated. The absorption of Mn^{2+} is usually low due to the spin-forbidden transition. To improve the absorption of Mn^{2+} , energy transfer mechanisms are employed to sensitize Mn^{2+} . Commonly used sensitizing ions are Sb^{3+} , Pb^{2+} , Sn^{2+} , Ce^{3+} , and Eu^{2+} , which absorb the ultraviolet (UV) light efficiently and then transfer the energy to Mn^{2+} .

Mn^{4+} : The emission of Mn^{4+} is usually in the red spectral region, due to the ${}^2\text{E} \rightarrow {}^4\text{A}_2$ transition, while the excitation spectral shows broad bands. In CaAl_2O_9 , the emission peak is located at ~ 656 nm, and it is positioned at ~ 715 nm in

YAlO_3 . Mn^{4+} -doped K_2TiF_6 can absorb blue or violet light efficiently and emits at ~ 632 nm, enabling it to be used in high-color-rendering white LEDs.

Crystal Field Splitting and Nephelauxetic Effect

As it is known, when aforementioned luminescence ions are incorporated in a crystal, the surrounding anions will affect them, especially the luminescence ions with d-electron orbits and transitions. Certainly, it affects the luminescence ions with f-f transitions as well by getting rid of forbidden transitions, which would result in luminescence and even high luminescence efficiency. This effect is called crystal field effect, which shown great influence on luminescence ions. Crystal field splitting is, in fact, the energy level splitting of luminescence ions by the influence of crystal field. Generally, the crystal field splitting is affected by symmetry of coordination anions and the distances between luminescence ions and coordination anions. And crystal field splitting is usually described by the point charge electrostatic model (PCEM), which can be usefully applied especially in the case of coordination in the form of an octahedron, cube, tricapped trigonal prism, and cuboctahedron.

Some electrons of the ligands move into the orbitals of the central ion and reduce the cationic valency. Due to this reduction, the d-electron wave functions expand toward the ligands to increase the distances between electrons, reducing the interaction between them. This effect is called the nephelauxetic effect. This effect may be considered to increase with ligands anion in the order: $\text{F}^- < \text{O}^{2-} < \text{N}^{3-} < \text{Cl}^- < \text{Br}^- < \text{I}^- < \text{S}^{2-}$.

Energy Transfer

Energy transfer means energy migrating from a higher energy ion to another lower energy ion or even from a host crystal to activators. Thus, three kinds of energy transfer should be included as following:

The energy transfer between the same ion species.

When the ionic separation becomes smaller, the energy transfer probabilities increases by

means of resonant, exchange interaction, and multipolar interaction. As results, the optical excitation is trapped at defects or impurity sites, enhancing non-radiative relaxation, which is the reason for concentration quenching.

The energy transfer between different ion species can take place when they have closely matched energy levels. The energy transfer result is that the luminescence of energy-accepted activators enhanced and the luminescence of energy-donated activators weaken. Typically, Ce^{3+} ions transfer energy to Tb^{3+} ions in the phosphors $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$; $\text{LaPO}_4:\text{Ce}^{3+}, \text{Tb}^{3+}$; etc.

The energy transfer from a host crystal to activators mainly refers to energy transfer from an excited oxy-anion complex to lanthanide ions. This kind of energy transfer is commonly seen in vanadate, molybdate, tungstate, etc., such as $\text{YVO}_4:\text{Eu}^{3+}$, $\text{CaMoO}_4:\text{Eu}^{3+}$, and $\text{Y}_2\text{WO}_6:\text{Eu}^{3+}$.

Characterization

Besides the mechanism of luminescence, characterization of the prepared phosphors is vital to evaluate the practical application. The principal characterization of phosphors can be listed as following:

- Spectrum
- Quantum efficiency
- Thermal quenching
- Fluorescent lifetime

Spectrum

When light interacts with a phosphor, refraction, absorption, scattering, and luminescence will occur. The luminescence process – excitation and emission – will be discussed and characterized by excitation and emission spectra.

Emission Spectrum

During the luminescence process, the distribution of emission photon number varies with emission wavelength or frequency, that is, the so-called emission spectrum, which is important to judge

what color of light emitted by the phosphor and to calculate the color coordinates. Generally, two kinds of emission spectrum are presented: line spectrum for activators with f-f transitions and band spectrum for activators with d-electron transitions. No matter what kinds of emission spectrum, it shows difference with activator species and compositions of host.

Excitation Spectrum

The distribution of emission photon number of the selected wavelength varies with excitation wavelength or frequency. Thus, it, in fact, is a specific emission spectrum, which is used to characterize the effectiveness of excitation. From this spectrum, it is convenient to judge the effective excitation wavelength for phosphors, which is important for users.

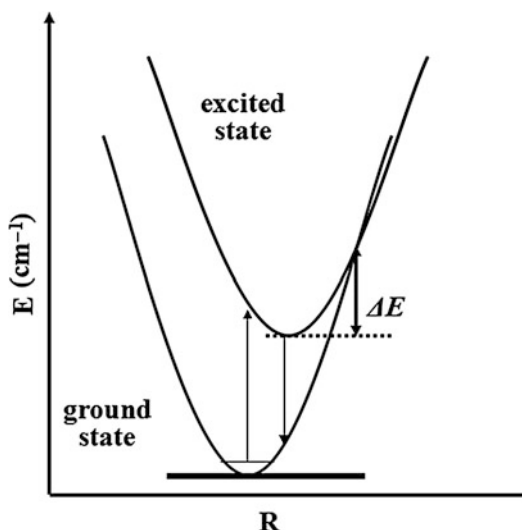
Quantum Efficiency

The emission and excitation spectra let people know what light the phosphor emits and the effectiveness of different excitation wavelength, while it still does not know how efficient the energy conversion is. Thus, the quantum efficiency is introduced. The quantum efficiency of a phosphor is how much light (photons) absorbed by the phosphor is converted into luminescence, which is in the range of 0 ~ 1. This is also called internal quantum efficiency. The external quantum efficiency denotes the ratio of the number of photons of the emitted light to that of the incident light on a phosphor. The internal (η_i) and external (η_0) quantum efficiency can be computed by the following equations:

$$\eta_i = \frac{\int \lambda \cdot P(\lambda) d\lambda}{\int \lambda \{E(\lambda) - R(\lambda)\} d\lambda}$$

$$\eta_0 = \frac{\int \lambda \cdot P(\lambda) d\lambda}{\int \lambda \cdot E(\lambda) d\lambda}$$

where $E(\lambda)/h\nu$, $R(\lambda)/h\nu$, and $P(\lambda)/h\nu$ are the number of photons in the spectrum of excitation,



Phosphors and Fluorescent Powders,
Fig. 2 Configurational coordinate diagram for thermal quenching

reflectance, and emission of the phosphor, respectively.

Thermal Quenching

Thermal quenching of luminescence is the phenomenon in which the luminescence efficiency decreases as the temperature increases due to the increasing phonon–electron interaction. This phenomenon can be explained by configuration coordinate (as shown in Fig. 2). The excited electron can go across the crossing point of ground state and excited state and relax to the ground state non-radiatively, the probabilities of which will increase with the increasing of temperature. The energy difference between the lowest vibrational level of excited state and the crossing point is called activation energy for thermal quenching, which is used to judge the thermal quenching properties of phosphors. The larger value of activation energy means higher energy barrier for thermal quenching. The activation energy can be calculated by the Arrhenius equation:

$$I_T = \frac{I_0}{1 + c \exp\left(\frac{-\Delta E}{kT}\right)}$$

Here I_0 is the initial emission intensity, I_T is the intensity at different temperatures, ΔE is the activation energy of thermal quenching, c is a constant for a certain host, and k is the Boltzmann constant.

Fluorescent Lifetime

Fluorescent lifetime is another important factor, which is not used to judge the phosphors good or bad but limits the applications of phosphors. For display and imaging applications, short fluorescent lifetime is needed as well as high luminescence efficiency and narrow full width at half maximum (FWHM) of emission spectrum. The decay process of the luminescence intensity $I(t)$ after the termination of excitation at $t = 0$ is generally represented by an exponential function of the elapsed time after the excitation:

$$I(t) = I_0 \exp(-t/\alpha)$$

where α is the decay time constant of the emission and $\tau = -1/\alpha$ is named fluorescent lifetime. For convenience, fluorescent lifetime is defined as the duration of $I(t)$ decreasing from I_0 to $1/e I_0$. Generally, light emission due to a forbidden transition has a long fluorescence lifetime, while that due to an allowed transition usually shows a short fluorescence lifetime, but that is not absolute since the electron may be captured by a trap or lattice defect, leading to long afterglow luminescence. Fluorescent lifetime is usually determined by measuring the time-resolved emission spectroscopy, which provides information on crystal structure and the kinetic behavior of luminescent excited states and intermediates as well.

Practical Phosphors

As mentioned above, phosphors are various and have shown varieties of applications in lighting, display, and imaging. However, the lighting industry constitutes the largest consumer of phosphors and produces the largest quantity of phosphor-related products. Thus, the widely used phosphors for fluorescent lamps and w-LEDs are chosen and introduced as following:

- Phosphors for fluorescent lamps
- Phosphors for w-LEDs

Phosphors for Fluorescent Lamps [1]

The invention of fluorescent lamp can be traced back to the nineteenth century while the calcium halophosphate phosphor was developed, and the age of fluorescent lamps was truly entered since 1950s. However, the problems of poor color-rendering and high color temperature of this lamp prompted the development of three-band fluorescent lamp, which shows higher color-rendering and even higher luminescence efficiency. The practical phosphors for fluorescent lamps will be introduced briefly as following:

Calcium Halophosphate Phosphors for Traditional Fluorescent Lamps

Crystal structure: These phosphors have an apatite-type structure belonging to the hexagonal system. The Ca^{2+} ions have two different sites: Ca (I) coordinated by six oxygen and Ca (II) coordinated by halogen ions. Sb^{3+} and Mn^{2+} ions are capable of replacing Ca^{2+} ions at both types of sites. In addition, the Ca^{2+} ions are possible to be partially replaced by Sr^{2+} , Ba^{2+} , and Cd^{2+} ions.

Luminescent properties: These series of phosphors are doubly activated by Sb^{3+} and Mn^{2+} . The Sb^{3+} emission peak is located at ~ 480 nm and is not influenced by the kind of halogen in the host. The emission peak of Mn^{2+} is located at $575 \sim 585$ nm, depending on the F:Cl ratio. Energy transfer from Sb^{3+} to Mn^{2+} can be observed and show great influence on emission colors by turning $\text{Sb}^{3+}:\text{Mn}^{2+}$ ratio. By adjusting appropriate compositions of these phosphors, the color temperatures of the white light can be obtained ranging from 3,000 to 6,500 K. In addition, these series phosphors have almost 90 % of the emission intensity retained, even at 200°C , without a shift in the emission peak.

Phosphors for Three-Band Fluorescent Lamps

Phosphors utilized for three-band lamps can be grouped into three categories as following:

1. Red phosphor: $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ is the only available red phosphor for three-band lamps mainly due to its nearly perfect luminescence properties, except for the high price.

Crystal structure: The crystal structure of this phosphor is cubic and Eu^{3+} ions occupy the two Y^{3+} sites of C_2 and S_6 symmetry.

Luminescent properties: The excitation spectrum has a broad band ranging from 200 to 300 nm, which matches perfectly with 254 nm light emitted by mercury vapor. The emission spectra contains a very narrow and intense emission band at 611 nm (originates from the $^5\text{D}_0 \rightarrow ^7\text{F}_2$ transition of Eu^{3+}) with a full width at half maximum of ~ 7 nm. This phosphor exhibits high color purity, high quantum efficiency (nearly 100 %), excellent thermal quenching properties, and stable chemical properties.

2. Green phosphors: There are three kinds of practicable green phosphors being developed successively, which are $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$; $\text{LaPO}_4:\text{Ce}^{3+}, \text{Tb}^{3+}$; and $\text{GdMgB}_5\text{O}_{10}:\text{Ce}^{3+}, \text{Tb}^{3+}$.

Crystal structures: $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$ belongs to the hexagonal system and has the same magnetoplumbite structure as the $\text{PbFe}_{12}\text{O}_{19}$ crystal. $\text{LaPO}_4:\text{Ce}^{3+}, \text{Tb}^{3+}$ belongs to the monazite crystal group. $\text{GdMgB}_5\text{O}_{10}:\text{Ce}^{3+}, \text{Tb}^{3+}$ has monoclinic symmetry and is isostructural with $\text{LaCoB}_5\text{O}_{10}$.

Luminescent properties: The common characteristics of these phosphors are activated by Tb^{3+} and adopting Ce^{3+} as sensitizer, because the absorption lines of Tb^{3+} in the ultraviolet region does not match the light emitted by mercury vapor, but Ce^{3+} can absorb the light emitted by mercury vapor efficiently and transfer the energy to Tb^{3+} with a high efficiency by overlapping the emission spectra of Ce^{3+} and absorption lines of Tb^{3+} . Thus, these phosphors exhibit the emission of Tb^{3+} with emission peak at ~ 543 nm (corresponding to the $^5\text{D}_4 \rightarrow ^7\text{F}_5$ transition) and a full width at half maximum of ~ 10 nm. But by comparing other properties, the $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$ presents much more excellent in quantum efficiency (as high as 97 %), thermal quenching

properties, and chemical stability; therefore, the $\text{CeMgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$ phosphor is still one of the most important and widely used green phosphors for three-band fluorescent lamps.

3. Blue phosphors: $(\text{Sr}, \text{Ca}, \text{Ba})_5(\text{PO}_4)_3\text{Cl}:\text{Eu}^{2+}$ and $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ are two main classes of blue-emitting phosphors.

Crystal structures: $(\text{Sr}, \text{Ca}, \text{Ba})_5(\text{PO}_4)_3\text{Cl}:\text{Eu}^{2+}$ has an apatite-type structure belonging to the hexagonal system. $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ has a crystal structure similar to hexagonal β -alumina.

Luminescent properties: $(\text{Sr}, \text{Ca}, \text{Ba})_5(\text{PO}_4)_3\text{Cl}:\text{Eu}^{2+}$ gives a sharp emission spectrum peaking at 447 nm and a broad band excitation spectrum overlapping 200 \sim 400 nm region. Partial replacement of Sr^{2+} by Ca^{2+} or Ba^{2+} results in red shift or blue shift, mainly ascribed to the change of crystal field. Incorporation of a small amount of Ba^{2+} serves to improve the degradation characteristics; incorporation of a small amount of borate may increase the emission intensity. $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ also has a broad band excitation spectrum in the ultraviolet region and an emission band with the peak located at ~ 450 nm. Generally, the luminescent properties can be adjusted by controlling the Al_2O_3 ratio, Eu^{2+} concentration, and Sr^{2+} replacing part of Ba^{2+} . By comparison, $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ shows higher quantum efficiency (appears to exceed 100 % measured by conventional methods). The relative poor thermal quenching properties are still challenges for both blue phosphors.

Phosphors for w-LEDs

W-LEDs are thought to have brought about a revolution in energy-efficient lighting due to the outstanding advantages such as long lifetime, high energy efficiency, and environmental friendliness. Currently, w-LEDs are commonly generated by the combination of the blue LED chips and yellow-emitting phosphors. But, the white light obtained by this approach would result in low color-rendering index and high correlated color temperature due to the red and green

emission deficiency in the visible spectrum. Generally, red-emitting phosphors are added to make up for this deficiency. Thus, the widely used phosphors for w-LEDs can be divided into two groups as following:

Oxide Yellow Phosphors

1. YAG:Ce³⁺ garnet phosphor

Crystal structure: YAG is the abbreviation of Y₃Al₂Al₃O₁₂, which has the classical garnet structure represented by the general formula A₃B₂X₃O₁₂, where A, B, and X are eight, six, and four coordinated with the surrounding O, forming dodecahedron, octahedron, and tetrahedron, respectively. The octahedron and tetrahedron do not share any edge among themselves, but they share edges with at least one dodecahedron.

Luminescent properties: The excitation spectrum of YAG:Ce³⁺ is mainly composed of two broad bands with excitation peaks located at 340 and 460 nm, respectively. The emission spectrum shows a very broad band with emission peak at ~535 nm and a full width at half maximum (FWHM) of ~100 nm. The high external quantum efficiency (more than 83 %), excellent thermal quenching properties (more than 87 % of the initial emission intensity remained at 150 °C), and the strong absorption of blue light emitted by blue LED chip made YAG:Ce³⁺ one of the most important and widely used phosphors for white LEDs. Actually, YAG:Ce³⁺ is a green-yellow-emitting phosphor, and the emission spectrum in the red region is very weak, which results in the produced white light having a high color temperature. Generally, La³⁺ and Gd³⁺ replacing part of Y³⁺ or increasing Ce³⁺ concentration may turn the emission spectrum shift toward longer wavelength, but the quantum efficiency decreases radically; meanwhile, the thermal quenching increases.

2. Sr₂SiO₄:Eu²⁺ silicate phosphor

Orthosilicate phosphor has outstanding advantages such as high luminescence efficiency, low cost, and strong excitation in blue to UV region, which have attracted lots of attentions.

Crystal structure: Sr₂SiO₄ has two types of crystal structures, which are orthorhombic corresponding to α' phase and monoclinic corresponding to β phase. The α' and β phases of Sr₂SiO₄ are isostructural with Ba₂SiO₄ and β-Ca₂SiO₄, respectively. Generally, in order to stable the α'-Sr₂SiO₄ phase, Ba²⁺ ions were introduced to replace part of Sr²⁺ ions. For α'-Sr₂SiO₄, Sr²⁺ ions have two kinds of sites, Sr(1) and Sr(2), coordinated by 10 and 9 oxygen atoms, respectively.

Luminescent properties: The excitation spectrum of Sr₂SiO₄:Eu²⁺ shows a broad band covering the spectral region of 200–550 nm with excitation peak at about 387 nm. The emission spectrum shows two bands centered at about 490 and 560 nm, assigning to the luminescence centers occupying two types of Sr sites: Sr(1) and Sr(2), respectively. Due to the weak 490 nm emission peak, the emission spectra of Sr₂SiO₄:Eu²⁺ shows emission peak at about 560 nm, which can be shift to longer wavelength by increasing the SiO₂ content. Thermal quenching of this kind of silicate phosphor is unsatisfactory for the emission intensity decreases by 62 % when the temperature rises to 150 °C compared to the initial intensity measured at room temperature. Though the luminescence efficiency of Sr₂SiO₄:Eu²⁺ is comparable with YAG:Ce³⁺, the thermal stability and structure stability are still limitations for its applications.

Nitride Red Phosphors [2]

1. M₂Si₅N₈:Eu²⁺ (M = Ca, Sr, Ba) phosphors

Crystal structure: Ca₂Si₅N₈ has a monoclinic crystal system with the space group of Cc, whereas both Sr₂Si₅N₈ and Ba₂Si₅N₈ belong to orthorhombic crystal system with the space group of Pmn2₁. There are two different crystallographic sites for alkaline earth metals. Each Ca²⁺ ion in Ca₂Si₅N₈ is coordinated by seven nitrogen ions, while Sr²⁺ and Ba²⁺ in Sr₂Si₅N₈ and Ba₂Si₅N₈ are coordinated by eight or nine nitrogen ions.

Luminescent properties: The M₂Si₅N₈:Eu²⁺ (M = Ca, Sr, Ba) phosphors show very intense red emission under blue light excitation. The emission color changes from orange to deep red depending on the alkaline earth metal ion

ratio. The emission spectra all exhibit a broad and asymmetrical band extending from 550 to 850 nm. Upon the 450 nm excitation, $\text{Sr}_2\text{Si}_5\text{N}_8:\text{Eu}^{2+}$ shows an absorption and external quantum efficiency of 87 and 71 %, respectively.

2. $\text{CaAlSiN}_3:\text{Eu}^{2+}$ phosphor

Crystal structure: CaAlSiN_3 belongs to orthorhombic crystal system with $\text{Cmc}2_1$ space group. The Ca^{2+} ions have only one crystallographic site, which is coordinated by five nitrogen ions.

Luminescent properties: The excitation spectrum covers a broad spectral range of 250–600 nm, showing an extremely broad band and strong absorption of visible light. The emission spectrum extends from 550 to 800 nm, showing a broad band centered at 655 nm with a FWHM value of 93 nm. When the phosphor is excited by the blue light ($\lambda = 450$ nm), the external quantum efficiency is up to 70 %. In addition, it has a small thermal quenching. The luminescence is quenched by 10 % of the initial intensity when the phosphor is heated to 150 °C and excited at 450 nm.

References

1. Yen, W.M., Hajime, Y. (eds.): Phosphor Handbook. CRC Press, Boca Raton (2012)
2. Xie, R.J., Li, Y.Q., Hirotsaki, N., Yamamoto, H.: Nitride Phosphors and Solid-State Lighting. Taylor & Francis, Boca Raton (2011)
3. Carnall, W.T., Goodman, G.L., Rajnak, K., Rana, R.S.: A Systematic Analysis of the Spectra of the Lanthanides Doped into Single Crystal LaF_3 . J. Chem. Phys. **90**, 3443 (1989)

Photodetector

Bao-Jen Pong

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

Light sensors; Photosensors

Definition

Photodetectors or photosensors are sensors that are used to transform incoming optical energy into electrical signals. A photodetector used in color applications typically covers the visible spectrum range (380–780 nm).

Overview

Figure 1 shows a typical photodetector. It includes three stages: (1) to generate electron–hole pairs or electrons in photosensitive material using an incident light; (2) to transform them through a bias provided by a power supply; and (3) to output the electrical signal directly or via an amplifier [1–7].

Photodiode and Photomultiplier Tube

Figure 1 shows a photodiode, a circuit including a p–n junction semiconductor and can output electrical signal. It can be seen that the photosensitive material includes a p–n junction, a boundary or interface between n-type and p-type semiconductors, and a depleted semiconductor region with a high electric field that can separate the electron–hole pair.

As incident light on a photocathode (photosensitive material is made of a multi-alkaline material, for example), it generates photoelectrons; next, the generated photoelectrons travel in a vacuum tube and hit the first dynode and generate secondary electrons. Then, the secondary electrons are multiplied by a series of dynodes and reach an anode to produce an electrical signal. As the electrical signal is multiplied many times and the whole process is performed in a vacuum tube to prevent electron absorption by the air, it is called a photomultiplier tube (PMT), as shown in Fig. 2.

Photodiode Array and Charge-Coupled Device

As the discrete photodiodes form an array with an integrated circuit board, it is called a photodiode array.

As an integrated circuit contains an array of capacitors that store charge of electron–hole pairs created by photosensitive material, it is called a charge-coupled device (CCD).

Applications

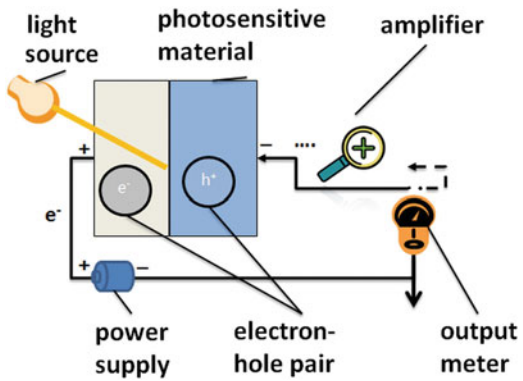
In color applications, photodetectors are included in many color measuring instruments, which measure parameters such as luminance, illuminance or chromaticity. Sze [1] divided photodetectors into four categories: photoconductor, photodiode, avalanche photodiode, and phototransistor. The photodiode and the phototransistor are typically used in color applications.

An Si photodiode together with a filter to mimic the luminous efficiency function (i.e., V

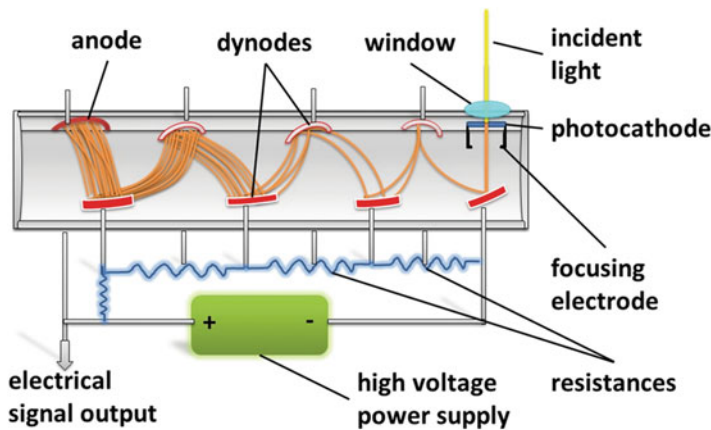
(λ) filter) becomes a $V(\lambda)$ photodetector, which is used for measuring the luminance (in cd/m^2 units). In a colorimeter, an Si photodiode combined with filters to mimic the color-matching functions of \bar{x} , \bar{y} , and \bar{z} are used to measure CIE tristimulus values (X , Y , and Z).

A PMT can detect dim light; it usually works with a monochromator. This arrangement can measure spectral information across the spectrum. The measured spectra multiplying the luminous efficiency function $V(\lambda)$ can obtain luminance.

In color applications, a photodiode array is placed in an array image plane to form an image-based colorimeter. A CCD combined with the desired color filters measures an array of chromaticity values. Recently, in the display and imaging industry, CCD has been replaced by a complementary metal–oxide–semiconductor (CMOS) as an image-type photodetector because of its low cost. Figure 3 shows an image luminance measurement device (ILMD) [5] that applies a commercial color CCD camera or color CMOS camera to measure the image quality of displays. ILMD is also used for measuring luminaires of road and tunnel lighting, indoor measurements (uniformity, contrast, glare, etc.), near-field goniophotometer, etc. [6].

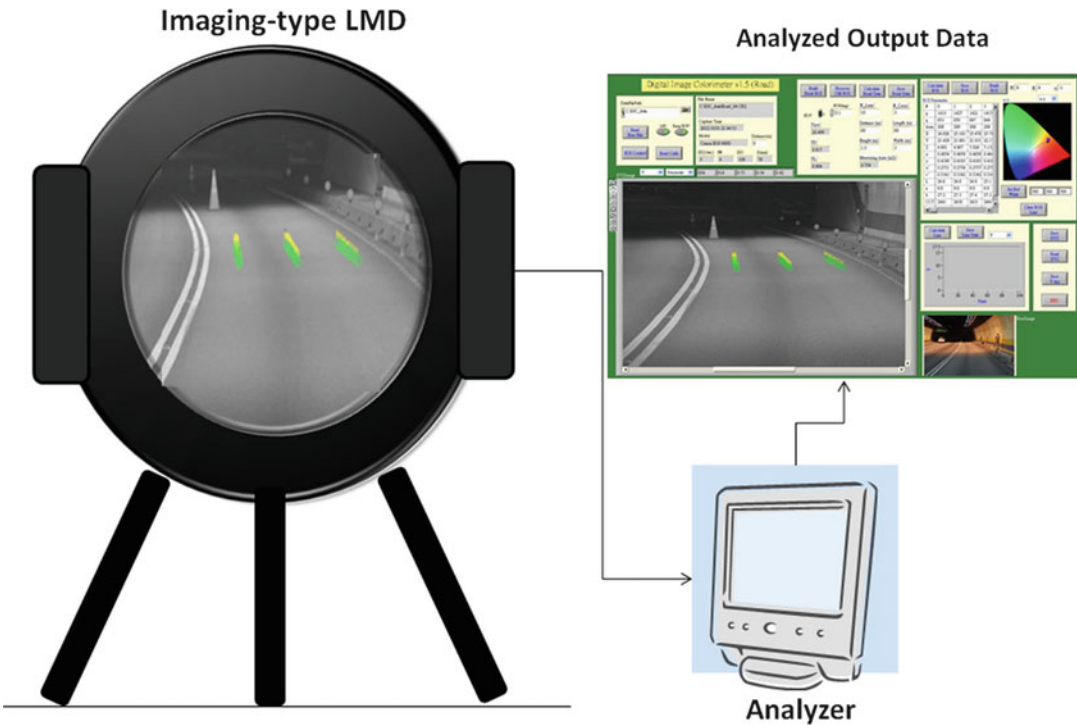


Photodetector, Fig. 1 A schematic diagram to show a photodetector (i.e., also a photodiode)



Photodetector, Fig. 2 A schematic diagram to show a photomultiplier tube. The incident light through the window hits a photocathode (photosensitive material is made by a multi-alkaline material, for example). It generates photoelectrons, which transmit in a vacuum, hit on the first dynode, and is then multiplied by a series of dynodes,

finally reaching an anode to produce an amplified electrical signal. To prevent the absorption of photoelectrons and generated electrons into the air, a vacuum tube is used. Finally, the output electrical signal is amplified 10^4 to 10^5 times and within in a pico-ampere to micro-ampere current range



Photodetector, Fig. 3 The schematic diagram of an image luminance measuring device (ILMD). The color image is captured through an ILMD. It is then transformed

into a luminance image through the analyzer, and the luminance distribution can be used for many applications, such as to calculate the glare index

Cross-References

- CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams
- Instrument: Colorimeter
- Instrument: Photometer
- Instrument: Spectrophotometer
- Luminous Efficiency Function
- Monochromator
- Spectroradiometer

References

1. Sze, S.M.: Physics of Semiconductor Devices, 2nd edn. Wiley, Taipei (1983)
2. Razeghi, M.: Fundamentals of Solid State Engineering. Kluwer, New York (2002). ISBN 978-0-7923-7629-3
3. Omnes, F.: Introduction to semiconductor photodetectors. In: Decoster, D., Harari, J. (eds.) Optoelectronic Sensors. ISTE, London (2009). ISBN 978-047-06-1163-0
4. Luna-López, J.A., Aceves-Mijares, M., Carrillo-López, J., Morales-Sánchez, A., Flores-Gracia, F. Vázquez Valerdi, D.E.: UV-vis Photodetector with silicon nanoparticles. In: Gateva, S. (ed.) Photodetectors. InTech, ISBN:978-953-51-0358-5 (2012)
5. Hsu, S.W., Chung, T. Y., Pong, B. J., Chen, Y. C., Hsieh, H. P., Lin, M. W.: Relations between Flicker, Glare, and perceptual ratings of LED billboards under various conditions. OP58, Proceeding of CIE Century Conference "Towards a New Century of Light", pp. 428–434, Pari, April 15 and 16, ISBN: 978-3-902842-44-2 (2013)
6. Available at: <http://www.cie.co.at/div2/documents/2003%20meeting%20materials/R2-29b%20Blattner.pdf>. Accessed 23 Jan 2014
7. Ready, J: Fundamentals of photonics: module 1.6 optical detectors and human vision. SPIE 2009. Available at: <http://spie.org/x17229.xml>. Accessed 23 Jan 2014

Photoreceptor Cells

- Photoreceptors, Color Vision

Photoreceptors, Color Vision

Marisa Rodriguez-Carmona
Optometry and Visual Science, City University
London, London, UK

Synonyms

Cone pigments; Opsin genes; Photoreceptor cells;
Visual pigments

Definition

It is the molecular structure of the genes that encode the human cone photopigments and their expression in photoreceptor cells. Alterations in these genes have consequences on the spectral sensitivity of the cone photopigments which in turn affect how the color-normal and color-deficient individual see the world.

Overview

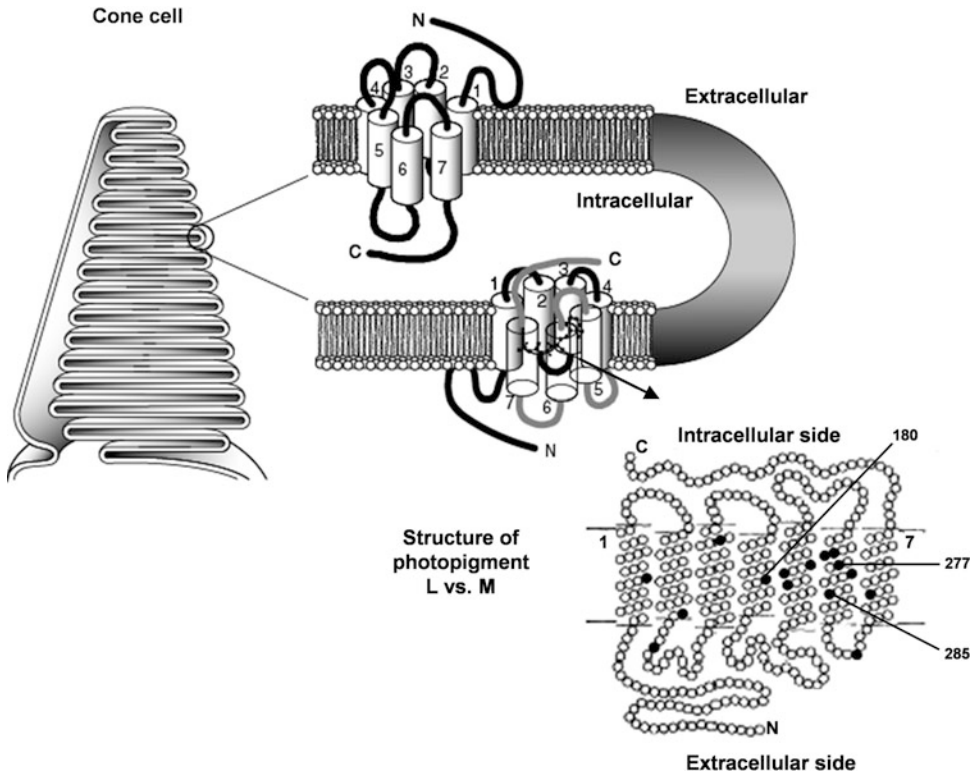
Normal trichromats can match any color by combining three suitably chosen primaries in appropriate proportions. Those that differ from normal trichromats with respect to the proportions of the primaries may either require that the three primaries be present in unusual quantities or may only need two primaries. The first type of variation is known as anomalous trichromatism, and it is assumed that it arises when three classes of cones are present, one of which contains a photopigment with an anomalous absorption spectrum. The second type of variation is known as dichromatism and occurs if only two of the three classes of cones are present. The type of deficiency involved causes differences in the chromatic sensitivity that can be identified easily by psychophysical tests of color vision. Different classes of phenotypes can be identified in this way and correspond to L-, M-, or S-cone deficiency. On the other hand genotype refers to characteristics of an individual's genes. Studies of the

empirical relation between a genetic code and an expressed phenotype have led to improved understanding in the physiological mechanisms underlying many individual differences in color vision.

Nathans and colleagues [1] were the first to use advances in molecular genetics to study human photopigment genes, involving the technique of recombinant DNA (deoxyribonucleic acid). Visual pigments are the light-absorbing molecules that mediate vision with absorption maxima at approximately 420 nm (S-), 530 nm (M-), and 560 nm (L-wavelength sensitive pigment). Structurally, they consist of opsin, transmembrane heptahelical proteins of a single polypeptide chain composed of either 364 (M- and L-cone pigment genes) or 348 (S-cone pigment gene) amino acids bound to a chromophore, 11-*cis* retinal. Photon absorption by the pigment molecules initiates visual excitation by causing an 11-*cis*-to-all-*trans* isomerization of the chromophore. The binding socket site of the chromophore in both the cone and rod opsins is located in helix 7 (Fig. 1).

The precise wavelength maxima (λ_{\max}) of a given pigment will be determined by the amino acid sequence of the opsin and the interaction of specific amino acids with retinal [3]. The exact amino acid sequence of opsin "spectrally tunes" the pigment to a specific wavelength [4, 5].

The genes encoding the S-cone pigments reside alone as a single copy on the q-arm (long arm) of chromosome 7 [6]. The genes encoding the L and M-cone pigments reside on the q-arm of the X-chromosome organized in a tandem array [1] of up to six exons (coding regions of DNA) separated by relatively long introns (noncoding regions of DNA) (Fig. 3). The S-cone pigment shows only 43 ± 1 % analogy in the amino acid sequence in comparison with either the M- or L-cone photopigments. In contrast, the M- and L-cone pigment genes are 96 % homologous, differing only in 15 amino acids [2]. The approximate 30 nm spectral difference between these two visual pigments must be attributable to substitutions at these particular amino acid positions. These differences are confined to exons 2–5. The largest shifts in λ_{\max} are produced by substitutions at two key sites within exon 5: amino acid positions 277 (~7 nm) and 285 (~14 nm) (Fig. 1).



Photoreceptors, Color Vision, Fig. 1 Arrangement of visual pigments in a cone photoreceptor. The infolding membrane of the cone outer segment is packed with photopigment molecules. The molecule consists of seven α -helices which span the membrane of the cell surrounding the chromophore, 11-*cis*-retinal. The NH₂ (N) terminus is extracellular, whereas the COOH (C) terminus is intracellular. Shown in the *bottom right* is the sequence of amino

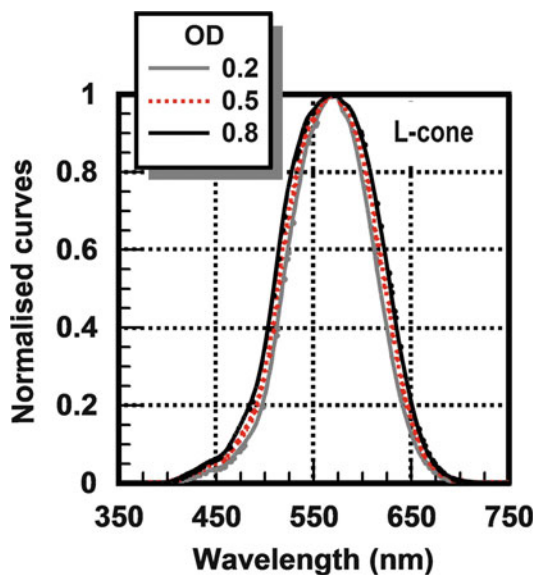
acids that make up this heptahelical molecule. Comparison between the L- and M-cone pigments shows the amino acids that are equal (*white circles*) and different (*dark circles*). The substitutions at positions 180, 277, and 285 are believed to contribute most of the spectral difference between the M- and L-cone pigments (Adapted from Sharpe [2] and Nathans [1])

P

These result in spectral shifts of 15–25 nm, the exact value depending on sequences in exons 2–4. Substitutions at the sites in exons 2–4 produce much smaller shifts of less than ~4 nm and may be responsible for the subtle differences underlying anomalous and normal color vision. Both *in vitro* [1, 4] and *in vivo* [7] methods used to investigate protein sequence variation and spectral sensitivity are in close agreement. These methods have also shown that amino acid substitutions in exon 2 contribute very little to spectral tuning (0.0–0.7 nm) [4, 7]. More recently, it has been speculated that the amino acid differences in exon 2 are involved in controlling the optical density of the L- and M-cone photopigments [8]. The optical density depends

on the concentration of the photopigment in the cone outer segment, the length of the cone outer segment, and also the extinction coefficient which describes the probability of a photon being absorbed [9]. Changes in the effective optical density of cones cause a broadening of the spectral sensitivity curve away from the absorption peak of the pigment (Fig. 2). If two genes that differ only in their exon 2 sequences are expressed, the resulting differences in the optical density of the photopigments may account for some residual color vision discrimination [8].

The close similarities between the L- and M-cone sensitive pigment genes suggest evolution from a single gene. Comparison of amino acid sequences suggests that S-cones and rod



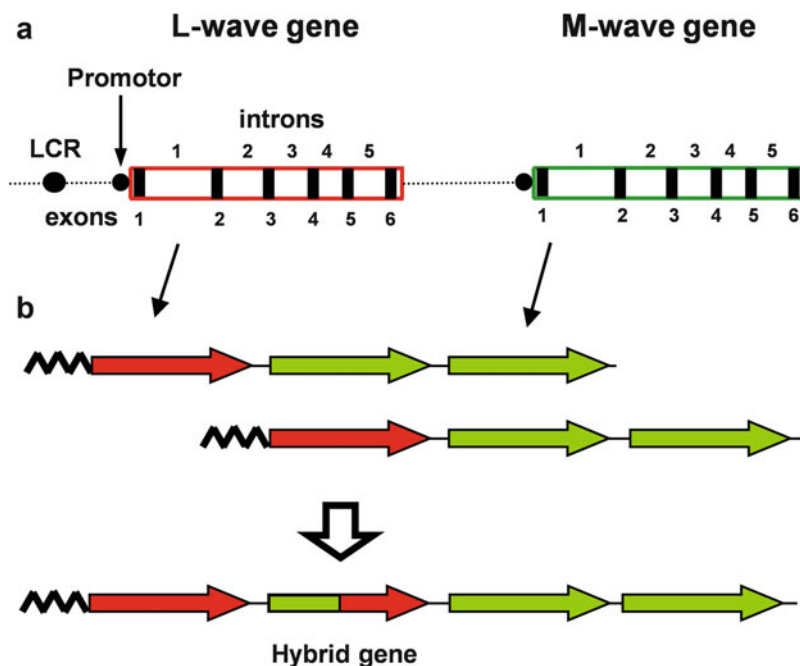
Photoreceptors, Color Vision, Fig. 2 Theoretical spectral sensitivity curves for the L-cone sensitive pigment. Changes in the optical density of the cone pigment may occur from amino acid substitutions in exon 2 altering the stability of the molecule or the efficiency with which it absorbs light. For wavelengths near the peak, the optical density qualitatively mimics the difference produced by a spectral shift

receptors arose first from a common ancestral receptor. From comparisons with contemporary New World monkeys, which only have two photopigments, it is thought that a long-wavelength gene duplicated and diverged to originate the red and green photopigments, at the time when Old World monkeys (trichromatic) separated from New World monkeys [1, 10]. The location of the genes for the M- and L-cone pigments on the X-chromosome can account for the larger number of (red/green) color-deficient men compared to women, as men have only one X-chromosome. The head-to-tail arrangement of the L- and M-cone genes on the X-chromosome is susceptible to mispairing during meiosis, leading to unequal crossing over between gene arrays. If the crossover occurs between genes, this will result in the deletion of a gene from one chromosome and its addition to the other, whereas a crossover within a gene will lead to the production of a hybrid gene that combines regions of the L and M genes into a single gene. Such hybrid

genes are thought to be the genetic basis responsible for the majority of color vision defects [1].

Figure 3 shows an example of how the formation of hybrid genes produces pigments with abnormal spectral sensitivity. The spectral sensitivity of the photopigment will depend on which L- or M-cone gene the crossover originated from [12]. Visual pigment gene arrays for people with normal color vision have an L gene in the most upstream or leftmost position and M genes downstream or to the right of the L gene. Color normals usually have only one copy of the L-cone gene, multiple copies of M, and possibly a number of hybrids [2]. There is good evidence from studies of gene expression in the retina and from studies involving phenotype-genotype relationships [13] that only the first two genes are expressed and have a significant role in color vision. While color normals have an L- and an M-cone gene occupying the two most proximal positions of the array from the locus control region (LCR), anomalous trichromats must have a hybrid gene positioned in either the first or second position. Deutan color vision arises if a hybrid gene encoding an M-cone gene sensitive pigment occupies second position. Protanomalous observers are missing normal L-sensitive photopigments, and color vision arises from two M-cone genes that differ subtly in spectral absorption properties [5]. The M- and L-cone opsin gene can take many forms giving hybrid variants and polymorphisms, and it is this variability underlying the huge differences in red/green color vision observed within the population. Even among people with normal color vision, differences in color matching behavior are associated with individual differences in L-cone pigments [11].

By contrast, the S-cone opsin gene sequence is nearly invariant in the human population; however, further phenotype-genotype studies are yet to be performed to prove this. Intragenic crossover, the mechanism that permits the frequent manifestation of anomaly in protan and deutan defects, has no analogy in tritan defects. No polymorphisms causing shifts in λ_{max} have been reported so far, and only one substitution was found in the coding sequences and exon-intron junctions [14]. Three mutations have been



Photoreceptors, Color Vision, Fig. 3 Schematic of the tandem array of L- and M-cones on the q-arm of the X chromosome. (a) The LCR (locus control region) can activate only one of the promoter regions just upstream of a gene. The promoters are regulatory units upstream of the transcription site and regulate the rate of DNA transcription into RNA and hence the amount of opsin gene expression

(for DNA, upstream is to the left and downstream to the right). The opsin-coding sequences are divided into exons (*black bars*). The intron sequences (*gaps*) are silent or noncoding and usually believed to have no apparent function. (b) Schematic of unequal intragenic crossover that would produce hybrid genes (Adapted from Neitz and Neitz [11])

established causing amino acid substitutions that perturb the structure or stability of the S-cone photopigment [6], therefore strongly affecting the performance of the S-cone photoreceptors. Further complications lie in the fact that any small variations in λ_{\max} would be difficult to dissociate from individual variations in macular pigment and lens density measured psychophysically or in vivo.

Future Directions

An important issue in color vision is how spectral differences between two photopigments actually translate to color discrimination performance and how other retinal and cortical factors (such as individual variability in postreceptoral signal gain that precedes the formation of color-opponent channels) act upon the photopigment

differences and influence discrimination. To distinguish anomalous trichromacy and dichromacy it is not enough to look at λ_{\max} differences. There are other factors such as optical density and differences in postreceptoral amplification of cone signals that may explain the observed phenotype. When color vision depends on subtle differences between the two pigments, as in color vision defects, relating genotype-phenotype requires consideration of genetic polymorphisms that might affect optical density as well as those that shift λ_{\max} . It has been suggested that a separation greater than ~ 20 nm would be sufficient to account for normal color vision. This implies that the majority of normal trichromats can in principle have hybrid genes encoding their visual pigments providing the λ_{\max} between L- and M-cones remains greater than 20 nm.

Changes in λ_{\max} , photoreceptor optical densities, and/or postreceptoral amplification of cone

signals appear to be the most important parameters that affect our red-green chromatic sensitivity. A complete description of the relative importance of each of these parameters in determining a subject's color discrimination performance remains a difficult task. The use of information derived from genetic analysis of pigment genes together with psychophysical data on chromatic sensitivity and modeling work could be used to understand and account for the large variability in chromatic discrimination and color matches observed in both normal and color-deficient observers.

Cross-References

- [Dichromacy](#)
- [Trichromacy](#)

References

1. Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B., Hogness, D.S.: Molecular genetics of inherited variation in human color vision. *Science* **232**(4747), 203–210 (1986)
2. Sharpe, L.T., Stockman, A., Jagle, H., Nathans, J.: Opsin genes, cone photopigments, color vision, and color blindness. In: Gegenfurtner, K.R., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 3–52. Cambridge University Press, Cambridge (1999)
3. Bowmaker, J.K.: Visual pigments and molecular genetics of color blindness. *News Physiol. Sci.* **13**, 63–69 (1998)
4. Asenjo, A.B., Rim, J., Oprian, D.D.: Molecular determinants of human red/green color discrimination. *Neuron* **12**(5), 1131–1138 (1994)
5. Neitz, M., Neitz, J., Jacobs, G.H.: Spectral tuning of pigments underlying red-green color vision. *Science* **252**(5008), 971–974 (1991)
6. Weitz, C.J., Went, L.N., Nathans, J.: Human tritanopia associated with a third amino acid substitution in the blue-sensitive visual pigment. *Am. J. Hum. Genet.* **51**(2), 444–446 (1992)
7. Sharpe, L.T., Stockman, A., Jagle, H., Knau, H., Klausen, G., Reitner, A., et al.: Red, green, and red-green hybrid pigments in the human retina: correlations between deduced protein sequences and psychophysically measured spectral sensitivities. *J. Neurosci.* **18**(23), 10053–10069 (1998)
8. Neitz, J., Neitz, M., He, J.C., Shevell, S.K.: Trichromatic color vision with only two spectrally distinct photopigments. *Nat. Neurosci.* **2**(10), 884–888 (1999)
9. Wyszecki, G., Stiles, W.S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, New York (1982)
10. Nathans, J.: The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments. *Neuron* **24**(2), 299–312 (1999)
11. Neitz, M., Neitz, J.: Numbers and ratios of visual pigment genes for normal red-green color vision. *Science* **267**(5200), 1013–1016 (1995)
12. Neitz, J., Neitz, M., Kainz, P.M.: Visual pigment gene structure and the severity of color vision defects. *Science* **274**(5288), 801–804 (1996)
13. Hayashi, T., Motulsky, A.G., Deeb, S.S.: Position of a 'green-red' hybrid gene in the visual pigment array determines colour-vision phenotype. *Nat. Genet.* **22**(1), 90–93 (1999)
14. Crognale, M.A., Teller, D.Y., Yamaguchi, T., Motulsky, A.G., Deeb, S.S.: Analysis of red/green color discrimination in subjects with a single X-linked photopigment gene. *Vision Res.* **39**(4), 707–719 (1999)

Photosensitive Retinal Ganglion Cells

- [Melanopsin Retinal Ganglion Cells](#)

Photosensors

- [Photodetector](#)

Physiologic Scotoma

- [Blind Spot](#)

Pigment, Ceramic

Guillermo Monros

Departamento de Química Inorgánica y Orgánica,
Universitat Jaume I, Edifici Científico-Tècnic,
Castelló, Spain

Synonyms

[Colorant](#); [Dye](#); [Paint](#); [Stain](#); [Tint](#)

Definition

A ceramic pigment is usually a metal transition complex oxide obtained by a calcination process which shows three main characteristics: (a) thermal stability, maintaining its identity when temperature increases; (b) chemical stability, maintaining its identity when fired with glazes or ceramic matrices; and (c) high tinting strength when dispersed and fired with glazes or ceramic matrices. Other characteristics as high dispersability in vehicles, high refractive index (in order to avoid transparency and to increase its tinting strength), acid and alkali resistance, and low abrasive strength are also suitable.

The main coloring methods of ceramics are based on dyes or pigments. Strictly a dye or soluble colorant is a colored substance that interacts with the matrix to which it is applied and usually is soluble in the media of application called matrix or substrate. Conversely, ideally a pigment or stain is a colored substance insoluble in the pigmented matrix, is stable in the matrix (it does not interact with the matrix), and, in the case of ceramic pigments used in stoneware or glazes, is stable when temperature increases (thermal stability). Obviously a pigment must have a high tinting strength relative to the materials it colors. As a general rule, ceramic stains and ceramic pigments look pretty much the same before and after firing, but not dyes, based on raw oxides as well as salts such as carbonates or nitrates, because they decompose in firing and dissolve in the glazes or stoneware (Fig. 1a) [1a].

Industrial Ceramic Pigments

In the industry there are several steps for the production of industrial ceramic pigments:

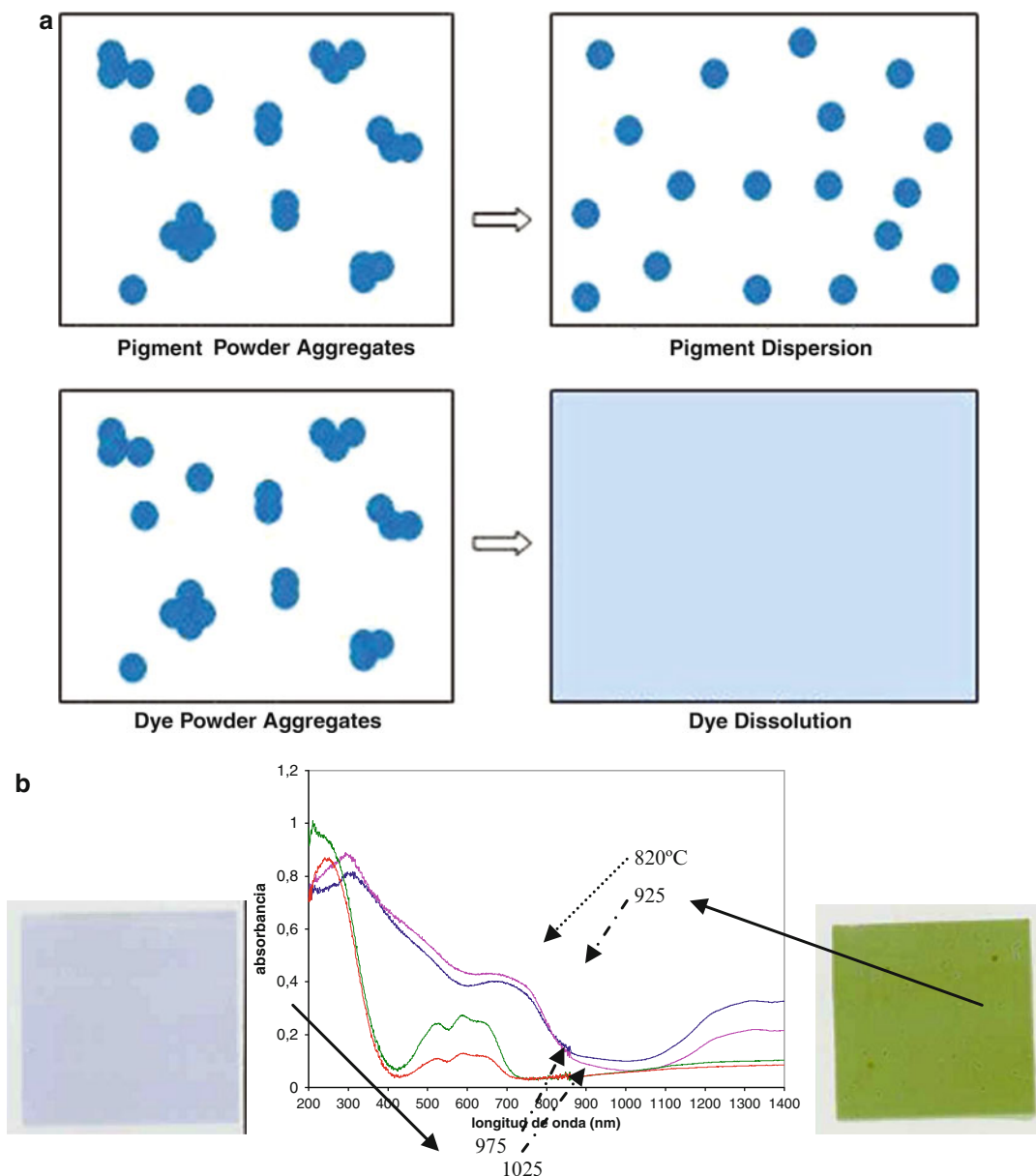
1. Grinding. The raw materials (usually oxides or carbonates) are carefully mixed in mills such as ball mills often with mineralizer agents (inorganic salts as NaCl, NaF, and KNO_3) in order to facilitate the pigment crystallization.
2. Calcination. The mixture is then calcined in either batch kilns or continuous calciners.

3. Washing. After calcination, the powders are washed to eliminate residual salts from mineralizers and other leaching components in order to avoid future firing defects and environmental affections.
4. Micronization. The mixture is then ground to the necessary fineness in mills. Micronizers and jet mills are used to break agglomerates.
5. Tinting strength control. The final production step involves careful control of the color tone.

Because these pigments are formed at high temperatures, they generally offer superb thermal stability and are relatively inert. This results in excellent weathering and light fastness properties. Most of these pigments have superior acid and alkali resistance. They are nonmigrating and nonbleeding in nature and do not interact with substrates. The principal disadvantage of ceramic pigments is their low tinting strength. In addition, some are relatively high in cost. This is particularly true with cobalt-containing pigments. Some of these pigments are difficult to disperse. However, the recent development of easily dispersed ceramic pigments should eliminate this problem. A final concern is the inherent hardness of these pigments. Their hardness can lead to processing system damage through abrasion. When using ceramic pigments, processing system components designed for use with abrasive materials should be considered [1b].

Overview

Stable fired pigments solve some of the problems found in using dyes such as: (a) the use of hazardous raw materials, because many of the unstable oxides or salts used as direct dyes are soluble or toxic, its calcination, by combining these elements along with clays, silica, alumina, etc., makes them stable giving a safety manipulation, is the case of vanadium pentoxide a toxic substance that is safely used in the zirconium vanadium blue zircon ceramic pigment, (b) volatilization in the kiln which is colored by the fumes such as it occurs using vanadium oxide



Pigment, Ceramic, Fig. 1 (a) Ideal behavior of interaction of pigments (dispersion) and dyes (dissolution) with the matrix. (b) Color evolution with the firing temperature of screen printing (90 threads/cm) deposition of a Co-Ti

ink (1 w.% in the organic polyol media) under a glazed CaO-ZnO-SiO₂ monoporosa tile (previously fired at 1,080 °C)

or chromium oxide, and (c) the own stability of the color avoiding chemical reactions of the dye with the components of the glaze, for example, eskolaite Cr₂O₃, used as natural green dye, reacts with tin oxide and silica of the glazes producing

malayaite Cr-CaSnSiO₅ pink precipitates, and then a pink undesired coloration is given, or reacts with zinc oxide of the glaze to crystallize ZnCr₂O₄ spinel, imparting undesired blue-brown color; the use of the cobalt-zinc-alumina-chromite blue-

green pigment gives a trustworthy stable range of color from green to blue.

How to give trustworthy and safety color to ceramics is the main proposal of pigment production because adding color to our ceramics is needed to produce ceramic art and it cannot be a tricky proposition. Working with paints, the color you put on your piece is the color of the final paint, but in ceramics the fire also builds the color. Pigment research is addressed to obtain new ceramic pigments in order to reach three main purposes:

- (a) High thermal stability pigments, addressed to new materials as porcelainized stoneware and correlated use of ink-jet printing techniques, that require obtaining adequate particle size of pigments or precursor solutions that produce the pigment “in situ,” during thermal processing of substrate.
- (b) High chemical resistance pigments, adapted to new glazes and ceramic matrices avoiding particle pigment solubilization or degradation when it reacts in ceramic processing.
- (c) Eco-friendly pigments, with low environmental affection evaluated on its complete life cycle: raw material preparation, transport, processing, product distribution, and waste management.
- (d) Pigments for industrial digital printing (ink-jet). Ink-jet decoration is a relatively recent methodology. Elmquist designed the first practice application of CIJ (continuous ink-jet) in 1951 for tape recorder. Winston developed the first teletype printer by CIJ, and in 1968 the first commercial printer 9600 from AB Dick appeared. A continuous flow of electrical conductive ink drops is thrown to substrate and selectively dispersed by an electric field in the CIJ ancient technology. The DOD (drop on demand) ink-jet was developed later in order to produce only the demanded drops. Firstly, in TIJ (thermal ink-jet) technique, the drops were produced from pressure of bubbles generated into the nozzles by a local heating mechanism. Now, in the PIJ (piezoelectric ink-jet) technique, the drops are produced by pressure due to the

deformation when an electric field is applied to a piezoelectric piece disposed in the nozzles.

Digital decoration technology is widely developed in other materials (textile, paper, etc.), basically using TIJ technique. It has been claimed for surface ceramic tile decoration due to several advantages based on the fact of absence of contact between the applicator and the decorated surface:

- (i) High resolution of images. The ink-jet technique allows to throw 2,000–5,000 drops/cm in a strictly controlled way. Using serigraphy, chalcography, or flexography methods, the resolution is very limited. Therefore photographic quality images could be carefully reproduced by ink-jet digital monitored application.
- (ii) Low raw strength of tile and short enameling line required. Due to the absence of contact applicator surface, the raw strength is limited only to the adequate manipulation of raw tile, and then the thickness of piece can be reduced. The number of decoration operations is reduced sometimes at one because it can be applied four colors in a selective way at the same time; likewise the reduced weight of ink deposited and relatively low water concentration reduce the drying time between successive applications, and therefore the length of industrial enameling line is drastically reduced.
- (iii) The topography of the tile surface is not limited and relief surfaces can be easily decorated.
- (iv) Reduction of necessary ink (1 g/m^2), fully recycling and washing operations eliminated.
- (v) Simplicity of the decoration process because it can be limited to only one decoration step using the simultaneous four-color application CMYK (cyan, magenta, yellow, black).

The main difficulty for ink-jet application in ceramic tile industry is the ink that must show specific properties, basically:

- (i) High stability, because the precipitation, agglomeration, or viscosity changes can obstruct the nozzles.
- (ii) High color strength is necessary due to reduced weight of ink deposited by dropping; on contrary, diffused and weak colors are obtained and poor decoration is produced.
- (iii) Neutral pH is required in order to prevent corrosive effects on the nozzles, but conductivity is not required in both PIJ or TIJ technologies; contrarily in ancient CIJ technique, the electric charge capacity of drops was necessary for their selective dispersion by an electric field.

There are three main families of digital inks: (a) soluble salts such as cobalt nitrate (*CYAN*), mixture of chromium nitrate and iron nitrate (*MAGENTA*), mixture of antimony nitrate, nickel and stabilized titania (*YELLOW*), and mixtures of cobalt and iron nitrates (*BLACK*) are used for producing “in situ,” when decorated ceramics are fired, classical ceramic pigments such as Co_2SiO_4 olivine blue pigment, $\text{Fe}(\text{Fe,Cr})_2\text{O}_4$ iron chromite brown spinel, $(\text{Ti,Ni,Sb})\text{O}_2$ chromium-antimony yellow rutile, and CoFe_2O_4 black spinel, respectively; (b) colloidal or ultramicrosized classical pigments particles dispersed using dispersant agents; and (c) nanoparticles of classical ceramic pigments, produced by “soft chemistry” such as polyol or sol-gel routes, or metals absorbed on stable sol particles of oxohydroxy compounds of Al, Ti, Sn or Zr (purple of Cassius and its analogous with Cu).

Since the nineteenth century, needs on color stability, glaze resistance, and sure management lead to stabilized fired pigments. On these pigmenting systems, chromophore agents (usually first row transition heavy metals and p metals) are inertized in high-thermal and low-solubility ceramic matrices.

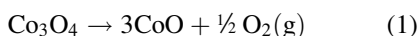
The main mechanisms of origin of color based on the stabilization-inertization of chromophores are the following:

- (a) Structural mechanism. In this case the chromophore cations (usually transition metals that absorb, as it is discussed below, selected visible wavelengths by d-d electron transitions) are forming the ceramic network such as Co(II) in Co_2SiO_4 olivine. In other structural cases, a solid solution is obtained: pigmenting cations substitute structural cations in ceramic matrix as V^{4+} substitutes Zr^{4+} in ZrSiO_4 zirconium vanadium blue zircon (a cyan pigment also named Turkish blue zircon or turquoise zircon) formulated by CPMA as $(\text{V,Zr})\text{SiO}_4$. The above-discussed classical pigments $\text{Fe}(\text{Fe,Cr})_2\text{O}_4$ (iron chromite brown spinel with iron in solid solution into chromite network) and $(\text{Ti,Ni,Sb})\text{O}_2$ (chromium-antimony yellow rutile with Sb^{+5} and Cr^{3+} in solid solution into rutile lattice), are also solid solutions.
- (b) Inclusion mechanisms. Colored nanoparticles protected into high stable host crystals (e.g. Fe_2O_3 hematite into zircon crystal particles that gives the pink coral of iron in the zircon ceramic pigment) are the origin of color in this case.
- (c) Mordant mechanism. The origin of color is in this case a metal-based dye stabilized by a mordant compound. The final color depends on the mordant used, due to the formation of metal complexes that cause a change in the molecular orbital energies and hence a shift in UV-Vis absorption bands; e.g. Au in $\text{Sn}(\text{OH})_4$ mordant substrate on Cassius purple pigment [1a].

In agreement with above-described purposes of research on new ceramic pigments, three main limitations can be pointed out: (a) thermal and chemical stability, (b) structure base, and (c) environmental safety. These limitations are described and illustrated with three examples.

Thermal and Chemical Stability: Cobalt Inks

Usually cobalt oxide Co_3O_4 used as cobalt precursor raw material is a mixed valence oxide of Co^{2+} and Co^{3+} with spinel structure $\text{Co}^{\text{II}}\text{Co}_2^{\text{III}}\text{O}_4$ which decompose on firing by reducing the unstable Co^{3+} to Co^{2+} producing a profuse pinhole in the glazes if the time of firing is not sufficient due to oxygen release:



The Co^{2+} ion dissolves in ceramic glazes and imparts an intense blue color to ceramic glazes up to 0.05 w.% and could be used in novel dye ink-jet application in ceramics. This color intensifies in the presence of ZnO , due to the crystallization of a solid solution of Co^{2+} in willemite ($\text{Zn}, \text{Co})_2(\text{SiO}_4)_2$, or in the presence of BaO which precipitates cobalt-celsian solid solution ($\text{Ba}, \text{Co})(\text{SiO}_4)_2$. In glazes containing MgO and TiO_2 , cobalt gives green colorations, due to crystallization of green cobalt-ilmenite CoTiO_3 more stable in boron glazes [1c].

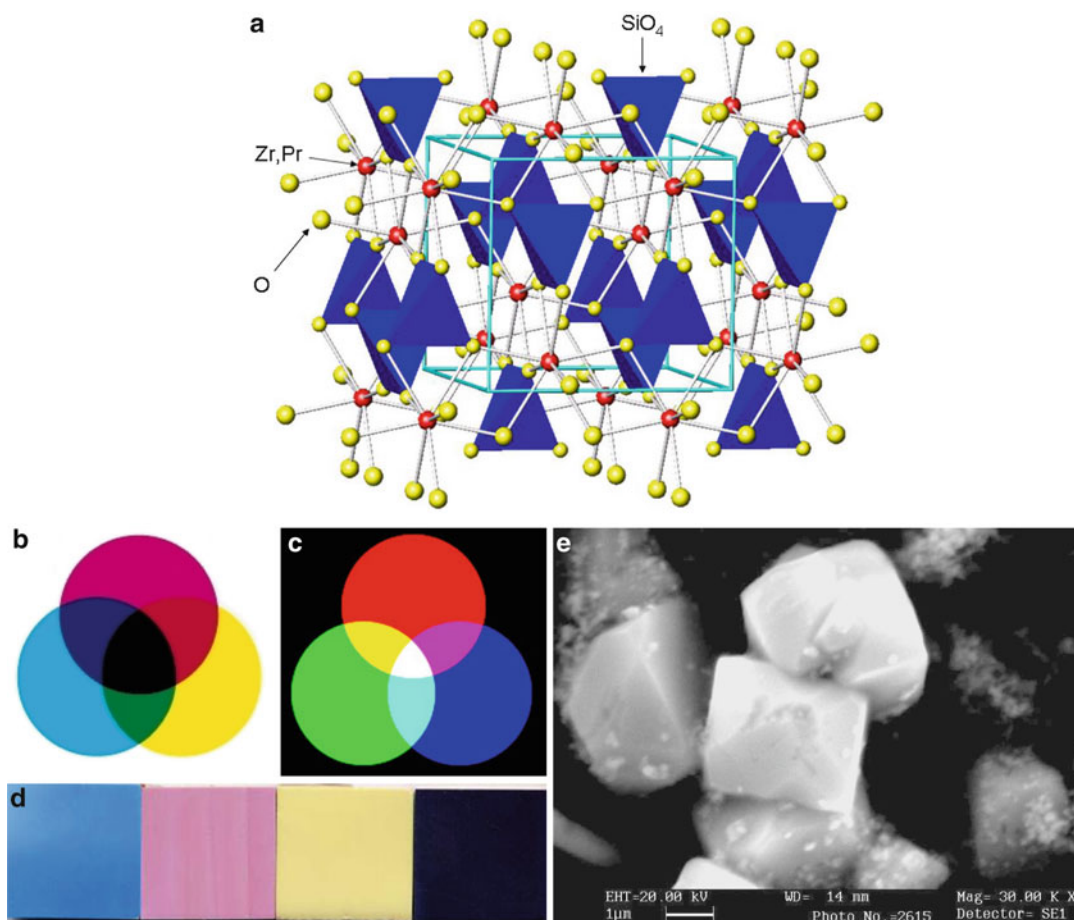
The color evolution with the firing temperature of screen printing (90 threads/cm) deposition of a Co-Ti ink (1 w.% in the organic polyol media) over a glazed “monoporosa” tile (previously fired a 1,080 °C) is shown in Fig. 1b. The green coloration at low temperatures (820, 925 °C) is associated to cobalt-ilmenite CoTiO_3 crystallizations with Co^{2+} in octahedral coordination (absorption bands at 350 nm in the UV range, 700 nm in the visible red wavelength, and 1,200 nm in the IR range, in the spectral and color rendering of Fig. 1). When temperature increases, the cobalt-ilmenite CoTiO_3 decomposes and Co^{2+} solubilizes in the glaze giving intense blue color associated to Co^{2+} in tetrahedral coordination (absorption bands at 530 nm green, 590 nm orange, and 650 nm red and high reflectivity at 430 nm blue wavelength). At low temperature the color is associated to a pigment mechanism (particles of ilmenite crystallization on the glaze surface), but at high temperatures,

the color is due to a dye mechanism (cobalt dissolution in glaze).

Structure Base: The Classical Pigment Palette

A useful colorant classification of the chemicals used as ceramic pigments was reached in 1977 from the requirements of the Toxic Substances Control Act, 94–469 US law, which includes the whole chemical substances used in the USA whether they are toxic or not. The Dry Colors Manufacturer’s Association (DCMA) commends to their Metal Oxide and Ceramic Color Subcommittee of DCMA Ecology Committee the classification of the commercial ceramic pigments giving a standard terminology. This DCMA committee applied a chemical-structural criteria for the classification of ceramic pigments in 14 structural families: (1) baddeleyite [1 ceramic pigment], (2) borate [1], (3) corundum-hematite [4], (4) garnet [1], (5) olivine [2], (6) periclase [1], (7) phenacite-willemite [1], (8) phosphate [2], (9) priderite [1], (10) pyrochlore [1], (11) rutile-cassiterite [11], (12) sphene [1], (13) spinel [19], and (14) zircon [3 ceramic pigments].

In 2010 the Color Pigments Manufacturers Association (CPMA) actualized the classification and chemical description of the complex inorganic color pigments in the fourth edition [2]. CPMA describes the industrial ceramic pigment by a code such as CPMA 1-01-4. The first number on the code is the corresponding number of the structure above discussed (in this case 1=baddeleyite), the second set is the CPMA category number identifying each pigment within a given crystal class (in this case 1=first ceramic pigment on the CPMA list), and the third set of numbers identifies the color of the pigment as follows: (1) violet and red-blue; (2) blue and blue-green; (3) green; (4) yellow and primrose; (5) pink, orchid, coral, and peach; (6) buff; (7) brown; (8) gray; and (9) black (in the above case 4=yellow). Finally, CPMA also gives a classification based upon the predominant use of these



Pigment, Ceramic, Fig. 2 (a) ZrSiO_4 structure from CPMA classification, S.G. $I_4/\bar{1}$ with D_{2d} symmetry for SiO_4 group and dodecahedral D_{2d} for ZrO_8 , (b) CMYK subtractive color model, (c) RGB additive color model, (d) the CMYK zircon palette (cyan $(\text{Zr,V})\text{SiO}_4$ DCMA 14-42-

2, pink $(\text{Zr,Fe})\text{SiO}_4$ DCMA 14-44-5, yellow $(\text{Zr,Pr})\text{SiO}_4$ DCMA 14-43-4) including a Cr,Co-ferrite as black component, all 3 % enameled in monoporosa glaze, (e) particles of black spinel pigment by SEM

pigments; three use categories were adopted for guidance. Category A deals with pigments suspended in glass matrices which require the highest degree of heat stability and chemical resistance to withstand the attack of molten glass. Category B deals with pigments suspended in plastics and other polymers which require only moderate heat stability. Category C deals with pigments suspended in liquid vehicles which require little or no heat stability.

Among the more successful ceramic pigments which gives a complete CMYK palette of color, the pigments based on zircon structure (number

XIV in CPMA classification) shown in Fig. 2a and discovered by C. A. Seabright in 1948 stand out.

(a) CYAN based on the vanadium blue zircon V-ZrSiO_4 (CPMA number 14-42-2). CPMA describes this pigment as an inorganic pigment and as a reaction product of high-temperature calcination in which zirconium (IV) oxide, silicon (IV) oxide, and vanadium (IV) oxide in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of zircon. Basic chemical formula: $(\text{Zr,V})\text{SiO}_4$. Its composition may

include any one or a combination of the modifiers alkali or alkaline earth halides. Use category A. Exceptionally suitable for coloring ceramic glazes and clay bodies. Not generally used in porcelain enamels.

- (b) **MAGENTA** based on the pink coral Fe-ZrSiO_4 (CPMA number 14-44-5). CPMA describes the zirconium iron pink (peach, coral) zircon, as an inorganic pigment, as a reaction product of high-temperature calcination in which zirconium (IV) oxide, silicon (IV) oxide, and iron (III) oxide in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of zircon. Basic chemical formula: $(\text{Zr,Fe})\text{SiO}_4$. Its composition may include any one or a combination of the modifiers alkali or alkaline earth halides. Use category A. Exceptionally suitable for coloring ceramic glazes. Not generally used in clay bodies or porcelain enamels.
- (c) **YELLOW** based on the yellow of praseodymium Pr-ZrSiO_4 (CPMA number 14-43-4). CPMA describes the zirconium praseodymium yellow zircon, as an inorganic pigment, as a reaction product of high-temperature calcination in which zirconium (IV) oxide, silicon (IV) oxide, and praseodymium (III, IV) oxide (Pr_6O_{11}) in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of zircon. Basic chemical formula: $(\text{Zr,Pr})\text{SiO}_4$. Its composition may include any one or a combination of the modifiers alkali or alkaline earth halides. Use category A. Exceptionally suitable for coloring ceramic glazes and clay bodies. Not generally used in porcelain enamels.
- (d) **BLACK** usually obtained from spinel structure in CPMA classification such as the ferrite $(\text{Mg}_{0.5}\text{Fe}_{0.5})(\text{Co}_{0.5}\text{Fe}_{1.4}\text{Cr}_{0.1})\text{O}_4$ shown in Fig. 2d. CPMA describes the iron cobalt chromite black spinel as an inorganic pigment and as a reaction product of high-temperature calcination in which iron (II) oxide, cobalt (II) oxide, iron (III) oxide, and chromium (III) oxide in varying amounts are homogeneously and ionically interdiffused to form a crystalline matrix of spinel. Basic chemical

formula: $(\text{Co,Fe})(\text{Fe,Cr})_2\text{O}_4$. Its composition may include any one or a combination of the modifiers Al_2O_3 , B_2O_3 , CuO , MnO , NiO , or SiO_2 . Use categories A, B, and C. Predominantly used for coloring category A substrates.

Adequate mixtures of above-described pigments may produce all range of visible colors, using the subtractive color vision model CMYK (where C=cyan, M=magenta, Y=yellow, and K=key associated to a mixture of C+M+Y which gives the black color) shown in Fig. 2b, in opposition to the additive color vision model RGB (R=red, G=green, B=blue) shown in Fig. 2c. Really the black color is not necessary because it could be built mixing CMY, but in ink and paints, it is used in order to simplify and to cheapen the printing manufacture. A detailed study of the CMY zircon pigments can be found in several doctoral theses [4, 5]. Really the color given by CMY zircon pigments is influenced by temperature and glaze or ceramic matrix; therefore usually a wider color palette must be used on pigmenting ceramics.

A total of 44 chemical substances are listed in the CPMA classification referred exclusively to industrial ceramic pigments. But the world of pigments is wider and, for example, on the industrial classification, the main red colors do not appear: the unstable purple of Cassius discovered by Johann Rudolph Glauber in 1659 and the sulfoselenide dated on 1909 and classified as harmful substance due to cadmium presence. The reddish ceramic pigments have been during centuries the driving force in the ceramic pigment field research [6]. The investigation on new host structures for pigmenting also has received attention, but usually, searching the red shades.

As example, a new color alternative to classical zircon palette can be obtained from the perovskite, a new and versatile lattice for ceramic pigments not considered on CPMA classification of industrial pigments. Neodymium perovskite pigments have been obtained by nonconventional methods [7–9]. Red, blue, green, and gray ceramic pigments based on FeNdO_3 perovskite (unmineralized sample or using $\text{BaF}_2+\text{MgF}_2$ flux

agent for red and blue pigments, respectively) and CrNdO_3 (unmineralized sample or using $\text{BaF}_2+\text{MgF}_2$ flux agent for green and gray pigments, respectively) can be obtained by ammonia coprecipitation from a solution of $\text{NdCl}_3 \cdot 6\text{H}_2\text{O}$, $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, $\text{CrCl}_3 \cdot 6\text{H}_2\text{O}$ salts in water media.

Dried powders were successively fired at 1,000, 1,100, and 1,200 °C during 6 h of soaking time, and the obtained powders were 5 % enameled in a conventional glaze (1,050 °C). Characterizations of samples are shown in Figs. 3 and 4:

- (a) Powder optical lens view ($\times 40$) (Fig. 3a) shows red, blue, green, and gray color of the respective perovskites.
- (b) XRD (X-Ray Diffraction) analysis carried out on a diffractometer, using Cu K_α radiation, 20–70 °2 θ range, scan rate 0.05 °2 θ /s, 10 s per step, and 40 kV and 20 mA conditions, shows that perovskite is the only crystalline phase detected in CrNdO_3 samples, but residual Nd_2O_3 peaks are always observed on FeNdO_3 samples (Fig. 3b).
- (c) 5 w.% enameled samples in a conventional glaze (1,050 °C) show the visual red, blue, green, and gray color of the respective perovskite color associated to $\text{CIE}^*a^*b^*$ color measurements (Fig. 3c), which were measured following the CIE (Commission Internationale de l'Eclairage) colorimetric method using a spectrophotometer, with standard lighting D65 (natural daylight) and standard observer of 10°. On this method, L^* is a measure of brightness (100=white, 0=black) and a^* and b^* of chroma ($-a^*$ =green, $+a^*$ =red, $-b^*$ =blue, $+b^*$ =yellow).
- (d) UV-Vis-NIR (Ultraviolet-Visible-Near Infra-red) spectroscopy of enameled samples, collected using a spectrometer through diffuse reflectance technique, shows bands associated to Fe^{3+} in octahedral coordination (in FeNdO_3) or Cr^{3+} in octahedral coordination (in CrNdO_3), respectively, along with sharp bands associated to Nd^{3+} f-f transitions [12]. Samples with $\text{BaF}_2+\text{MgF}_2$ flux agent

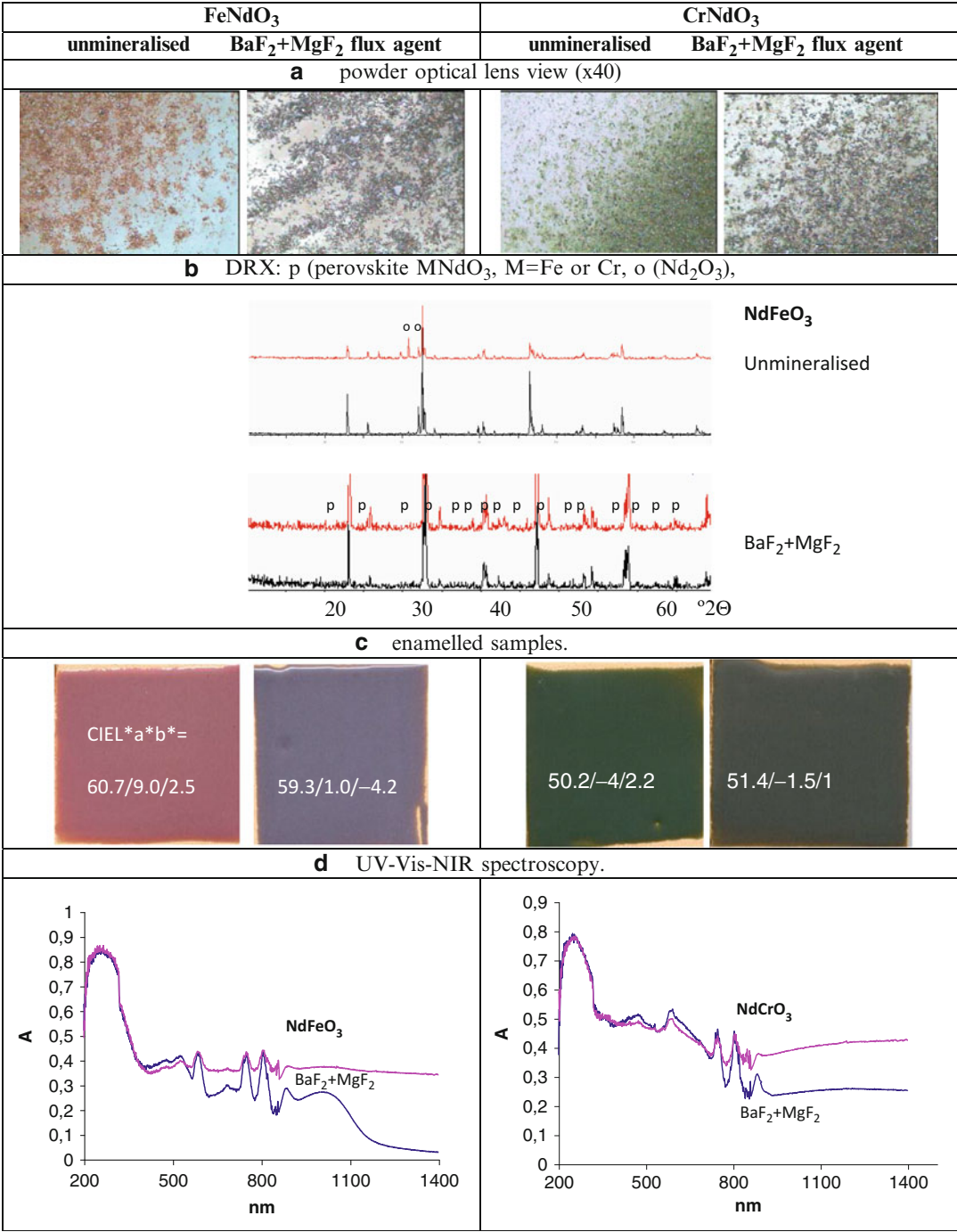
addition show similar bands to unmineralized samples, but the absorbance level increases in the Vis-NIR range producing dark colors (dark blue and black color, respectively, on Fig. 3c).

The micrographs of samples of unmineralized and $\text{BaF}_2+\text{MgF}_2$ mineralized samples fired at 1,200 °C, carried out by scanning electron microscopy (SEM), are shown in Fig. 4. Both samples show higher cubic crystalline morphology, but unmineralized sample shows higher size and heterogeneous particles that show low aggregation; in contrast $\text{BaF}_2+\text{MgF}_2$ mineralized sample shows a bimodal distribution of particle size (small 0.5 μm edge cubic particles and big 1.5 μm particles).

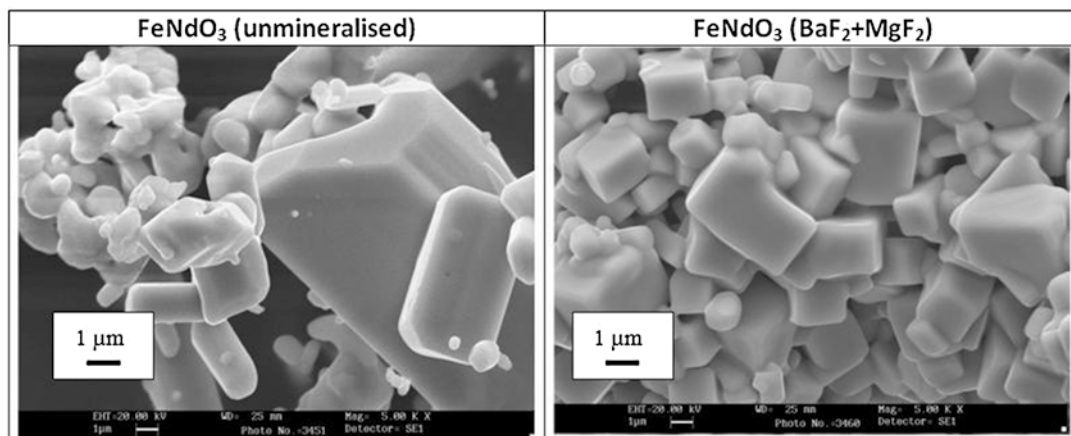
The preparation method of ceramic pigment allows obtaining new and best ceramic pigments. There are different new methods of preparing ceramic pigments that produce new microstructures and solid solutions [1, 8].

Chemical methods as alternative to solid oxide reaction can be illustrated with above-discussed CrNdO_3 perovskite. A polyol route was used from $\text{NdCl}_3 \cdot 6\text{H}_2\text{O}$ and $\text{CrCl}_3 \cdot 6\text{H}_2\text{O}$: salts were dissolved in 200 ml of DEG (diethylene glycol) to prepare 5 g. of final product, then refluxed at 150 °C during 2 h, and then charred at 500 °C/1 h and fired at 1,100 °C with soaking time of 6 h. SEM micrograph of powders charred at 500 °C shows aggregates of incipient cubes (Fig. 5c), well developed in fired sample at 1,100 °C (Fig. 5d). Powder develops, 5 w.% enameled in conventional ceramic glaze (1,080 °C), good black shades ($L^*a^*b^* = 50.5/-0.6/1.0$) without mineralizer addition (compare Figs. 5b and 3c).

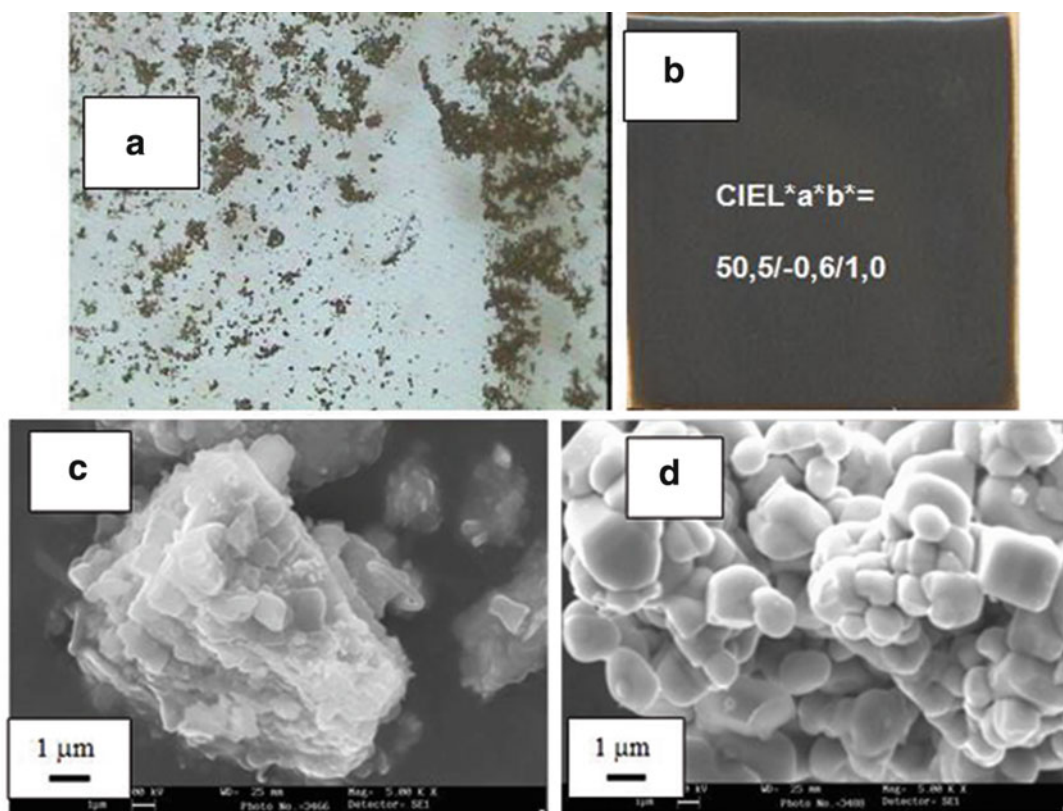
Therefore above neodymiums could be an example of alternative to cyan, pink, and black pigments to the classical zircon color palette. Finally an alternative to yellow of praseodymium in zircon could be a $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_7$ yellow pigment based on pyrochlore crystal structure (Fd-3m) which describes materials of the types $\text{A}_2\text{B}_2\text{O}_6$ and $\text{A}_2\text{B}_2\text{O}_7$, where the A and B species



Pigment, Ceramic, Fig. 3 Samples fired at 1,200 °C: (a) powder optical lens view (×40), (b) DRX, crystalline phases. p(perovskite MNdO₃, M = Fe or Cr), o(neodymium oxide Nd₂O₃), (c) enameled samples, (d) UV-Vis-NIR spectroscopy



Pigment, Ceramic, Fig. 4 SEM micrographs of samples fired at 1,200 °C



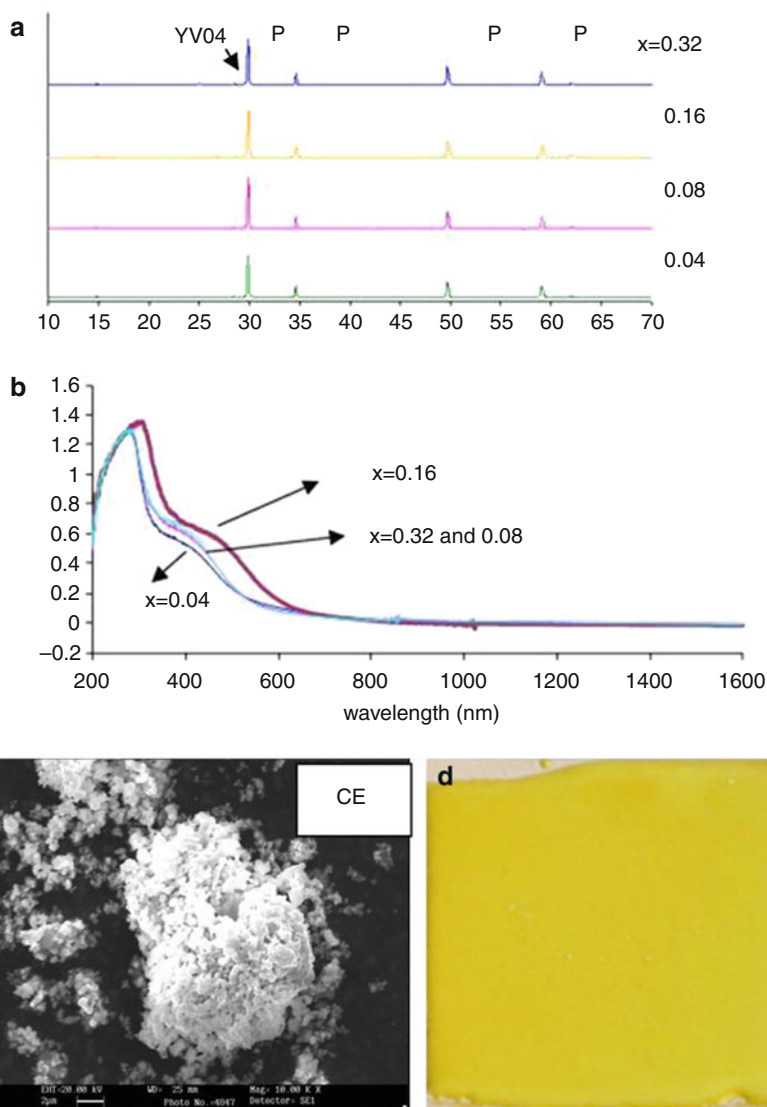
Pigment, Ceramic, Fig. 5 CrNdO₃ fired at 1,100 °C: (a) powder optical lens view (×40), (b) enamelled sample, (c) SEM micrograph of powder (500 °C), (d) SEM micrograph of powder (1,100 °C)

are generally rare-earth or transition metal species, e.g., Y₂Ti₂O₇ [9]. A yellow pigment with the pyrochlore structure Ca_xY_{2-x}V_xTi_{2-x}O₇ is known since 1993 as a substitute for the

decreasing variety of available yellow ceramic pigments due to the severe regulation of toxic lead and cadmium. The solubility limit of vanadium in this pigment was found to be 1.5 w.% as

Pigment, Ceramic,

Fig. 6 (a) XRD diffractograms of $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_{7.0,14}\text{Na}_2\text{SiF}_6$ powders, CRYSTALLINE PHASES: P = $\text{Y}_2\text{Sn}_2\text{O}_7$, S = SnO_2 , Y = Y_2O_3 . (b) UV-Vis-NIR spectra of $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_{7.0,14}\text{Na}_2\text{SiF}_6$ powders, (c) SEM micrograph of $x = 0.16$ sample, (d) 5w% $x = 0.16$ glazed sample



V_2O_5 or 0.13 as x in the above formula expression. Characterization of vanadium in the vanadium pyrochlore yellow pigment by electron spin resonance showed that the oxidation state of vanadium was V^{5+} and its yellow color mostly originated from V^{5+} substituted for Ti^{4+} [10, 11].

XRD (X-Ray Diffraction) results (Fig. 6a) indicate the crystallization of $\text{Y}_2\text{Sn}_2\text{O}_7$ as the only crystalline phase in all samples except in

$x = 0.32$ that shows very weak peaks associated to YVO_4 along with perovskite peaks. Spectral and color rendering of samples by UV-Vis-NIR reflectance diffuse spectroscopy (Fig. 6b) shows a sharp band centered at 250 nm associated to the $\text{Sn}^{4+}-\text{O}^{2-}$ band transfer and an additional band centered at 500 nm responsible of yellow color and associated to $\text{V}^{5+}-\text{O}^{2-}$ band transfer. The maximum of absorbance is observed for $x = 0.16$ sample. CIEL*a*b* results summarized in

Pigment, Ceramic, Table 1 CIEL a^*b^* colorimetric coordinates for the $\text{Ca}_x\text{Y}_{2-x}\text{V}_x\text{Sn}_{2-x}\text{O}_7$ yellow pigments

Sample (x)	L*	a*	b*
0.04	84.3	1.2	26.6
0.08	87.2	0.2	30.8
0.16	85.7	1.1	34.5
0.32	85.8	0.8	33.5

Table 1 indicate that the yellow b^* parameter increases with x until $x = 0.16$ and decreases for $x = 0.32$ indicating a solubility limit of vanadium in this pigment around $x = 0.16$. The enameled powders were 5 w%, glazed in a CaO-ZnO-SiO_2 (1,080 °C) single firing glaze but do not produce color; in $\text{Na}_2\text{O-CaO-PbO-SiO}_2$ (1,000 °C) double firing glaze, the color of all samples is similar (CIEL $*a^*b^* = 78/5/35$) (Fig. 6d).

In order to increase the reactivity of the system, the ceramic optimized composition $x = 0.16$ was prepared by ammonia coprecipitation (CO) and Pechini route (CI). In both routes $\text{Y}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$, $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$, $\text{SnCl}_2 \cdot 2\text{H}_2\text{O}$, and $\text{VOSO}_4 \cdot 8\text{H}_2\text{O}$ (ALDRICH) were used as precursors. In the CO route the precursors were solved in water (250 ml for 5 g of final product), then concentrated ammonia was dropped at 70 °C in continuous stirring until pH = 8; finally powder was obtained by drying in oven at 110 °C. In CI route, a molar ratio metallic cations/ethylene glycol/citric acid = 1:1:1 was used. Firstly precursors were solved in water (250 ml for 5 g of final product) and citric acid was solved at 70 °C in vigorous stirring, and then ethylene glycol was added and maintained 8 h for esterification. The obtained ester was dried at 110 °C and submitted to charring treatment at 250 °C. Na_2SiF_6 flux agent was added to CO or CI powders by manual mixture in an agatha mortar using acetone media. XRD results obtained at 1,000 °C/6 h for both CI and CE show peaks associated to pyrochlore along with unreacted SnO_2 and Y_2O_3 peaks; CO sample only shows weak peaks associated to SnO_2 along with pyrochlore. At 1,200 °C all samples show pyrochlore as the only crystalline phase detected. CO at 1,200 °C shows the best pigmenting results in double firing glaze ($L^*a^*b^* = 80.0/5.0/43.7$) than CE (79.3/2.6/35.4) and CI (81.7/0.1/31.8).

SEM micrographs of powders fired at 1,200 °C indicate the presence of aggregates of submicrometric particles in all samples (Fig. 6c). Size of aggregates in CE and CO sample (2–6 μm) is similar but more compact in CO sample and higher than in citrate powder (1–4 μm). Also size of particles forming aggregates is higher for CE (500 nm) than for CO (300 nm) and higher than CI powder (150 nm). Likewise SEM-EDX (Scanning Electron Microscopy-Energy Dispersive X-Ray Spectroscopy) microstructural analysis of powders fired at 1,200 °C (not shown) indicates that codopants, vanadium, and calcium present a homogeneous distribution on samples. However the global content of calcium is higher than vanadium, probably due to vanadium evaporation during firing. Likewise the content of vanadium is slightly higher in CO and CE sample than in CI sample in agreement with color intensity observed on glazed samples.

Environmental Safety: Ecotoxicity Evaluation

The environmental safety using ceramic pigments is an important limitation that gives the severe regulation of toxic components such as lead or cadmium. It is necessary to check the safety of the pigments. The leaching test is usually used for determining the ecotoxicity of solids. For example, the European leaching test was performed at a ratio of liquid to solid sample $L/S = 16 \text{ L/kg}$. A mixture of 10 g solid sample and 160 mL distilled water was combined in borosilicate glass bottles and then agitated for 24 h at 10 rpm with a rotary agitator. The leachate was then collected by filtration through a 0.45 μm membrane filter. The toxicity of the leachates is evaluated using a battery of bioassays including the photobacterium *Vibrio fischeri* or *Photobacter phosphoreum* (Microtox test) and the crustaceans *Daphnia magna*. Microtox test was based on the bioluminescence measurement of the marine bacteria *V. fischeri*, within exposure time to leachates of 15 min., using the DeltaTox PS1 Analyzer.

The Microtox ecotoxicity test results for the above-discussed pigments are shown in Tables 2

Pigment, Ceramic, Table 2 EC50 Microtox results for pyrochlore of tin pigments

Sample	EC50 (ppm)
Ca _{0,16} Y _{1,84} V _{0,16} Sn _{1,84} O ₇ CE 1,200 °C/6 h)	296,400
Ca _{0,16} Y _{1,84} V _{0,16} Sn _{1,84} O ₇ SONO 1,200 °C/6 h	65,639

Pigment, Ceramic, Table 3 EC50 Microtox results for perovskite of neodymium pigments

Sample	EC50 (ppm)
FeNdO ₃	86,664
2 %BaF ₂ +8%MgF ₂ SONO fired at 1,100 °C/3 h	54,398
CrNdO ₃	
2 %BaF ₂ +8 %MgF ₂ SONO fired at 1,100 °C/3 h	
(Mg _{0,5} Fe _{0,5})(Co _{0,5} Fe _{1,4} Cr _{0,1})O ₄ SONO fired at 1,000 °C/3 h	2,336

and 3. In the case of the yellow of the pyrochlore Ca_{0,16}Y_{1,84}V_{0,16}Sn_{1,84}O₇, both unmineralized ceramic CE and sonocoprecipitated SONO powders fired at 1,200 °C/6 h are compared in Table 2. In the SONO method, the chloride precursors were dissolved in water/diethylene glycol = 1:1 v/v media and were coprecipitated in ultrasonic bath by dropping 12 % ammonia solution until pH = 8. In agreement to European laws, the limit of EC50 (equivalent concentration of leached that decreases the bioluminescence of *V. fischeri* to 50 %) is 3,000 ppm. Therefore the pigment accomplishes the European regulations.

In the case of the yellow of the neodymiums above described (violet of FeNdO₃ with 2% BaF₂+8%MgF₂ SONO fired at 1,100 °C/3 h and black of CrNdO₃ with 2%BaF₂+8%MgF₂SONO fired at 1,100 °C/3 h), the Microtox results are compared with the reference black of spinel (Mg_{0,5}Fe_{0,5})(Co_{0,5}Fe_{1,4}Cr_{0,1})O₄ SONO fired at 1,000 °C/3 h) in Table 3. In agreement with European limit of EC50 the perovskite of neodymium pigments accomplish the European regulations but not the spinel black due to released chromium. The pigments based on rare-earth components usually give eco-friendly pigments [12].

Acknowledgments Authors acknowledge the financial support given by FUNDACION CAJA CASTELLÓN-UIJ, P1-1B2010-09 project.

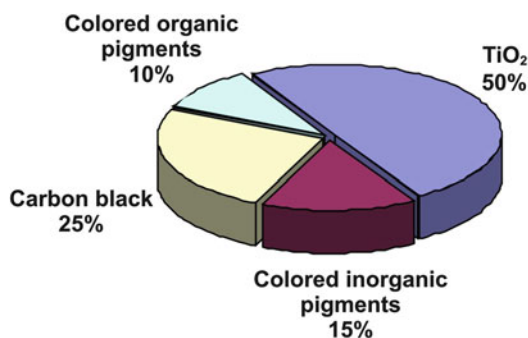
Cross-References

- CIELAB
- Color Vision Testing
- Colorant, Natural
- Dye, Ink-Jet

References

- (a) Monrós, G., Badenes, J.A., García, A., Tena, M.A.: El color de la Cerámica. Nuevos mecanismos en pigmentos para los nuevos procesados de la industria cerámica, pp. 146–179. Edited by Universitat Jaume I, ISBN 84-8021-449-X (2003). (b) R. A. Eppler Selecting Ceramic Pigments. Am. Ceram. Soc. Bull. **66**(11), 1600–1604 (1987) (c) Bucher, W., Schliebs, R., Winter, G., Buchelin, K.H.: In: Industrial Inorganic Chemistry, VCH, Berlin (1989)
- (a) CPMA: Classification and chemical description of the complex inorganic color pigments, 4th edn. Dry Color Manufacturers Association, Alexandria (2010) (b) DCMA: Safe Handling of Color Pigments, Color Pigments Manufacturers Association Inc., 300 North Washington Street, Suite 102, Alexandria (1993)
- Seabright, C.A.: Ceramic Pigments, US Patent, 2, 407-441 (1948)
- (a) Monrós, G.: Soluciones sólidas V-ZrSiO₄, Tesis Doctoral, Universidad de Valencia (1991) (b) Lahuerta, J.: Colorantes Cerámicos Rojos Basados en Fe₂O₃, Tesis Doctoral, Universidad de Valencia (1993)
- (a) Badenes, J.: Estudio de los sistemas Pr-ZrSiO₄ y Pr-ZrO₂, Tesis Doctoral, Universidad Jaume I de Castellón (2000) (b) Calbo, J.: Desarrollo de ecopigmentos negros de espinela dopada alternativos a los ferritos tradicionales mediante presión y procesados sol-gel, Tesis Doctoral, Universidad Jaume I de Castellón (2003)
- Hopper, R.: The Ceramic Spectrum, 2nd ed. Chilton Book Co, ISBN 978-1-57498-302-9, Radnor, Penna (1984)
- García, A.: Pigmentos cerámicos CMYK de base nanoestructurada obtenidos por sonocoprecipitación y ruta poliol, Tesis Doctoral, Universidad Jaume I de Castellón (2011)
- García, A., Llusa, M., Sorli, S., Tena, M.A., Calbo, J., Monrós, G.: Eu³⁺-Nd₂O₃ blue pigmenting solid solutions. Brit. Ceram. Trans. **101**, 242–246 (2003)
- Gargori, C., Galindo, R., Cerro, S., García, A., Llusa, M., Monrós, G.: Synthesis of a new Ca_xY_{2-x}V_xSn_{2-x}O₇ yellow pigment. Phys. Procedia. **8**, 84–87 (2010)

10. Subramanian, M.A., Aravamudan, G., Subba Rao, G. V.: Oxide pyrochlores: a review. *Prog. Solid State Chem.* **15**, 55–143 (1983)
11. Ishida, S., Ren, F., Takeuchi, N.: New yellow ceramic pigment based on codoping pyrochlore-type $\text{Y}_2\text{Ti}_2\text{O}_7$ with V^{5+} and Ca^{2+} . *J Am Ceram. Soc.* **76**(10), 2644–2648 (1993)
12. García, A., Llusar, M., Calbo, J., Tena, M.A., Monrós, G.: Low toxicity red ceramic pigments for porcelanised stoneware from lanthanide-cerianite solid solutions. *Green Chem.* **3**, 238–242 (2001)



Pigment, Inorganic, Fig. 1 Structure of pigment consumption

Pigment, Inorganic

Teofil Jesionowski and Filip Ciesielczyk
Institute of Chemical Technology and
Engineering, Poznan University of Technology,
Poznan, Poland

Synonyms

Coatings; Dyes; Paints

Definition

Inorganic pigments, which have been widely used since prehistoric times, include naturally occurring substances prepared from minerals or their combustion products as well as synthetic compounds produced from appropriate raw materials and also hybrid pigment types derived from organic dyes and selected mineral supports. Inorganic pigments are insoluble in the surrounding medium and their optical effect arises from selective light absorption. They are classified according to their color, origin, method of production, and the character of the pigment material. Inorganic pigment technology is directly related to the development of new synthesis methods and the preparation of environmentally friendly colorants.

Introduction

This entry presents the main features of the classification and production of the most common

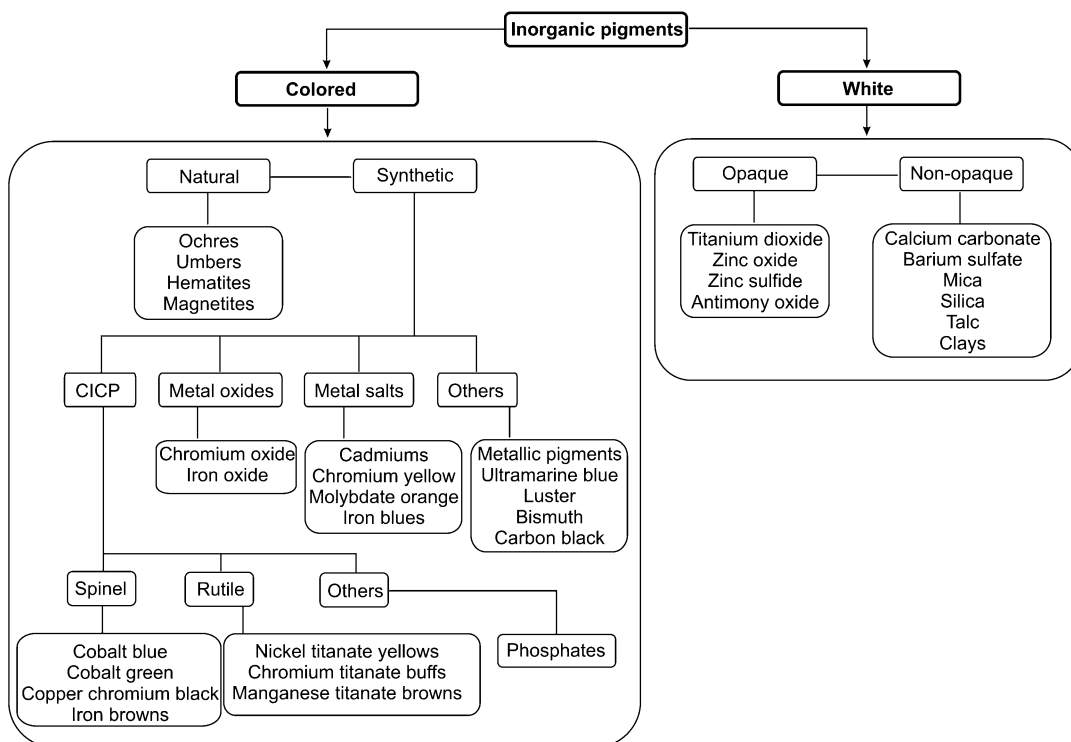
inorganic pigments. The most important physico-chemical and application properties of a wide range of pigments, including white pigments, are presented. Also discussed are issues relating to specialized pigments, including hybrid pigments.

The majority of inorganic pigments are fully suitable for use in a variety of consumer products such as paints and lacquers, plastics, inks, construction materials, paper, glass, and ceramics. The global consumption of pigments for 2012 shown in Fig. 1 clearly confirms the greater importance of inorganic pigments compared to organic pigments [1].

Inorganic pigments are more inert and insoluble and are more resistant to both temperature and extreme pH than their organic counterparts. Characteristic features of inorganic pigments include excellent opacity, good color saturation, and stability. Inorganic pigments enable the production of durable and high-quality coatings.

General Information

Pigments are colored materials which, in a purified form, are also used for the dyeing of synthetic fabrics, rubber, and paper and which impart color as a result of wavelength-selective light reflection, absorption, or interference. They improve the aesthetic value of products and also possess practical functions such as protecting metal objects against corrosion by means of painting. Pigments are insoluble in oil, water, organic solvents, and



Pigment, Inorganic, Fig. 2 Inorganic pigment classification

resins. They are classified according to their color, origin (inorganic or organic), method of production (natural or synthetic), and the character of the pigment material (e.g., light and weather resistance, high opacity, color strength, ease of dispersion, etc.). Inorganic pigments usually display better resistance to light and atmospheric conditions, higher dispersion, and opacity [2, 3].

There are pigments which, apart from coloration, also display luminescence (luminescent pigments) or change color as a function of temperature (thermosensitive pigments); these are used for the production of luminescent or thermosensitive paints.

Pigments can include different components; according to the character of the material from which they are made, they are divided into:

- Homogeneous pigments – comprise pigment particles of the same type, e.g., oxides of metals

- Mixed or hybrid pigments – produced by mechanical mixing (vigorous mixing of the coloring substance with an extender) or by chemical synthesis (organic dyes or pigments grafted onto a mineral support)

Classification of Inorganic Pigments

Pigments can be classified on the basis of their chemical structure, optical properties, or technological properties.

A detailed classification of inorganic pigments according to their chemical structure and color subgroups is shown in Fig. 2.

Properties

Pigments are mainly used for coloration and for endowing specific properties to paint coatings. They are important components of lacquer and

paint coatings, influencing their final mechanical properties and resistance; they affect the paints' rheology, stability, and some properties related to application. The physicochemical properties of pigments chiefly influence the application and technological conditions of their commercial use [3–5].

From an application point of view, the following functional properties of pigments are important for specific target demands:

- Optical properties – color, tinctorial strength, opacity
- Physical/ solid state properties – crystalline structure, grain morphology, character of surface, light refraction index, and defined structural properties
- Resistance and protective properties

All of these properties are to some degree determined by chemical structure [3].

The pigment color, shade, opacity, resistance to chemicals, as well as viscosity depend mainly on the chemical and crystalline structure, light refraction index, size and shape of the pigment particles, as well as the character of the particle surface. Crystalline structure determines such properties as dispersion, rheology, solubility, color strength, and shade.

Exemplary morphological structures of selected inorganic pigments and extenders are shown in Fig. 3.

The fundamental properties of inorganic pigments include the following:

- *Color and intensity of coloration*

The color of a pigment depends on chemical composition and degree of purity, as these two aspects define the pigment's ability to absorb certain wavelengths and to reflect the complementary color. The tinctorial strength or color strength of a pigment is defined as the pigment's ability to change the original color of a given material. This is related to the pigment's introduction to a given material and depends on the pigment's degree of refinement, purity, and homogeneity of chemical composition.

- *Opacity*

Pigment opacity (hiding power) depends on both the size and structure of the pigment particles as well as the difference in the light refraction index of the pigment and the binding agent. The covering ability of a pigment is greater when the pigment comprises small particles and the difference in light refraction index is then significant.

- *Shape and size of pigment particles*

The size of pigment particles is controlled so as to obtain optimum visible light scattering. The size of pigment particles impacts on the stability, coating resistance, and covering power. Pigment particle size can vary from 1 to 10 μm , with particles of the smallest diameters ensuring higher density and protective properties.

- *Resistance to UV radiation*

Pigments should display the highest possible resistance to light. A change in color of the pigment, usually brightening or darkening as a result of exposure to visible light or other radiation, is mainly caused by changes in the dispersion of the pigment in the binding agent and changes in that agent, rather than to changes within the pigment.

- *Thermal resistance*

Thermal resistance is of particular importance in the cases of paint coatings for use on hot substrates or when dyed plastic products are subjected to hot processing. The pigments used for such purposes should be resistant to high temperatures.

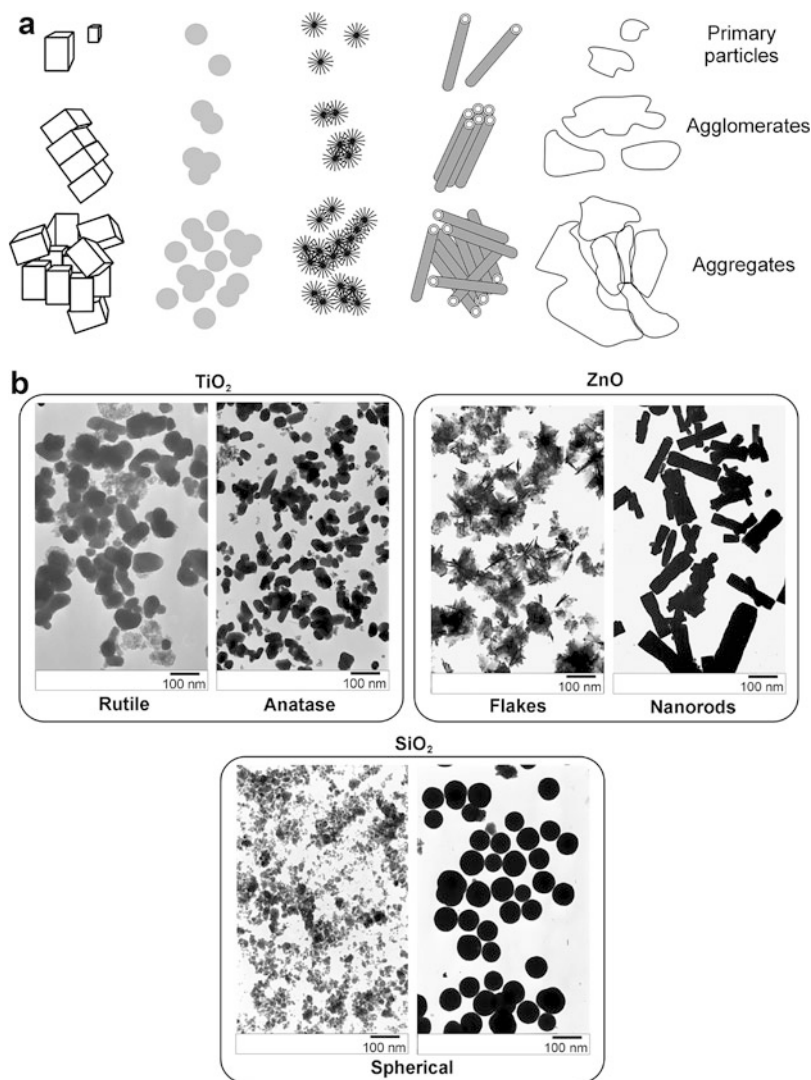
- *Resistance to chemical agents*

Of particular importance is a pigment's resistance to both acids and alkalies. A pigment's resistance to the binding agent, substrate, and exploitation conditions should also be analyzed. When pigments are mixed, their interaction must also be established.

- *Optical properties*

The optical properties of pigments are reflected in their color and light absorption coefficient. The color of a pigment is determined by the adsorbed or reflected wavelengths of light.

Pigment, Inorganic,
Fig. 3 (a) Particle morphological structure. (b) TEM images of commonly known inorganic white pigments and extenders (Created based on Ref. [3] with permission from VincentzVerlag Publisher)



- *Wettability by binding agent*

The extent of the wettability of a pigment by a binding agent depends on the degree of the pigment particles' refinement, pigment structure, and the wetting properties of the binding agent. For oil products, it is described by the so-called oil absorption number.

- *Toxicity*

Although mineral pigments often contain heavy metals, their presence is not a threat to human health or the natural environment. In the production of such pigments, the heavy metals used are so strongly bonded in the pigment structure that they are not

released into the environment. In many countries, the content of heavy metals in pigments for use in special products, such as those that are in contact with food, toys for children, and cosmetics, is controlled by regulations that specify admissible heavy metal levels.

Most of the aforementioned physicochemical properties of pigments determine their potential uses. Among the most common types displaying these physicochemical properties are white and black pigments and more rarely brown, yellow, red, blue, and green pigments.

White pigments – the optical effect of these is achieved by nonselective light scattering; technologically the most important examples are titanium dioxide, zinc white, zinc sulfide, and lithopone.

Titanium dioxide (C.I. Pigment White 6; TiO_2) occurs in three crystalline varieties, namely, brookite (not used as a pigment), anatase, and rutile. Irrespective of the technology of titanium dioxide production, the pigments are characterized by high purity and defined crystalline and particle size. To improve the functional properties of the pigment in a coating, the titanium white surface is modified with different chemical compounds. The attractive features of titanium white pigment include high light refraction index, chemical and physical stability, possibility of regulation of particle size distribution, and possibility of surface modification. The crystallites and particle size are the main parameters determining the paint properties. The optimum size of TiO_2 crystallites to ensure maximum light scattering, and hence high opacity of paint coatings, is $\sim 0.23 \mu\text{m}$. As the polymers used in paint formulations usually undergo degradation under the influence of ultraviolet radiation, the addition of substances to absorb UV radiation increases a paint coat's lifetime. Moreover, titanium dioxide has the highest refractive index (2.55 for anatase and 2.70 for rutile) of all inorganic pigments and commonly known extenders. Because of the universal properties of this pigment, it is produced by industrial methods on a wide scale [6, 7].

Production of TiO_2

Titanium white pigments are produced on an industrial scale using two methods, namely, sulfate and chloride:

Sulfate process – TiO_2 is precipitated from a solution of ilmenite ore using concentrated sulfuric acid; both rutile and anatase forms can be obtained.

Chloride process – TiO_2 is obtained as a result of the oxidation of titanium tetrachloride (TiCl_4) synthesized by reduction and chlorination of

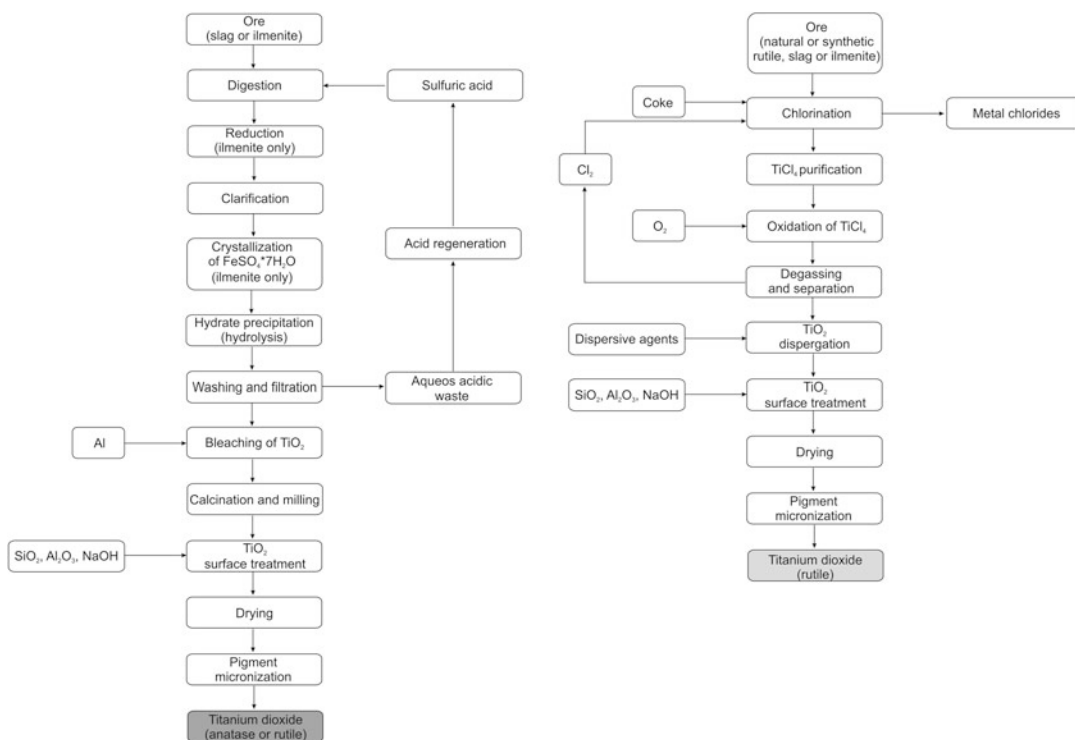
ilmenite ore or rutile sources; only the rutile form is obtained.

The chloride process is more often used; indeed, it is estimated that over half of the titanium white produced worldwide is obtained using this method. Although this process requires higher-quality and preliminarily enriched ore and employs more complex technology, it produces less waste (thus being less burdensome on the environment) and is cheaper than the sulfate method.

Sulfate Method

The sulfate method of titanium white production is complex and divided into two main technological parts, referred to as the “black” part (alluding to the color of the ore) and the “white” part (alluding to the color of white pigment).

The process aims to extract TiO_2 from raw titanium ore at elevated temperature by means of concentrated sulfuric acid (85–95 %). The most often used titanium ores are ilmenite (FeTiO_3) containing 45–65 % TiO_2 , and titanium slag. The prepared ore is mixed with concentrated sulfuric acid, and compressed air is blown through the reactor, as this accelerates etching. The reaction starts above 140°C and is exothermic, which causes a further increase in temperature to $200\text{--}220^\circ\text{C}$. The reaction mixture is left for 1–12 h to mature and the ensuing mixture is cooled to obtain a porous sinter, which is dissolved either in water or diluted sulfuric acid at low temperature ($<85^\circ\text{C}$) to prevent premature hydrolysis. The solution obtained (so-called titanium liquor) is reduced using steel scrap to convert Fe^{3+} into Fe^{2+} ions and partly also Ti^{4+} into Ti^{3+} ions. The solution is subjected to sedimentation, and the unreacted substrate is filtered out from the ensuing suspension. The content of TiO_2 in the product differs depending on the type of titanium ore: when ilmenite is used, it is 8–12 %, and when titanium slag is used, it is 13–18 %. When the temperature is decreased to 10°C , the main by-product of the process [hydrated iron sulfate ($\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$)] undergoes crystallization. When titanium slag with a TiO_2 content $>75\%$ is used, the iron sulfate



Pigment, Inorganic, Fig. 4 Production of TiO₂ by the sulfate and chloride methods

crystallization stage is not necessary. The purified solution is concentrated to 200–250 g TiO₂dm⁻³. Hydrolysis of titanium sulfate is then performed (at 94–110 °C) by diluting the solution and maintaining the suspension at its boiling point. This process influences the size of the product particles and the degree of flocculation. The hydrated titanium dioxide formed is filtered off, washed with water, and subjected to whitening by hydrogen. The hydrated titanium dioxide obtained contains 20–28 % sulfuric acid and sulfates which are removed by filtration and washing with water, followed by reduction in the process of whitening, in which most of impurities are removed; at the end of this process, the content of colored impurities is in the order of ppm, while the content of sulfuric acid is still significant, at around 5–10 %.

In the “white” part of the process, the hydrated titanium dioxide is again filtered, washed, and prepared for calcination. When the aim is to produce the rutile form, nuclei of rutile particles and

substances favoring transformation of anatase to rutile are introduced into the mixture, and calcination is performed at ~950 °C. When the aim is to obtain the anatase form of the pigment, no admixtures are introduced and calcination is performed at a lower temperature. The hydrate formed is calcined to remove water and to obtain TiO₂. When rotary filters are employed, the content of titanium dioxide in the product is 30–40 %, but when pressure rotary filters or automated filtration press is used, the content reaches 50 %.

A schematic presentation of the sulfate method of producing titanium white pigment is given in Fig. 4, and the reactions at each step are listed in Table 1.

Chloride Process

In the chloride process, the substrates are either rutile or enriched titanium ore containing >90 % TiO₂. This method is more modern, more technologically advanced, and less burdensome on the natural environment than the sulfate process;

Pigment, Inorganic, Table 1 Characteristics of commonly known inorganic pigments

Pigment type	Production	Advantages	Disadvantages	Color
Carbon black	Decomposition of carbonaceous precursors	Rich color, color strength, light and weather resistant	Thickens paint, color limits application	Black
$\text{CH}_4 \xrightarrow{\text{temp.}} \text{C} + 2\text{H}_2$				
Titanium dioxide	Synthesized mainly by sulfate or chloride technology (alternatively using sol-gel, hydrothermal, or solvothermal methods)	High opacity and color strength, relatively cheap, very good UV resistance, very high refractive index, good resistance to alkaline solutions, mostly chemically inert pigment	Tendency for agglomeration of primary particles, forms radicals that degrade the binder/polymer matrix	White
$\text{FeTiO}_3 + 2\text{H}_2\text{SO}_4 \longrightarrow \text{TiOSO}_4 + 2\text{H}_2\text{O} + \text{FeSO}_4$ $\text{TiOSO}_4 + 2\text{H}_2\text{O} \longrightarrow \text{TiO}(\text{OH})_2 + \text{H}_2\text{SO}_4$ $\text{TiO}(\text{OH})_2 \xrightarrow{\text{temp.}} \text{TiO}_2 + \text{H}_2\text{O}$				
$\text{TiO}_2 + \text{C} + 2\text{Cl}_2 \longrightarrow \text{TiCl}_4 + \text{CO}_2$ $\text{TiCl}_4 + \text{O}_2 \longrightarrow \text{TiO}_2 + 2\text{Cl}_2$				
Iron oxides	Synthesized, obtained as minerals and also prepared from chemical wastes	Light and weather resistant, mostly unreactive	Cannot produce clean shades	Yellow, red, brown, black
$\text{C}_6\text{H}_5\text{NO}_2 + 2\text{Fe} + \text{H}_2\text{O} \longrightarrow \text{C}_6\text{H}_5\text{NH}_2 + \text{Fe}_2\text{O}_3$ $3\text{Fe}_2\text{O}_3 + \text{H}_2 \longrightarrow 2\text{Fe}_3\text{O}_4 + \text{H}_2\text{O}$ $3\text{Fe}_2\text{O}_3 + \text{CO} \longrightarrow 2\text{Fe}_3\text{O}_4 + \text{CO}_2$ $2\text{Fe}_3\text{O}_4 + 1/20_2 \xrightarrow{\text{temp.}} 3(\gamma - \text{Fe}_2\text{O}_3)$ $2\text{Fe}_3\text{O}_4 + 1/20_2 \xrightarrow{\text{temp.}} 3(\alpha - \text{Fe}_2\text{O}_3)$				

Zinc oxide	Chemical or metallurgical synthesis	Excellent heat stability, light and weather resistant, very good UV resistance, anticorrosive properties	Expensive chemical synthesis, tendency to form agglomerations	White
$\left. \begin{array}{l} \text{ZnSO}_4 + 2\text{NaOH} \longrightarrow \text{Zn(OH)}_2 + \text{Na}_2\text{SO}_4 \\ \text{ZnCl}_2 + 2\text{NaOH} \longrightarrow \text{Zn(OH)}_2 + 2\text{NaCl} \\ \text{Zn(CH}_3\text{COO)}_2 + 2\text{NaOH} \longrightarrow \text{Zn(OH)}_2 + 2\text{CH}_3\text{COONa} \\ \dots\dots\dots \\ \text{ZnCO}_3 \longrightarrow \text{temp. ZnO} + \text{CO}_2 \end{array} \right\} \text{Zn(OH)}_2 \longrightarrow \text{temp. ZnO} + \text{H}_2\text{O}$				
Zinc chromates	Synthesized	Anticorrosive properties	Thickens paints, dangerous synthesis with regard to reagent	Yellow
$4\text{ZnO} + 2\text{CrO}_3 + \text{K}_2\text{Cr}_2\text{O}_7 + \text{H}_2\text{O} \longrightarrow \text{temp. K}_2\text{CrO}_4 + 3\text{ZnCrO}_4$	Zn(OH)_2			
Ultramarine	Synthesized, seldom obtained as mineral	Rich color	Low stability in acidic environment	Blue
$6[\text{Al}_2\text{Si}_2\text{O}_7] + 7\text{Na}_2\text{CO}_3 + \text{nS} \longrightarrow \text{temp. } 2[\text{Na}_6\text{Al}_6\text{Si}_6\text{O}_{24}\text{NaS}_6] + 7\text{CO}_2$				
Chromium oxides	Synthesized as native pigments or produced from wastes	Light, weather, acid and alkali resistant, thermal stability	Does not give clear colors	Green
$\begin{array}{l} 4\text{CrO}_3 + 3\text{N}_2\text{H}_4 \longrightarrow 4\text{Cr(OH)}_3 + 3\text{N}_2 \\ 2\text{Cr(OH)}_3 \longrightarrow \text{temp. Cr}_2\text{O}_3 + 3\text{H}_2\text{O} \\ (\text{NH}_4)_2\text{Cr}_2\text{O}_7 \longrightarrow \text{temp. Cr}_2\text{O}_3 + \text{N}_2 + 4\text{H}_2\text{O} \\ \text{Na}_2\text{Cr}_2\text{O}_7 + 2\text{S} \longrightarrow \text{temp. Cr}_2\text{O}_3 + \text{Na}_2\text{SO}_4 \end{array}$				
Cadmium sulfides	Synthesized	High opacity and color strength	Poor weather resistance, very expensive	Greenish yellow, red, deep red
$\text{CdCl}_2 + (\text{NH}_2)_2\text{CS} + 2\text{NH}_4\text{OH} \longrightarrow \text{CdS} + (\text{NH}_2)_2\text{CO} + 2\text{NH}_4\text{Cl} + \text{H}_2\text{O}$				

however, it enables the production only of rutile. The problems related to this method are the high temperature of the process; the aggressive medium; the used toxicity of chlorine, carbon monochloride, and titanium tetrachloride; as well as the risk of uncontrolled emission of chlorine gas. The advantages of this process are that the pigments obtained have higher resistance to light and are brighter than those obtained by the sulfate method.

The substrate, which is maintained in the fluid phase by a flux of air, is heated to 650 °C and then refined coke (250–300 kg per tonne of TiO_2) is introduced. The substrate should be dry and should not contain small particles that could be carried by the stream of forming TiCl_4 and contaminate the product. The burning coke increases the temperature to about 1,000 °C at which point chlorine gas and a continuous supply of titanium ore (substrate) and coke are provided.

The ensuing reaction leads to the formation of titanium tetrachloride gas, which is the main product as well as by-products that include chlorides of contaminating metals (e.g., iron, manganese, and chromium), which together with the unreacted substrate are separated from the reaction mixture by fractional condensation. The titanium tetrachloride formed and the accompanying chlorides are distilled off from the fluid bed, and these gases are then cooled with liquid TiCl_4 . As some metal chlorides (mainly iron chloride) can deposit on the surface of the TiCl_4 upon crystallization, the chlorides are condensed off, thereby removing the impurities of the TiCl_4 . Cooling is carried out in stages; firstly, the gases are cooled to below 300 °C and then the chlorides co-occurring with titanium tetrachloride are separated by condensation or sublimation. After this stage, the gas mostly comprises TiCl_4 and is cooled to <0 °C, at which point, condensation of TiCl_4 occurs. The raw titanium tetrachloride is carefully purified, at first, by a chemical method to separate the substances that could not be removed upon distillation, such as VCl_4 and VOCl_3 , and then by vacuum distillation, which gives highly pure TiCl_4 . The pure titanium tetrachloride is evaporated and combusted in a stream of oxygen or air to obtain TiO_2 and chlorine; the reaction involved

is weakly exothermic and requires a high temperature in the range 900–1,400 °C. The required pigment is obtained together with a mixture of gases (Cl_2 , O_2 , CO_2) which are cooled and then separated using either an indirect or direct method. The stream of gases is recirculated to the cooling zone of the furnace and to the chlorination stage.

A schematic presentation of the chloride process of titanium white production is given in Fig. 4, and the reactions within each stage are listed in Table 1.

After cooling, titanium dioxide obtained via either the sulfate or chloride method is ground using a wet or dry method and dispersed in water. The pigment can be subjected to surface modification with hydrated oxides (e.g., silica, alumina, or zirconia) to obtain varieties of pigment for special applications. The main purpose of this surface treatment is to improve the pigment's stability, reactivity, and dispersibility; treatment is carried out in the water phase, and solutions of the substances used to modify the particles of titanium dioxide are added in a defined sequence or simultaneously. By regulating the pH of the suspension, hydrated oxides are precipitated on the surface of the pigment particles; the form of the oxides strongly depends on the conditions of precipitation. Alumina, silica, zirconia, or other oxides do not occur in the pigment as an admixture, but as a molecular layer bonded to the pigment surface. After treatment, the pigment is filtered again, dried, and ground. The product is finally packed and stored.

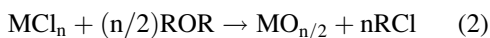
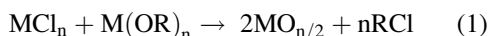
Other alternative methods of TiO_2 production, which allow the control of process parameters as well as the physicochemical properties of the products, include the sol–gel, hydrothermal, and solvothermal methods, as well as chemical vapor deposition.

Sol–Gel Method

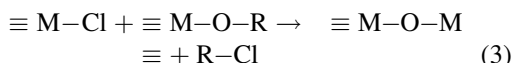
In a typical sol–gel process, a colloidal suspension or sol is formed as a result of hydrolysis and polymerization of precursors, which are usually inorganic metal salts or organometallic

compounds such as metal alkoxides. Polymerization and loss of solvent result in conversion of the liquid sol to a solid state gel phase.

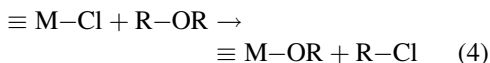
The sol-gel method is based on hydrolysis and condensation of metal alkoxide or metal salt. In the reaction of the metal chloride with either metal alkoxide or organic ether, the latter two compounds act as oxygen donors, according to Eqs. 1 and 2:



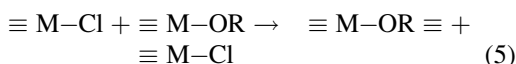
In these reactions, M–O–M bond formation is promoted by condensation between M–Cl and M–OR, according to Eq. 3:



In the reaction with ether, Eq. 4, metal alkoxide is formed as a result of alcoholysis with M–Cl:



At room temperature, the reactions are slow, and oxide formation is usually favored at an elevated temperature in the range 80–150 °C. The main reaction (Eq. 5) between metal chloride and metal alkoxide takes place at room temperature:



An important advantage of the sol-gel method is the possibility of obtaining TiO₂ particles of nanometric size and controlled shape.

Hydrothermal and Solvothermal Methods

This method is commonly used for obtaining titanium dioxide of small-diameter particles required for the ceramic industry. TiO₂ nanoparticles are

obtained by peptization of titanium precursors with water. The hydrothermal method is commonly used for the production of different morphological forms of TiO₂ such as nanotubes, nanowires, and nanorods.

The solvothermal method is very similar to the hydrothermal method, the difference being that in the former process, an anhydrous solvent is used instead of water. The range of temperatures used can be much greater than in the hydrothermal method and depends on the boiling point of the organic solvent used. In the solvothermal process, it is easier to control the size, shape, and crystallinity of the TiO₂ nanoparticles than in the hydrothermal process. The solvothermal method is regarded as a versatile method for the synthesis of nanoparticles with narrow particle size distribution.

Chemical Vapor Deposition

Chemical vapor deposition (CVD) is used for the thermochemical processing of materials. The aim of the method is to deposit thin films of TiO₂ on a particular material so as to change its physical, chemical, or mechanical properties. In the reaction chamber, titanium precursors, mostly in the gaseous phase, deposit on the hot substrate surface. Usually CVD methods require temperatures of ≥900–1,100 °C or higher, necessary for film formation, and this condition significantly restricts the use of the method. In order to obtain the target products, different gas or liquid substrates (precursors) are used, e.g., halogens, chlorides, carbonyls, and volatile organic compounds of titanium.

The wide range of production methods, as well as the outstanding physicochemical and utilizable properties of TiO₂, makes this pigment an extremely attractive product in the paint and lacquer industries. The greatest amount of titanium white is used for the production of paints. Thanks to very high light refraction index, paint coatings containing titanium dioxide are intensely white and have high opacity, which means that only small amounts of TiO₂ are needed to produce an excellent white opaque coating. Titanium white

also increases coating durability and ensures high covering power, so that the coating can be thinner, which improves its functionality. This pigment is used to obtain high-quality lacquers, both white and in pastel colors, in mixtures with other pigments and emulsion paints. Paint production uses a wide variety of titanium white pigments, namely, rutile, anatase, pigments with modified and unmodified surface, or pigments modified for special needs. The type of pigment used depends on the type of paint and its application. For external use (paints exposed to sunlight), rutile varieties are used, while for indoor applications, anatase is preferred. Anatase varieties with lower resistance are used in cheap dispersion paints, in self-cleaning paints, and in paints for road markings. Rutile varieties have much wider use and can be employed in all types of paints.

Although TiO_2 is also used for the production of printing inks, less than 3 % of global production of titanium white is used for this purpose. TiO_2 pigments in printing inks ensure opacity and high brightness of the coating. Because of the thinness of printing coatings, TiO_2 pigments employed for this purpose should have a narrow particle size distribution and should not contain agglomerates. The choice of pigment is very important as it affects many properties of the ink, such as luster, color, covering power, susceptibility to sedimentation, wearing, and rheological properties. The varieties of titanium white pigment should be chosen depending on the final intended use of the ink.

The second best-known and widely used white pigment is *zinc oxide* (C.I. Pigment White 4; ZnO) which is produced by a metallurgical process based on zinc ore calcination [8].

Metallurgical Process

Zinc oxide products are divided into:

- Type A – obtained by a direct process (American process)
- Type B – obtained by an indirect process (French process)

In the direct (American) process, zinc ore is reduced by heating with coal (e.g., anthracite) followed by oxidation of the resulting zinc vapors in the same reactor. This single production cycle was proposed by Wetherill. In the indirect (French) process, molten, metallic zinc is evaporated at $\sim 907^\circ\text{C}$, and the immediate reaction of the zinc vapor with oxygen from air gives ZnO . Zinc oxide particles are transported through a cooling channel and collected by cyclones and sock filters. This process was popularized by LeClaire in 1844 and since that time it has been called the French process. The product is obtained in the form of agglomerates of mean size from 0.1 to a few μm . The particles of ZnO are of mainly spheroidal shape. Zinc oxide of type B is of higher purity than type A.

Of great importance in the technology of production of ZnO with defined physicochemical and structural properties are the chemical processes described below.

Controlled Precipitation

This is widely used as it enables products with reproducible properties to be obtained. The idea of the method is to perform the fast and spontaneous reduction of a zinc salt using a reducing agent that halts the growth of particles at a certain size, followed by precipitation of ZnO precursor from solution; in the next step, the precursor is subjected to thermal treatment and grinding. Breaking up of the agglomerates formed is very difficult, so the calcined powders show a high level of agglomeration. The precipitation stage is controlled by pH, temperature, and time. The most important zinc precursors are zinc acetate and zinc hydroxide. A necessary stage of this process is the calcination of hydrated ZnO above 120°C (see Table 1).

Mechanochemical Process

This cheap and simple method for the synthesis of nanoparticles on a large scale is based on high-energy dry milling, which initiates a reaction upon

ball–powder collisions in a ball mill at low temperature. The main difficulty in this method is the uniform grinding of the powder and reduction of the pigment grains to the target size. Although pigment size decreases with increasing time and energy of milling, longer milling times lead to the greater amounts of impurities. The benefits of this method are that the particles are of low cost, of small size, and of limited tendency to form agglomerations, as well as their highly uniform crystalline structure and morphology.

The initial materials used in this method are mainly anhydrous ZnCl_2 and Na_2CO_3 . The reactor is charged with NaCl , which is the reaction medium and which separates the forming nanoparticles. The zinc precursor formed (ZnCO_3) is subjected to calcination at 400–800 °C.

Other significant methods for the synthesis of zinc oxide are the hydro- and solvothermal methods and synthesis of morphologically defined particles in emulsion and microemulsion systems, as described above for TiO_2 [8].

Besides TiO_2 and ZnO , commonly known white pigments are *lead white* (C.I. Pigment White 1; $2\text{PbCO}_3\text{Pb(OH)}_2$) and *zinc sulfide* (C.I. Pigment White 7; ZnS).

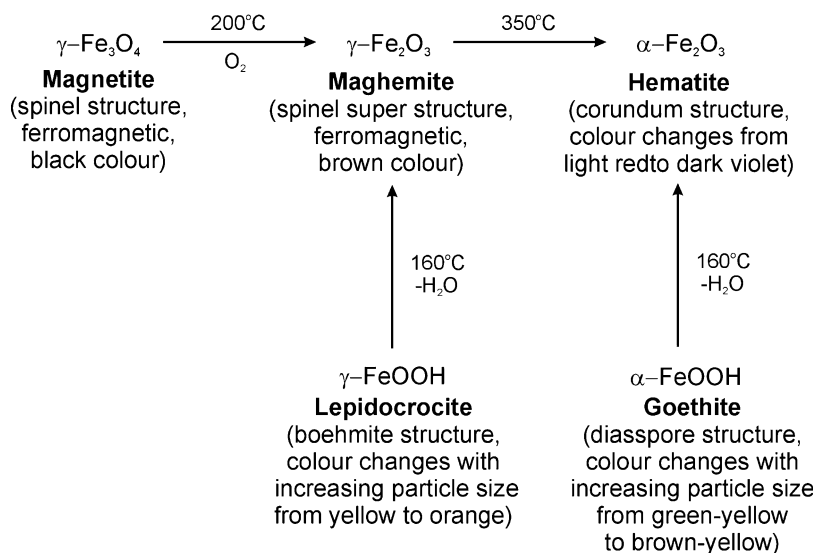
Black pigments – the optical effect of these is achieved by nonselective light absorption; among

this group of pigments, the most common are carbon black and black iron oxide.

The most common black pigment is *carbon black* (C.I. Pigment Black 6, 7, 8). The majority of carbon black pigments have highly developed surface area, which means that their oil absorption numbers are high. Their high surface activity is responsible for their poor wettability, the effect of flotation, and adsorption of paint components on the surface. Carbon black is an inorganic pigment, but it has many properties of organic compounds. The paint industry uses only modest amounts of black pigments compared to the rubber industry. In the production of rubber for tires, carbon black not only provides color but also increases mechanical strength. Black pigments based on carbon black have excellent resistance to light and solvents, although certain solvents can extract some colored contaminants from cheap types of pigments. Black pigments show excellent chemical and thermal stability. The intensity of the black color depends on the size of the particles: the smaller the size, the more intense the color.

A very attractive black pigment is also *iron oxide* (C.I. Pigment Black 11; Fe(II) and Fe(III) oxide; FeO , Fe_2O_3). In natural form, it is found as magnetite (also maghemite, hematite, goethite, lepidocrocite – Fig. 5), which can be used as a pigment, although most iron oxide black pigments

Pigment, Inorganic,
Fig. 5 Typical forms of
iron oxide-based pigments



are nowadays synthesized. Iron oxide is a low-cost, inert pigment of excellent chemical resistance, insoluble in organic solvents, and of excellent durability, though unfortunately it also has limited thermal stability and low tinctorial strength. This group of pigments also includes *graphite* (C.I. Pigment Black 10), formed by crystallization of carbon in hexagonal form.

Color pigments – the optical effect of these is achieved by selective light absorption often in combination with selective scattering; examples include brown, yellow, red, blue, and green pigments.

Brown pigments, exemplified by brown iron oxide (C.I. Pigment Brown 6 synthetic and 7 natural; Fe_2O_3), which naturally occurs as burnt sienna or burnt umber. Different shades can be obtained depending on the presence of impurities, in particular, manganese oxide. The pigments are more often used for artists' paints than for commercial paints. They have low tinctorial strength; synthetic pigments give an intense brown color of excellent resistance. They are not widely used for paint production, as similar shades can be obtained by mixing together cheaper pigments.

Lead chromates (C.I. Pigment Yellow 34, Orange 21; $\text{PbCrO}_4/\text{PbSO}_4$; orange $\text{PbCrO}_4/\text{Pb}(\text{OH})_2$) are a good example of yellow pigments. The pigments can be composed of pure lead chromate or of a mixture of lead chromate and lead sulfate. They provide colors from green yellow to orange. Chrome yellow displays excellent opacity, low oil absorption number, high brightness, and high color saturation, which make it an ideal pigment for the production of paints in a wide gamut of yellow shades. Yellow pigments have very high resistance to solvents and relatively good thermal stability, which can be improved by chemical stabilization. They are sensitive to both alkaline compounds and acids, which cause them to fade. Their resistance to light is usually satisfactory, although they can become slightly darker under exposure to light. These drawbacks can be counteracted by surface modification with such materials as silica, antimony, aluminum oxide, or different metals. Other yellow pigments are cadmium yellow (C.I. Pigment Yellow 37, 35 lithopone; CdS ; lithopone – CdS

containing ZnS) and yellow iron oxide (C.I. Pigment Yellow 42 synthetic, 43 natural; $\text{Fe}_2\text{O}_3 \cdot \text{H}_2\text{O}$ or more exactly $\text{FeO}(\text{OH})$).

Excellent functional properties, such as thermal and chemical stability, and resistance to solvents are provided by red iron oxides (C.I. Pigment Red 101 synthetic, 102 natural; Fe_2O_3), which belong to the category of red pigments. They are also highly opaque and have quite low tinctorial strength. Despite the serious drawback of its dull brown-red shade, it has the great advantage of low cost.

Ultramarine (C.I. Pigment Blue 29, Violet 15; general formula $\text{Na}_7\text{Al}_6\text{Si}_6\text{O}_{24}\text{S}_3$), which belongs to the group of blue pigments, has attractive features, including dazzling color and a pure shade that can vary from pink through violet to green. It is also characterized by excellent thermal stability, resistance to solvents, and good resistance to light and alkaline compounds. It is not resistant to acids, and thus it cannot be used in exterior paints. Other well-known blue pigments are Prussian blue (C.I. Pigment Blue 27; $\text{FeK}_4\text{Fe}(\text{CN})_6$) and cobalt blue (C.I. Pigment Blue 36; $\text{Co}(\text{Al,Cr})_2\text{O}_4$).

Last in sequence but not least in importance are the green pigments, such as green chromium oxide (C.I. Pigment Green 17; Cr_2O_3) and hydrated chromium oxide (C.I. Pigment Green 18; $\text{Cr}_2\text{O}(\text{OH})_4$). These pigments have an opaque shade, good opacity but low tinctorial strength. They also offer excellent thermal stability and chemical resistance to solvents and light [2, 9, 10].

Because of their specific properties, inorganic pigments are used mainly in the paint and lacquer industries. However, in order for the final products (paints and coatings) to meet set requirements, they must be formulated with other additives that do not provide any staining or opacity. These additives are known as extenders (Table 2). They may be used to change typical pigment properties as well as specific functions [3, 11].

Inorganic pigment technology is directly related to the development of new synthesis methods and the preparation of environmentally friendly colorants. The manufacture of pigments containing metals such as $\text{Pb}(\text{II})$, $\text{Cr}(\text{VI})$, $\text{Cd}(\text{II})$, etc. in aqueous media involves the risk of releasing harmful substances to the environment and a

Pigment, Inorganic, Table 2 Example pigment extenders

Extender type	Formula	Uses
Calcium carbonate	CaCO_3	Additive for paints containing solvent, printing inks
Silica	SiO_2	Additive for rheology and thixotropy control, anti-settling, thickening, and anti-sagging agent
Talc	$\text{Mg}_3\text{Si}_4\text{O}_{10}(\text{OH})_2$	Gives high hiding power and a stain finish, provides a matting effect and decreased moisture permeability
Barytes	BaSO_4	Additive for waterborne paint systems
Kaolin	$\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$	Decreases viscosity
Mica	$\text{AB}_{2-3}(\text{Al,Si})\text{Si}_3\text{O}_{10}(\text{OH})_2$; A = K, Na, Ca; B = Al, Fe, Mg, Li	UV and chemically resistant, improves water resistance, provides pearl effect

need to dispose of materials which contain them; this restricts or prohibits their application. Such pigments cannot be used in paints, but are potentially useful and environmentally safe in other industries, such as catalysis and ceramics, but only as stable spinels. The manufacturers and users of such materials are still obliged to comply with environmental and regulatory restrictions in their synthesis and use. The production of the new pigment group targets first of all safe materials; waste materials, though, such as chromium (VI) compounds, can also be used for the direct synthesis of Cr_2O_3 -based pigments. Those aspects by no means diminish interest in inorganic pigments and their use; on the contrary, due to environmental reasons, the focus on and the industrial role of TiO_2 -based pigments or hybrid pigments prepared using alternative methods continue to increase.

A special group of inorganic pigments consists of *pigments with luster (mirror effect)*, *luminescent pigments*, and also *hybrid pigments*

[12–16]. The first group of these special pigments includes the following:

Metallic pigments – (only inorganic) contain lamellar metal particles of high refraction index, e.g., flakes of copper, gold, aluminum.

Pearlescent pigments – contain transparent lamellar particles, which are responsible for a pearl shine which is the result of multiple reflections from particles oriented in parallel; examples include specially treated alkaline lead carbonates, bismuth oxychloride, and titanium dioxide supported on mica.

Interference pigments – here the optical effect is a result of the thickness of the oxide layer deposited on supports (similarly as in pearlescent pigments); examples are oxides of iron or chromium supported on mica; pigments in this group also display opalescence.

Also very important are luminescent pigments, including:

Fluorescence pigments – of which the optical effect is based on selective absorption of light and simultaneous luminescence and is initiated by high-energy radiation from the ultraviolet or shortwave visible range; examples are radioactive luminescent pigments.

Phosphorescence pigments (only inorganic) – here the optical effect is based on selective light absorption and scattering followed by delayed luminescence and is also initiated by high-energy radiation from the ultraviolet or shortwave visible range; examples are zinc sulfide and alkaline earth metal sulfides “dotted” with heavy metal ions.

Hybrid Pigments

Recently, much interest has been paid to the synthesis of hybrid pigments. These can occur in three main combinations, namely, organic/organic, inorganic/inorganic, and organic/inorganic. This group of pigments includes those obtained on the basis of organic dyes and oxide support such as SiO_2 . Substances of this type are

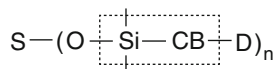
produced mainly as a result of adsorption of organic molecules on to a synthetic or mineral oxide support. Adsorption of organic compounds such as dyes is determined first of all by the type of interactions involved (covalent bonds, electrostatic force, hydrogen bonds, etc.) and the character of the inorganic support (SiO_2 , TiO_2 , $\text{SiO}_2\text{-TiO}_2$, Al_2O_3 , aluminosilicates, and carbonaceous materials), such as its surface properties and dispersion abilities [14–27].

The most often used selective adsorbent of dyes is silica synthesized by the sol–gel method. Alternative supports include silicas synthesized in the precipitation reaction from water solutions of alkali metal silicates, mainly sodium silicate. A large number of selective adsorbents are provided by mineral substances such as kaolin and montmorillonite. The core of the pigment composites can also consist of biopolymers of natural origin (chitin, chitosan, lignin) or synthetic compounds.

Hybrid pigments of this type can be obtained in different colors depending on the color of the initial dye. They display many advantages over conventional pigments. The particle size of such pigments depends on those of the initial support. Moreover, from the same core, it is possible to secure pigments with the same particle size, density, and surface properties, but differing in color and hue.

Pigments Obtained by Grafting of an Organic Dye to a Modified Silica Surface

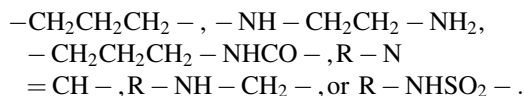
Such pigments are composed of a core that is a fine powder – SiO_2 , TiO_2 , Al_2O_3 , etc. – and a dye chemically bonded with the silica surface by a coupling agent, e.g., aminosilane. They display high color strength, excellent durability, and resistance to solvents, water, temperature, and light. Their structure can be described by the formula:



where:

S is the support.

CB is a coupling bridge – an alkyl group (3–6 carbon atoms) containing $-\text{NH}-$ or $-\text{NHCO}-$ in a chain like:



D is the organic dye.

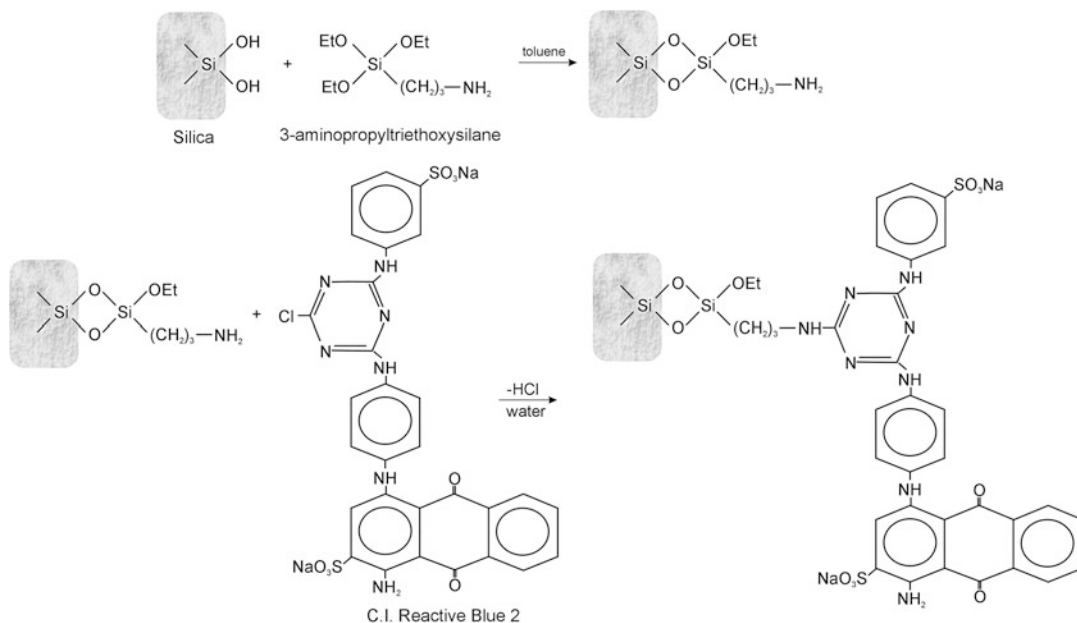
The most popular inorganic support for hybrid pigment formulation is silica. Hydrophilic silicas, which are usually colorless and have silanol groups on their surface, can react with many functional groups through covalent bonds. The process of dyeing these silicas is multistaged: silica is first reacted with a coupling agent, and then the dye (reactive, acid, or cationic) reacts with the coupling agent (e.g., silane) to give silica particles that contain covalently bound dye. The dye is permanently bonded to the silanol coupling agent and, therefore, cannot be leached by a solvent, which leads to reduced toxicity of the components within a paint or ink.

The sequence of reactions in this method is shown schematically in Fig. 6 and includes modification of the silica surface with 3-aminopropyltriethoxysilane in order to introduce amino groups, which subsequently react with the reactive dye (e.g., C.I. Reactive Blue 2) to give dyed silica particles [28].

Through simple and economical processes, it is possible to obtain colored particles suitable for use as pigments for inks which display high resistance to water and temperature and low toxicity. The inks can be made in a wide range of colors and can be used for printing on foil, smooth, or coated paper.

This process can be also realized by the multistep synthesis of a chromophore on the surface of an inorganic support.

Pigment synthesis involves modification of the support surface – which becomes the core of the pigment – with an aminosilane coupling agent to introduce surface amino groups and then bonding the dye with amino groups. For this purpose, silicas of small diameters are used, as when silica



Pigment, Inorganic, Fig. 6 Mechanism of hybrid pigment formation by adsorption of organic dye onto aminosilane-functionalized silica support (Created based on Ref. [28] with permission from Elsevier Publisher)

particles are small, the area of contact is usually larger, which is beneficial for mixing with other materials.

Pigments Obtained in the Process of Silica Synthesis

Pigments of this type are obtained using silica synthesized in a process of hydrolysis and condensation of tetraethoxysilane (TEOS) in a mixture of ethanol/water/ammonia and in the presence of cationic dyes such as C.I. Basic Blue 9 (Fig. 7).

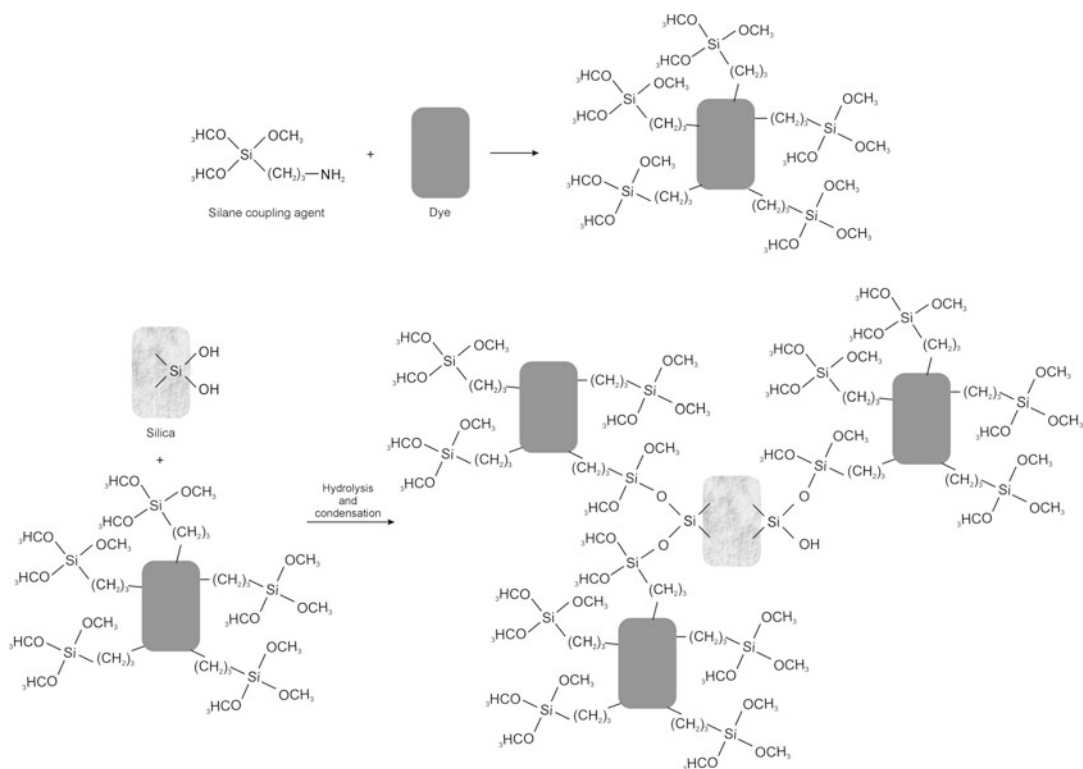
The initial dye solution is prepared in a water medium and is then filtered off through a suitable membrane. In order to synthesize inorganic particles containing a dye, the silica sol and dye solution are mixed to obtain a permanent product. Often, the whole amount of the dye used in the process is incorporated into the structure of the pigment, as indicated by a colorless solution remaining after the process. The color particles are filtered off and washed with distilled water until no dye is left in the filtrate.

Retention of dye in the support particles is tested by dispersion in ethanol or acetone, followed by ultrasound treatment and centrifugation. This procedure is repeated a few times to check the stability of the bonding of the dye to the support surface.

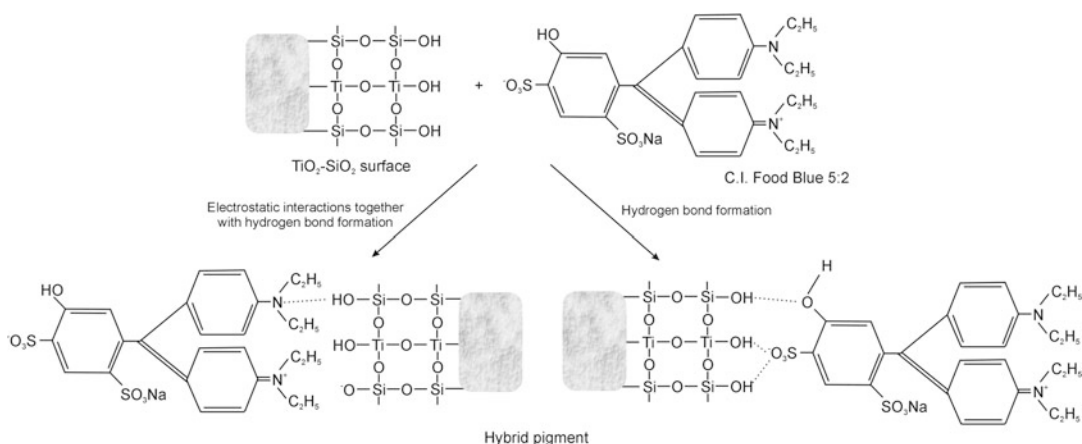
There is also interest in combining the functional features of inorganic pigments (TiO₂) or their derivatives (TiO₂-SiO₂) with selected dyes used in the food and/or pharmaceutical industries (e.g., C.I. Food Yellow 4, C.I. Food Blue 5:2, C.I. Food Red 9) [16]. An example mechanism of synthesis of such a hybrid pigment is illustrated in Fig. 8.

Such a combination makes it possible to obtain a homogeneous hybrid system, which is a great advantage from the application point of view, as separate introduction of dye and pigment at the stage of drug formulation (e.g., tablets) is no longer needed and problems with their steric and electrokinetic stability are eliminated.

The physicochemical and structural parameters of the hybrid pigments depend on many factors, among which the most important are the chemical structure of the dye and the support, concentration of reagents, reaction time, pH, etc.



Pigment, Inorganic, Fig. 7 Modification of the pigment and binding with the silica surface



Pigment, Inorganic, Fig. 8 Mechanism of C.I. Food Blue 5:2 interactions with $\text{TiO}_2\text{-SiO}_2$ surface (Created based on Ref. [16] with permission from Elsevier Publisher)

Summary

Technologies for the production and application of a wide range of inorganic pigments are closely

interrelated and are constantly being developed. In an era in which many branches of industry are undergoing dynamic development, there is a need to improve and obtain new compounds of various

kinds, including inorganic pigments, with specific physicochemical and application properties. New trends in the development of pigments relate mainly to:

- The introduction of new pigments or improvement of the properties of existing ones
- Broadening of the assortment and range of uses of pigment composites
- The introduction of nanoparticle pigment structures

The above is confirmed by the numerous R&D projects carried out at scientific centers and in industry, which to date have published more than 25,100 scientific papers relating to inorganic pigments.

The main reason for the high level of interest, particularly in the use of nanoparticles, is the possibility of improving the barrier properties of coatings and thereby also corrosion resistance and mechanical properties. Other drivers of innovation in the field of pigments are environmental awareness, economic pressures, and the constantly changing needs of the industries which consume these products. Hence the greatest level of development is observed in relation to pigments with special metallic and interference effects, including hybrid pigments. All of these aspects ensure unceasing demand and continued efforts to obtain a wide color range of inorganic pigments, whose importance can be expected to continue to increase.

Cross-References

- [Colorant, Environmental Aspects](#)
- [Dye, Functional](#)
- [Pigment, Ceramic](#)

References

1. Garo, D.: Pigments – a growth market. *Eur. Coat. J.* **1**, 10–13 (2012)
2. Endriß, H.: *Inorganic Coloured Pigments Today*. Vincentz Verlag, Hannover (1998)
3. Golgschmidt, A., Streitberger, H.J.: *BASF-Handbuch Lackiertechnik*. Vincentz Verlag, Hannover (2002)
4. Lambournle, R., Strivens, T.A. (eds.): *Paint and Surface Coatings Theory and Practice*, 2nd edn. Woodhead Publishing, Cambridge (1999)
5. Oyarzún, J.M.: *Pigment Processing Physico-chemical Principles*. Vincentz Verlag, Hannover (2000)
6. Bauxbaum, G. (ed.): *Industrial Inorganic Pigments*. WILEY-VCH Verlag GmbH, Weinheim (1998)
7. Chen, X.S., Mao, S.S.: Titanium dioxide nanomaterials: synthesis, properties, modifications, and applications. *Chem. Rev.* **107**, 2891–2959 (2007)
8. Moezzi, A., McDonagh, A.M., Cortie, M.B.: Zinc oxide particles: synthesis, properties and application. *Chem. Eng. J.* **185–186**, 1–22 (2012)
9. Klapiszewska, B., Krysztafkiewicz, A., Jesionowski, T.: Highly dispersed green silicate and oxide pigments precipitated from model systems of postgalvanic waste. *Environ. Sci. Technol.* **37**, 4811–4818 (2003)
10. Krysztafkiewicz, A., Klapiszewska, B., Jesionowski, T.: Precipitated green pigments: products of chromate postgalvanic waste utilization. *Environ. Sci. Technol.* **42**, 7482–7488 (2008)
11. Wypych, G.: *Handbook of Fillers*, 3rd edn. ChemTec Publishing, Toronto (2010)
12. Geddes, C.D., Apperson, K., Birch, D.J.S.: New fluorescent quinolinium dyes – applications in nanometer particle sizing. *Dyes Pigments* **44**, 69–74 (2000)
13. Malinge, J., Allain, C., Brosseau, A., Audebert, P.: White fluorescence from core-shell silica nanoparticles. *Angew. Chem. Int. Ed.* **51**, 8534–8537 (2012)
14. Jesionowski, T., Przybylska, A., Kurc, B., Ciesielczyk, F.: The preparation of pigment composites by adsorption of C.I. Mordant Red 11 and 9-aminoacridine on both unmodified and aminosilane-grafted silica supports. *Dyes Pigments* **88**, 116–124 (2011)
15. Jesionowski, T., Przybylska, A., Kurc, B., Ciesielczyk, F.: Hybrid pigments preparation via adsorption of C.I. Mordant Red 3 on both unmodified and aminosilane-functionalised silica supports. *Dyes Pigments* **89**, 127–136 (2011)
16. Siwińska-Stefańska, K., Nowacka, M., Kołodziejczak-Radzimska, A., Jesionowski, T.: Preparation of hybrid pigments via adsorption of selected food dyes onto inorganic oxides based on anatase titanium dioxide. *Dyes Pigments* **94**, 338–348 (2012)
17. Wu, G., Koliadima, A., Her, Y.-S., Matijević, E.: Adsorption of dyes on nanosize modified silica particles. *J. Colloid Interface Sci.* **195**, 222–228 (1997)
18. Parida, S.K., Mishra, B.K.: Adsorption of styrylpyridinium dyes on polyethyleneglycol-treated silica. *Colloids Surf. A* **134**, 249–255 (1998)
19. Binkowski, S., Jesionowski, T., Krysztafkiewicz, A.: Preparation of pigments on modified precipitated silicas. *Dyes Pigments* **47**, 247–257 (2000)
20. Harris, R.G., Wells, J.D., Johnson, B.B.: Selective adsorption of dyes and other organic molecules to

- kaolinite and oxides surfaces. *Colloids Surf. A* **180**, 131–140 (2001)
21. Neumann, M.G., Gessner, F., Schmitt, C.C., Sartori, R.: Influence of the layer charge and clay particle size on the interactions between the cationic dye methylene blue and clays in an aqueous suspension. *J. Colloid Interface Sci.* **255**, 254–259 (2002)
 22. Rytwo, G., Nir, S., Crespín, M., Margulies, L.: Adsorption and interactions of methyl green with montmorillonite and sepiolite. *J. Colloid Interface Sci.* **222**, 12–19 (2000)
 23. Hajjaji, M., Kacim, S., Alami, A., Bouadili, A.E., Mauntassir, M.E.: Chemical and mineralogical characterization of a clay taken from the Moroccan Meseta and a study of the interaction between its fine fraction and methylene blue. *Appl. Clay Sci.* **20**, 1–12 (2001)
 24. Bujdák, J., Iyi, N., Hrobáriková, J., Fujita, T.: Aggregation and decomposition of a pseudocyanine dye in dispersions of layered silicates. *J. Colloid Interface Sci.* **247**, 494–503 (2002)
 25. Bujdák, J., Iyi, N.: Visible spectroscopy of cationic dyes in dispersions with reduced-charge montmorillonites. *Clays Clay Minerals* **50**, 446–454 (2002)
 26. Jesionowski, T., Binkowski, S., Krysztafkiewicz, A.: Adsorption of the selected organic dyes on the functionalized surface of precipitated silica via emulsion route. *Dyes Pigments* **65**, 267–279 (2005)
 27. Jesionowski, T., Andrzejewska, A., Krysztafkiewicz, A.: Adsorption of basic dyes from model aqueous solutions onto novel spherical silica support. *Color Technol.* **124**, 165–172 (2008)
 28. Winnik, F.M., Keoshkerian, B., Fuller, J.R., Hofstra, P. G.: New water-dispersible silica-based pigments: synthesis and characterization. *Dyes Pigments* **14**, 101–112 (1990)

Pigments

► Colorant, Natural

Plasmonics

► Surface Plasmons

Polychromatism

► Iridescence (Goniochromism)

Polychromy

José Luis Caivano

Secretaria de Investigaciones FADU-UBA,
Universidad de Buenos Aires, and Conicet,
Buenos Aires, Argentina

Definition

Polychromy refers to the combination of many colors in a visual scene, whether it is a natural landscape or a man-made arrangement of things, for instance, a piece of art, a design, an object, a building, an urban landscape, etc. The word is formed by the Greek words *poly* (many) and *chroma* (color).

Overview

Colors can be divided into two separate classes: *chromatic colors* (such as red, green, blue, yellow, and any other tone), and *achromatic colors* (such as white, black, and grays). The characteristic of chromatic colors is that they have a definite hue, in addition to a certain saturation and lightness. Achromatic colors do not have a definite hue, and their saturation has a zero value; they are only distinguished by their lightness. When many chromatic colors of different hue appear together, this is called a *polychromatic* composition. If a composition uses colors of different saturation and lightness but of the same hue (for instance, different kinds of yellow that could be lighter or darker and more or less saturated), then this arrangement is said to be a *monochromatic* color composition or a monochrome arrangement (*mono*: one; *chroma*: color). If a composition uses only gray colors of different lightness (which may also include black and white), it is called an *achromatic* composition.

While it may seem from the Greek etymology that the words *color* and *chroma* are equivalent, modern conceptions make a distinction between them. *Color* refers to a visual sensation in general, which includes not only the categories of red,

green, blue, yellow, etc. but also gray, black, and white. The term *hue* is used more specifically to name the quality of a chromatic color that makes yellow look different to red, red to green, green to blue, etc. In opposition, gray colors, black, and white have no hue. In modern usage, *chroma* is considered a synonym of saturation, i.e., the strength, the colorfulness, or the degree of purity of a certain hue. As previously said, gray colors, black, and white have no chroma (saturation), or have chroma zero. However, in order to avoid confusion with the original Greek etymological meaning (*chroma* = color), the term *saturation* is recommended, rather than *chroma*. The concept of *chromaticity* includes the two parameters: hue (dominant wavelength) and saturation (purity), such as in the CIE system of 1931 (also called chromaticity diagram). That is to say, hue and saturation define the chromaticity of a color, independently of its lightness.

Then, it has to be taken into account that the word *color* is often used with two different semantic extensions: with a narrow sense, meaning “chromatic color,” and with a wide sense, including also grays, white, and black as color sensations. The narrow meaning is evident when some people say that black is not a color (because it is considered “the absence of color”), white is not a color (because it is “the sum of all colors”), and grays are not colors either. From a more comprehensive point of view, and in modern usage, this is a misconception.

All the previous clarifications are important in order to understand that *polychromy* means not just “many colors” but many chromatic colors of different hue. Polychrome compositions may include, of course, achromatic colors, but their presence must be accompanied by chromatic colors of different hue in order to be truly polychromatic.

Polychromy in Color Harmony

From the quantity of hues that take part in a color composition or arrangement, Antal Nemcsics states five types of harmonies ([1], pp. 274–280):

- *Monochrome harmonies*, with colors taken from the same hue section in a color order system
- *Dichrome harmonies*, with colors taken from two sections of hue, preferably complementary hues
- *Trichrome harmonies*, with colors taken from three different hues, which may be evenly spaced in the whole chromatic circle or grouped forming a domain in one sector of it
- *Tetrachrome harmonies*, with colors taken according to regular series from four hue triangles, which may be in four well-separated domains in the chromatic circle, in only one domain, in a group of three hues in one domain against one hue opposite to them, or in two complementary pairs
- *Polychrome harmonies*, with colors belonging to more than four different hues, preferably making groups of hues in definite domains

Polychromatic Compositions Along History

The use of polychromy in the visual arts, decoration, and architecture has a long history. The small book by Thomas Goodwin reviews some of the earlier examples and concentrates mainly around the Middle Ages and the religious architecture in England [2]. Goodwin starts the introduction with a definition and then refers to the practice of polychromy in Egypt, Persia, and other antique civilizations:

The art of decorating the interiors of buildings with colour and gilding —Polychromy, as it is sometimes called,— is one which has been practised in all ages and among all nations. Its universal use among all people, as well barbarous as refined, is a satisfactory proof that it is founded upon the best and most natural principles. ([2], p. 1)

The Greeks painted their temples and family dwellings with many vivid colors of different hue, such as red, yellow, and blue. The idea that Greek temples had just the appearance given by white marble has been superseded long time ago, even when for some people it is still a novelty. Greeks started making wooden temples, and of course

they painted them not only for decorative reasons but also to protect the wood. This usage created a style, and centuries later in the Greek culture and artistic practice, when the wood was replaced by marble for building temples, the custom of using colors applied by paints remained. And then, Greeks painted over the marble, producing a polychromatic architecture. The painted colors faded with time, the pigments disappeared, and for centuries the buildings remained as ruins of white marble for the later generations.

By the middle of the nineteenth century, Jacques-Ignace Hittorff succeeded in conveying the attention toward the discovery, made previously to him but neglected for many years, that the Greek architecture was not white – as it was believed for centuries based on the color of the ruins. It was polychromatic. Hittorff published these findings in his book of 1851 on the polychromy of Greek architecture, where he made a case of the temple of Empedocles in Selinus, Sicily [3]. The more general acceptance of this evidence came to change a long-held view about the Greek sense of beauty and harmony. However, it took some time to change the traditional notion, and the neoclassicist architects of the nineteenth century continued to make neoclassic buildings (based on the orders of Greek architecture) in gray, white, or with a monochromatic appearance. As a practicing architect, Hittorff was an exception, since he made polychromatic buildings [4].

Polychromy was also present in statuary in antique Greece. At the beginning of the nineteenth century, Quatremère de Quincy presented a detailed essay on the taste for polychrome sculpture among Greeks and Romans [5]. More recent studies provide additional evidence that Greek artists used to paint with referential colors the marble statues, imitating the colors of skin, hair, representing the colors of clothes, etc. The analysis by Manzelli starts from the first studies on the polychromy of Greek and Roman sculptures, goes through the pigments and painting techniques of statues, and comes to consider the symbolic meanings of color in the ancient classical culture [6].

Roman culture inherited the practice of polychromy from the Greeks and applied it also to

statuary and architecture. In the present times, there is still physical evidence of the polychromy used in the interior of houses and buildings in the old Roman city of Pompeii, which remained preserved for centuries under the ashes of the Vesuvius volcano.

There are evidences that the first Christians decorated the catacombs with color. This practice naturally continued with the first Paleochristian churches and during the Romanic style, and acquired a high degree of perfection in the Byzantine period, particularly with the mosaics. This technique of employment of polychromy has made possible the preservation of beautiful abstract decorations and figurative representations over a 1000 years until nowadays.

During the Middle Ages the polychromy in architecture was particularly prominent in the stained glass windows of the cathedrals. The colored windows also remained until now without deterioration. But, even when it is not always visible today, because pigments have faded, the interior of Gothic cathedrals exhibited architectural elements, such as columns and arches, painted in many colors, in addition to polychrome sculptures. What seems more surprising, from recent discoveries, is that not only the interior spaces but also the facades of many Gothic cathedrals were painted with polychromatic combinations. A restoration of the western facade of the Amiens cathedral made in the 1990s revealed that it was originally painted in various colors and that by the thirteenth century it had a polychromatic appearance.

In the Renaissance, most of the practice of polychromy applied directly on structural or decorative elements of architecture started to be neglected, and was partially replaced by representative frescoes and paintings on the walls and ceilings that endorsed architectonic spaces with color and images “borrowed” from the art of painting. According to Leon Solon, it was in this period that color began to be avoided and disassociated from architecture ([7], pp. 9–10). Even when the Renaissance implicated a classical revival, at this time, around the fifteenth century, very little or nothing was known about the polychromy of the Roman and Greek classical

architecture, because the only remnants were achromatic or monochromatic ruins, and archeological studies had to wait until the nineteenth century to flourish. However, evidence of polychromatic architecture of the medieval churches was supposedly present in the fifteenth century, but since the Renaissance was also a strong reaction against the Gothic style, this may have been another reason for abandoning color. Solon presents various other reasons and makes some speculations about this feature of Renaissance architecture ([7], pp. 11–15).

Owen Jones and Gottfried Semper were two nineteenth-century architects who deserve a special mention due to their contributions to the study of polychromatic architecture. In 1852, Jones published *An Attempt to Define the Principles which Should Regulate the Employment of Colour in the Decorative Arts*, in which he makes 22 propositions or rules for the use of color [8]. Jones was also responsible for coloring a Greek monument exhibited in the Crystal Palace, and he did this task based upon the researches of his time. In a book published in 1854, he explains and defends the criteria with which he selected and applied colors to the different parts of the monument [9]. An appendix of the book includes a fragment written by Gottfried Semper on the origin of polychromy, reviewing cultures such as Indians, Jews and Phoenicians, Greeks, Assyrians, Persians, Egyptians, Chinese, etc. Semper sets forth the hypothesis that this origin relies in textiles, which constitute the oldest manufactures used as covering, shelters, fences, or partitions of space in most cultures. Since textiles were usually dyed in many colors, it is supposed that this practice moved in a natural way to architecture, when textiles were replaced by strongest materials in order to build living spaces ([9], pp. 47–53). *The Grammar of Ornament*, published by Jones in 1856, constitutes an extraordinary source of historical color designs, systematically arranged and covering different geographies and periods, from the ancient times to the early seventeenth century [10].

Another source of information on polychromatic decoration, in this case concentrated on the medieval-style architecture, and particularly on

the Gothic revival and Victorian style of the nineteenth century in England, is the book published in 1882 by the Scottish architects and brothers George and William Audsley, with plenty of illustrations in the form of colored plates [11].

Architectural Polychromy in the Twentieth Century

The Case of Le Corbusier

Without any doubt, the richest and most complex personality of twentieth-century architecture has been Le Corbusier (Charles Edouard Jeanneret). His relationship with color has been complex too, and sometimes contradictory [4]. Between Bruno Taut, uncompromising defender and propagandist of polychrome architecture, and Walter Gropius, who theorizes about color but in his works uses it in extremely moderate doses, Le Corbusier appears to be midway. Precisely, he maintains some differences with Taut regarding color. About the chromatic audacity of Taut, Le Corbusier goes on to say, “My God, Taut is color blind!”

The early writings about color of Le Corbusier were published in the articles on purism and cubism in the journal *L'Esprit Nouveau*, in collaboration with the painter Amédée Ozenfant. A 1918 article reads,

The idea of form precedes that of color. The form is preeminent, color is but one of its accessories. Color depends entirely of the material shape: the concept of sphere, for instance, precedes the concept of color; it is conceived as a colorless sphere, a colorless plane, color is not conceived independently of some support. Color is coordinated with form, but the reciprocal is not true. ([12], p. 42)

Some other articles published in 1921, 1923, and 1924 in the same journal proceed more or less in the same vein, that is, denying any importance that color might have in the construction of space in painting [13]. The curious fact is that a few years later, in his writings on architectural polychromy of 1931, Le Corbusier seems to have changed his mind completely, to the extent of quoting and agreeing with Fernand Léger, who said, “Man needs colors to live, it is an element

as necessary as water and fire.” In addition, Le Corbusier describes examples of his own use of color in order to drastically change the spatial perception of architecture, as in the neighborhood designed and built in Pessac [14].

In his monograph written for the exhibition of the *Pavillon des Temps Nouveaux*, Le Corbusier includes a chapter entitled “Polychromy = Joy,” in which he associates the creative ages of architecture with the vitality of chromatic color and relates the stagnant academicism to sad gray ([15], pp. 22–23). It seems that in both his theories and his works, Le Corbusier evolved toward a more conscious and thorough consideration of the power that color has to modify the spatial environment. This is especially evident in the buildings projected and built after World War II, in what is called his “brutalist” period.

As with the works by other architects of the modern movement, Le Corbusier’s architecture (which was reproduced in black-and-white photography in its time) is being reconsidered in the light of recent restorations and studies. The *Ville Savoye*, a paradigm of modern architecture of the 1920s and 1930s, was not white (as it appeared in many publications) but developed a complex polychromatic scheme [16].

Le Corbusier’s text on architectural polychromy, written in the 1930s, remained as an unpublished manuscript. The first edition of this text, including also the color collections produced for the Salubra wallpaper company, was edited by Arthur Rüegg and published in 1997 [14, 17]. As for the polychromy in the purist architecture of Le Corbusier, and his evolution, see also the recent book by Jan De Heer [18].

Polychromy in Modern Architecture and Beyond

Apart from the particular case of Le Corbusier, there are many other examples of polychromy in architecture in the twentieth century. At the beginning of the century, the Art Nouveau and Art Deco styles were particularly profuse with colors. The Neoplasticist movement is recognized by the use of pigmentary primary colors red, yellow, and blue in high saturation and combined with black and white in different spatial planes. The best

examples of Neoplasticist architecture and design can be found in the works by Gerrit Rietveld. The followers of the organicist architecture (whose paradigm is Frank Lloyd Wright), supporters of not covering surfaces with paints but leaving the materials to express their inherent color, were not less aware of the value of color. Perhaps with the aim of providing a rich polychromy without using paints, Wright made a plentiful use of stained glass windows, as a way of “painting” with light.

The main idea behind the polychromy of modern architecture is that color is not used with a decorative purpose. Instead, it should appear to enhance or modificate the space, to express functions, or to differentiate materials or construction elements. Within some architectural trends or contexts of the modern movement it was evident already in the early 1960s that color entailed new solutions but also new problems. It is for this reason that the Scientific and Technical Center for Building (Centre Scientifique et Technique du Batiment) in France organized a seminar on architectural polychromy under the direction of Guy Habasque, and with the participation of color consultants Antoine Fasani, Jacques Fillacier, Bernard Lassus, Georges Patrix, and Erasme Saffre. The discussions of this seminar were published in 1961 in an issue of the *Cahiers* of this Center [19].

After the modern movement (represented mainly by Le Corbusier, Frank Lloyd Wright, Mies van der Rohe, and Walter Gropius), the postmodernist reaction of the 1970s and 1980s was accompanied by a different view about color and polychromatic compositions. Postmodernism brought about a host of architects concerned with the references to history and the environment, and polychromy in architecture also acquired a new meaning under these orientations. As examples, the works by Charles Moore, Robert Venturi, Michael Graves, and Stanley Tigerman (in the USA), Paolo Portoghesi and Aldo Rossi (in Italy), Aldo van Eyck (in the Netherlands), and Mario Botta (in Switzerland) can be mentioned. In Latin America, there have been also outstanding cases of recognized masters of architecture that are paradigmatic for their conception and application of polychromy. Three of them deserve a special mention: Carlos Raúl Villanueva

in Venezuela, Luis Barragán in Mexico, and Clorindo Testa in Argentina [4].

Cross-References

- [Anchoring Theory of Lightness](#)
- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE Chromaticity Coordinates \(xyY\)](#)
- [Color Circle](#)
- [Color Harmony](#)
- [Color Order Systems](#)
- [Complementary Colors](#)
- [Mosaics](#)
- [Pigments](#)
- [Primary Colors](#)
- [Unique Hues](#)

References

1. Nemesics, A.: *Colour Dynamics*. Akadémiai Kiadó, Budapest (1990) (1993)
2. Goodwin, T.G.: *A Short Account of the Art of Polychrome Historical and Practical*. Joseph Masters, London (1860)
3. Hittorff, J.-I.: *Restitution du temple d'Empédocle à Sélinonte, ou l'architecture polychrome chez les Grecs*. Firmin Didot, Paris (1851)
4. Caivano, J.: Research on color in architecture and environmental design: brief history, current developments, and possible future. *Color Res Appl* **31**(4), 350–363 (2006)
5. Quatremère de Quincy, A.: *Le Jupiter olympien; ou l'art de la sculpture antique . . . ouvrage qui comprend un essai sur le goût de la sculpture polychrome . . . , et l'histoire de la statuaire en or et ivoire chez les grecs et les romains*. Firmin Didot, Paris (1814)
6. Manzelli, V.: *La policromia nella statuaría greca arcaica*. L'Erma di Bretschneider, Rome (1994)
7. Solon, L.V.: *Polychromy: Architectural and Structural Theory and Practice*. The Architectural Record, New York (1924)
8. Jones, O.: *An Attempt to Define the Principles Which Should Regulate the Employment of Colour in the Decorative Arts*. G. Barclay, London (1852)
9. Jones, O.: *An Apology for the Colouring of the Greek Court in the Crystal Palace, with Arguments by G.H. Lewes and W. Watkiss Lloyd, and a Fragment on the Origin of Polychromy by Professor Semper*. Crystal Palace Library and Bradbury & Evans, London (1854)
10. Jones, O.: *The Grammar of Ornament*. Day and Son, London (1856)
11. Audsley, W.J., Audsley, G.A.: *Polychromatic Decoration as Applied to Buildings in the Medieval Styles*. Henry Sotheman & Co., London (1882)
12. Ozenfant, A., Le Corbusier: *Après le Cubisme*. Edition des Commentaires, Paris (1918). Spanish translation, *Después del Cubismo*. In: Pizza, A. (ed.) *Acerca del Purismo. Escritos 1918–1926*, pp. 8–47. El Croquis Editorial, Madrid (1994)
13. Ozenfant, A., Le Corbusier: *Le Purisme*. *L'Esprit Nouveau* 4 (1921). *Nature et creation*. *L'Esprit Nouveau* 19 (1923). *Idées personnelles*. *L'Esprit Nouveau* 27 (1924). Spanish translation. In: Pizza, A. (ed.) *Acerca del Purismo. Escritos 1918–1926*. El Croquis Editorial, Madrid (1994)
14. Le Corbusier: *Polychromie architecturale*. Foundation Le Corbusier, Paris, manuscript (1931). Posthumous edition. In: Rüegg, A. (ed.) *Polychromie architecturale. Les claviers de couleurs de Le Corbusier de 1931 et de 1959*. Birkhäuser, Basel (1997)
15. Le Corbusier: *Des canons, des munitions? Merci! Des logis . . .* S.V.P. Editions de L'Architecture d'Aujourd'hui, Paris (1938)
16. Cram, N.: It was never white, anyway: despite his reputation as the godfather of white architecture, Le Corbusier developed complex ideas about color. *Architecture (USA)*, February, 88–91 (1999)
17. Schindler, V.M.: *Prefabricated rolls of oil paint: Le Corbusier's 1931 colour keyboards*. In: Caivano, J. (ed.) *AIC 2004 Color and Paints, Proceedings of the Interim Meeting of the International Color Association, Porto Alegre*, pp. 198–202 (2015). http://www.aic-color.org/congr_archivos/aic2004proc.pdf. Accessed 10 July 2015
18. De Heer, J.: *The Architectonic Colour: Polychromy in the Purist Architecture of Le Corbusier*. 010 Publishers, Rotterdam (2009)
19. CSTB, Centre Scientifique et Technique du Batiment: *La Polychromie Architecturale*. Cahier Nr. 423. CSTB, Paris (1961)

Preference for White in Asia

- [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)

Prelinguistic Color Categorization/Categorical Perception

- [Infant Color Categories](#)

Premetallized Acid Dyes

► [Dye, Metal Complex](#)

pRGCs

► [Melanopsin Retinal Ganglion Cells](#)

Primary Colors

Paul Green-Armytage
School of Design and Art, Curtin University,
Perth, Australia

Synonyms

[Basic colors](#); [Elementary colors](#); [Fundamental colors](#); [Primitive colors](#); [Principal colors](#); [Pure colors](#); [Simple colors](#); [Unique hues](#)

Definition

There are two kinds of primary colors that are often confused. They can be called “physical primaries” and “visual primaries.” Sets of three “physical primaries” in the form of lights, paints, or inks can be mixed to produce a comprehensive range of other colors. These sets are “primary” because they have been found to deliver the most extensive and useful range of other colors. Colors are also recognized as “primary” by virtue of their appearance. These are the “visual primaries.”

Introduction

Much of the confusion that surrounds the topic of color can be exposed in a discussion about primary colors. Primary colors, for some people, are just colors that are particularly vivid. For those

who work with color, there are two kinds of primary colors that are often confused. They can be called “physical primaries” and “visual primaries.”

Sets of three “physical primaries” in the form of lights, paints, or inks can be mixed to produce a comprehensive range of other colors. These sets are “primary” because they have been found to deliver the most extensive and useful range of other colors. The particular lights or paints that work best can be recognized by their appearance: for lights, they appear orange-red, yellow-green, and violet-blue, but they are usually identified simply as red, green, and blue. The mixing process for lights is additive. For paints and inks, the mixing process is subtractive. Best results are obtained with paints or inks that appear purplish red, slightly greenish blue, and yellow. These are best identified as magenta, cyan, and yellow, but they are often loosely identified as red, blue, and yellow. The reds and blues for additive mixing appear different from those for subtractive mixing, but confusion can arise because of the range of different “reds” and “blues,” where an orangish red and a purplish red are both “red.”

Colors are also recognized as “primary” by virtue of their appearance. These are the “visual primaries.” In the continuous circle of hues, there are four that are recognized as “unique”: a yellow that appears neither greenish nor reddish, a red that appears neither yellowish nor bluish, a blue that appears neither reddish nor greenish, and a green that appears neither bluish nor yellowish. These visual primaries serve as conceptual reference points and, together with white and black, anchor the Natural Color System (NCS), which orders colors according to how they appear. Confusion can arise when visual primaries are thought of as the same thing as physical primaries. Confusion can be overcome when the distinction is maintained between visual appearances on the one hand and, on the other hand, the lights, paints, and inks which give rise to those appearances.

Before going deeper into the two main questions – what are the primary colors, and what makes them “primary” – it will be helpful to consider the nature of color itself.

The Nature of Color

Color means different things to different people. There are the colors that physicists measure, the colors that chemists analyze, the colors that psychologists study, the colors that artists brush onto paper or canvas, and the colors that people talk about in general conversation. All these may be related, but they are not the same kind of thing.

It is possible to identify seven different “kinds of color” [1], but for present purposes a broad distinction can be made between the physical aspects of color and color as it is experienced in perception. In general conversation a rose or the setting sun might be described as red, and for most people that red would be a physical property of the rose or the sunlight. They might also recognize as red the subjective visual experience they have when they look at the rose or the setting sun. The one word “red” is used for two different kinds of thing. The lack of distinction between the physical and visual aspects of color is responsible for much of the confusion that exists in connection with color in general and with primary colors in particular.

Confusion Between the Physical and Visual Aspects of Color

The confusion is evident in what might be assumed to be an authoritative definition. In the *Concise Oxford Dictionary* [2] a primary color is defined as “any of the colours red, green, and blue, or (for pigments) red, blue, and yellow, from which all other colours can be obtained by mixing.” In this definition the meaning of the word “colour” changes halfway through. Pigments are physical substances. A particular pigment might appear red, but that redness is not the same thing as the pigment itself. It can help to remove the confusion if (physical) and (visual) are added after the word “color” to clarify the distinction between color meaning something physical on the one hand and color meaning visual experience on the other. Then that definition could be reworded: “any of the colors (physical) . . . from which all other colors (visual) can be obtained by mixing.”

The same change of meaning can be recognized in another definition, this from *The Color*

Compendium [3]: “Simply stated, primary colors are ones that cannot be mixed or formed by any combination of other colors.” This could also be reworded: “. . . primary colors (visual) are ones that cannot be mixed or formed from any combination of other colors (physical).”

Evolution of Ideas about Primary Colors

Although the definitions above confuse the physical with the visual aspects of color they do illustrate the two main ideas about what should be understood by primary colors. One is that primary colors are physical and are used in mixing processes; the other is that primary colors are visual and serve as perceptual and cognitive reference points. The evolution of these ideas can be traced through history.

Greek Philosophers

In the fifth century BC Empedocles recognized four principal colors as corresponding with the four elements: black, white, red, and green with earth, air, fire, and water [4, p. 15]. For Aristotle white and black, combined in different proportions, were the source of other colors. In a line between white and black he identified five other principal colors – yellow, red, violet, green, and blue. “. . . from these all others are derived by mixture.” [5, p. 698].

Leonardo da Vinci

Closer to the present time, but with echoes of Empedocles, Leonardo da Vinci wrote that

white is the first among the simple colors, and yellow the second, green the third of them, blue is the fourth, and red is the fifth, and black the sixth. And white is given by light, without which no color may be seen, yellow by earth, green by water, blue by air and red by fire, and black by darkness . . . [6, p. 70]

These six colors would seem to correspond with visual primaries as reference points. On the mixture of colors Leonardo wrote,

I call simple colours those which are not compounded and cannot be compounded by means of a mixture of other colours . . . After

black and white come blue and yellow, then green and tan, that is to say, tawny, or if you wish to say ochre, and then deep purple and red. These are the eight colours and there are no more in nature, and with these I begin the process of mixing, first with black and white, and then black with yellow . . . [6, p. 72]

It seems likely that these eight colors represent the pigments which Leonardo used in his paintings. His eight simple colors would correspond with physical primaries, but their number would have had more to do with the limitations of available pigments and his own approach to the practicalities of painting than any of the theories of primary colors that are more familiar today.

Robert Boyle

Something much closer to present ideas about mixing a comprehensive range of colors (visual) from a limited set of colors (physical) can be found in the writings of Robert Boyle:

. . . there are but few Simple and Primary Colours (if I may so call them) from whose Various Compositions all the rest do as it were Result. For though Painters can imitate the Hues (though not always the Splendor) of those almost Numberless differing Colours that are to be met with in the Works of Nature, and of Arte, I have not yet found, that to exhibit this strange Variety they need employ any more than *White*, and *Black*, and *Red*, and *Blew*, and *Yellow*; these *five*, Variously *Compounded*, and (if I may so speak) *Decompounded*, being sufficient to exhibit a Variety and Number of Colours, such, as those that are altogether Strangers to the Painters Pallets, can hardly imagine. [7, pp. 219–220]

There are two significant features of this passage. First, it shows that the possibility of producing a complete sequence of hues with only three paints (physical colors) was well established. Second, it introduces, by implication, the concept of color gamut. A gamut is the range of colors (visual) that can be produced with a given set of colors (physical). Although with Boyle's *Red*, *Blew*, and *Yellow* it was possible to produce all the hues, it was not possible to produce all the colors – colors of greater “splendor” were out of reach. With *Red* and *Blew* it was possible to produce a range of purples but not the vivid purples of some flowers. The word “hue” is often used as a synonym for “color,” but in a discussion of

primary colors it is important to recognize that hue is just one of the dimensions of color. Boyle's “splendor” is another dimension which refers to how vivid a color may be. Words used today for this dimension include intensity and saturation as well as the more specialist terms “chroma” and “chromaticness.”

Jakob Christoph Le Blon

While Boyle was a scientist, Jakob Christoph Le Blon was an artist. Le Blon developed a system of printing with transparent inks from three or four plates on white paper that is essentially the same as the process used by printers today. Three plates, one inked with yellow, one with red, and one with blue, were printed one over the other. “Painting can represent all *visible* Objects, with three Colours, *Yellow*, *Red*, and *Blue*; for all other Colours can be compos'd of these *Three*, which I shall call *Primitive*” [8]. Sometimes Le Blon used a fourth plate, inked with black, much as printers today use black to deepen the colors and sharpen the image.

Isaac Newton

Isaac Newton experimented with prisms and demonstrated that different wavelengths of light are refracted by a prism to a greater or lesser extent and that the refracted light appears in different colors. When projected onto a white surface the refracted light is spread out to reveal the colors of the spectrum. In Newton's words, “The Light whose Rays are all alike Refrangible, I call Simple, Homogeneal and Similar . . . The Colours of Homogeneal Lights, I call Primary, Homogeneal and Simple . . .” [9, p. 4]. For Newton, therefore, a primary color was light of a single wavelength.

The spectrum can be divided and subdivided many times over, so the number of primary colors in this physical sense is very large indeed. But Newton chose to name just seven: “The original or primary colours are red, yellow, green, blue and a violet-purple together with orange, indigo and an indefinite variety of intermediate gradations” (Newton, quoted in [10, p. 225]). Newton developed a circular diagram, divided into seven segments, which could be used to determine the likely appearance of a given mixture of his primary lights.

Thomas Young

The fact that it was possible to mix a comprehensive range of colors (visual) from a limited set of primary colors (physical) led George Palmer and Thomas Young, working independently, to similar insights about what there must be in the eye that enables humans to experience such a wide range of color appearances. Young's speculation has been confirmed in its essential details:

Now, as it is almost impossible to conceive each sensitive point of the retina to contain an infinite number of particles, each capable of vibrating in perfect unison with every possible undulation, it becomes necessary to suppose the number to be limited, for instance, to the three principal colors, red, yellow, and blue . . . and that each of the particles is capable of being put in motion less or more forcibly by undulations differing less or more from perfect unison . . . each sensitive filament of the nerve may consist of three portions, one for each principal color. [11, p. 147]

James Clerk Maxwell

Where Le Blon had demonstrated that an image could be reproduced in a print by the subtractive mixing of three transparent inks, James Clerk Maxwell demonstrated how an image could be reproduced by the additive mixing of colored light. The first color photograph, taken by Thomas Sutton, was used as an illustration for a lecture by Maxwell in 1861 [12, p. 321]. A ribbon had been photographed successively through a red, a green, and a blue filter. The three photographs were then projected through corresponding filters so that the three images were superimposed on the screen.

Maxwell also developed a means of measuring color. When a disk with segments in different colors is spun at high speed the colors blend and a new color is seen. Three paper disks, one in each of the additive primary colors (physical), are interleaved in such a way that the visible proportions of each can be adjusted. A disk in the color that is to be measured is fixed over the three interleaved disks. The superimposed disks are then spun in a series of trials, with the proportions of the three primary colored disks being adjusted after each trial, until the combination of three primary colors matches the color to be measured. The measurement is then expressed in terms of the relative

amounts of each of the primary colors needed to achieve the match. In this process colors (physical) and colors (visual) come together. The colored papers are physical, but it requires a visual judgment to establish a match.

Ewald Hering

During the nineteenth century the idea that color vision depends on three different kinds of sensitive cell in the eye, with three corresponding primary colors, became established. It was challenged by Ewald Hering, who used his own observations as a starting point. He concluded that there are six fundamental colors (*Ürfarben*) that must have their counterpart in the nervous system associated with vision. Hering's *Ürfarben* can be called primary colors (visual). The claim is that any color can be described in terms of its relative resemblance to these *Ürfarben* which serve as reference points. They are the basis of the Swedish standard Natural Color System [13]. After black and white Hering defines what are now recognized as the "unique hues":

. . . there are four outstanding loci in the series of hues that make up the closed circle: first the locus of the yellow that shows no remaining trace of redness, and yet reveals no trace of green; second the locus of the blue for which the same is true. These two hues may be called *primary yellow* and *primary blue* (*Ürgelb* and *Ürblau*). Likewise we can name, third, the red, and, fourth, the green that are neither bluish nor yellowish primary red and primary green (*Ürrot* and *Ürgrün*). [14, p. 42]

The CIE System

In 1931 the International Commission on Illumination (CIE) established the system of color measurement that is today's world standard [15, pp. 59–63]. All possible colors (physical) can be plotted on the CIE chromaticity diagram. The position of a color on the diagram is determined by the relative output of three primary lights that, when mixed together, would match the color. The measurement can be expressed in terms of tristimulus values. As with Maxwell's system of spinning disks the matches that underpin the CIE system were initially the result of visual judgments made by human observers. But there is no set of three colored lights (physical) that can

match all possible colors (visual). The most vivid colors (visual) are those of Newton's homogeneous lights – the pure colors of the spectrum. In order to measure such colors it was necessary to reduce their “splendor” by introducing negative values. This led to complexities that were overcome by clever mathematics and the introduction of theoretical “super light sources” that could not be physically realized but that were able to deliver matches for the spectral colors.

Primary Colors in the Twenty-First Century

The different ways in which the term “primary colors” is used today can be related to these ideas that evolved in the past.

Further Problems with Standard Definitions

The definitions of primary colors, quoted in the section “[Confusion Between the Physical and Visual Aspects of Color](#)” above, not only confuse the physical and visual aspects of color, they are also misleading in other ways. As Boyle observed, and the CIE confirmed, there is no set of primary colors (physical) from which all other colors can be obtained by mixing. It is possible to obtain all the hues but not all the colors – only those within a limited gamut. The CIE had to resort to its theoretical super lights. A full gamut of all other colors, right out to Newton's homogeneous lights, can only be achieved in theory, with the CIE super lights, and not in practice.

The second definition, that primary colors cannot be mixed from a combination of other colors, is also misleading. It would be true of the CIE theoretical super lights but not in the case of normal lights, paints, or inks. It would make no sense to claim that primary colors (physical) cannot be mixed from a combination of other colors (physical). The primary colors that do make sense in the context of this definition are visual colors. Leaving aside black and white the visual primaries should be the unique hues – Hering's *Urfarben* – and there is nothing unmixable about

these. A paint that appears yellow can be mixed with paint that appears magenta to produce a red that is neither bluish nor yellowish and with a paint that appears blue to produce a green that is neither bluish nor yellowish.

Primaries for Mixing other Colors in Print and on Screen

Primary colors (physical) for mixing a large range of other colors (visual) have particular application for the reproduction of colored images in print and on the television or computer screen. There is scope for further confusion here because the mixing processes are not the same. When layers of transparent ink are superimposed in a print the result is a subtractive mixture. Additive mixture occurs when pools of colored light are projected to overlap or when the small points of colored light blend on a screen.

It was not until the nineteenth century that the difference between subtractive and additive processes was recognized. Then it became clear that the primaries (physical) for each process were not only different in that pigments and ► [dyes](#) are physically different from colored lights but also that the most useful sets of three colors (physical) appear different. The two sets are listed in the first definition set out in the section “[Confusion Between the Physical and Visual Aspects of Color](#).” Additive primaries are listed first, then the subtractive primaries. They are described as red, green, and blue or (for pigments) red, blue, and yellow. This suggests that the only difference between the two sets is that one set has green and the other set has yellow. However, the lights that work best for additive mixture appear orange-red, yellow-green, and violet-blue, while the inks for subtractive mixture appear purplish red (magenta), slightly greenish blue (cyan), and yellow. The differences between the orange-red and magenta, and between the violet-blue and cyan, are masked by the way the names “red” and “blue” stretch to cover a range of different hues. What makes these two sets primary is that they have been found to deliver the most useful gamuts of other colors (visual). The gamuts could be

increased for each process by adding further inks or lights, but any slight improvement would not justify the very large extra costs that this would involve.

Mixing with a Larger Number of Colors (Physical)

When a color is required in a print that is beyond the reach of the cyan, magenta, and yellow process inks it may be achievable with a mixture of inks formulated according to the Pantone Matching System®. This system has 14 separate inks that can be premixed before being used on the press.

A similar number of different colors (physical) are used to mix the very large range of different colors (visual) that are available for painting the interior and exterior of buildings. The tinting machines for Dulux Australia have 4 main bases and 13 different tinters that add pigments to a base according to given formulas.

Today's artists, like Leonardo before them, need not restrict themselves to a set of three primary colors (physical). They use as many different paints as they need to achieve the results they want.

Primary Confusion

The confusion surrounding the different notions of primary colors, especially that between primary colors (physical) and primary colors (visual), was the subject of a paper that Green-Armytage gave at a conference on color education [16]. This confusion is especially evident in what he calls traditional art school color theory.

Among the most widely used and influential books on color theory for artists are *The Art of Color* [17] and the condensed version *The Elements of Color* by Johannes Itten [18]. Itten's account should be read carefully [17, p. 34, 18, p. 29]:

... let us develop the 12-hue color circle from the primaries – yellow, red and blue. As we know, a person with normal vision can identify a red that is neither bluish, nor yellowish; a yellow that is neither greenish nor reddish; and a blue that is neither greenish, nor reddish.

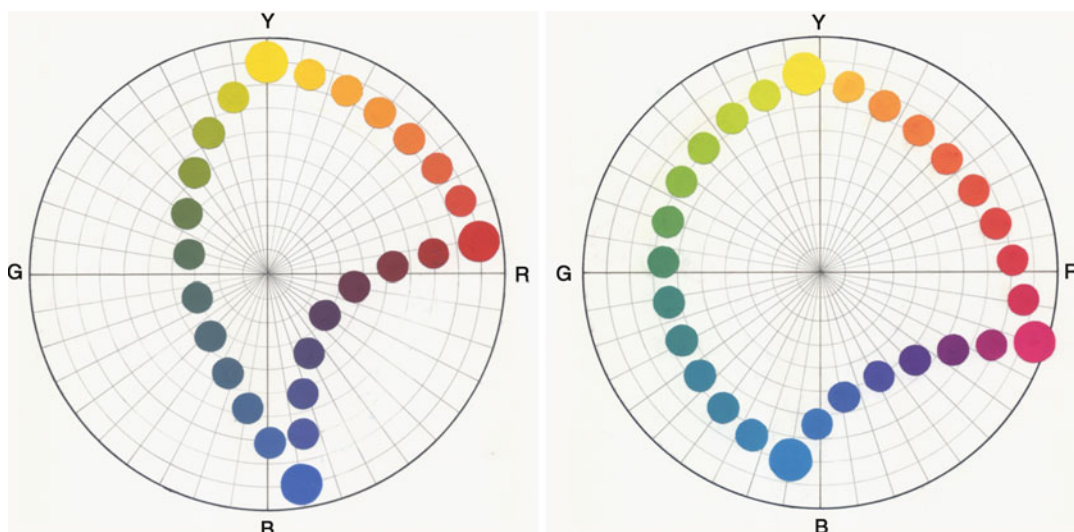
These definitions are reminiscent of Hering's definitions except that green is not included as a primary. However, green is included in the definitions of primary yellow and primary blue. Itten did not (could not?) describe primary blue as neither reddish nor *yellowish* or primary yellow as neither reddish nor *bluish*.

Itten goes on to describe secondary colors as "... three mixed colors, each composed of two primaries." Later in the books [17, p. 136, 18, p. 88] he describes green as "... the intermediate between yellow and blue. According as green contains more yellow or more blue, the character of its expression changes." Did he mean that there could be more or less yellow or blue paint in the mixture which resulted in the perception of green? Or did he mean that in the green percept there might be more or less resemblance to yellow and blue as primary colors (visual)?

It is not clear from the books whether Itten associated his primary yellow, red, and blue with any particular pigments. There is certainly no claim that with three such paints all other colors can be obtained by mixing. But what is clear from experiment is that three paints which might fit Itten's definitions would not work so well for producing a comprehensive range of other colors (visual) as would paints that look the same or similar to the cyan, magenta, and yellow process inks used by printers. Itten's primaries would not be a good choice as primary colors (physical). And without green they are not satisfactory as primary colors (visual).

Plotting the Gamut for a Given set of Colors (Physical)

The ► [color circle](#) of the Natural Color System (NCS) can be used as a diagram on which the appearance of paints can be plotted. The unique hues, as reference points, are spaced at equal intervals round the circle. The achromatic colors – the pure whites, grays, and blacks – are located in the center of the circle. The most vivid colors (visual) that can be imagined are located on the circumference. The less vivid colors (visual) are located between the circumference and the



Primary Colors, Fig. 1 Mixture lines connecting two sets of “primary” paints plotted on the *color circle* of the Natural Color System (NCS)

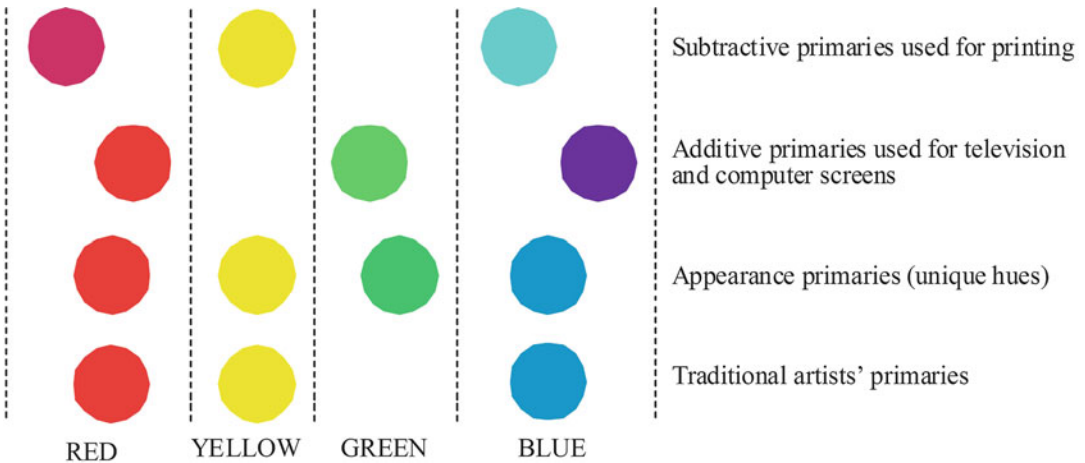
center, closer to the circumference or to the center as they are more or less vivid. Points on the diagram representing two paints can be connected by mixture lines to indicate the changing appearance of the mixtures as they contain more or less of each paint. The mixture lines connecting three paints trace an irregular triangle and mark the gamut of appearances that can be achieved with that set of paints. This makes it possible to compare the relative merits of different sets of primary colors (physical). The gamuts achievable with different sets of “primaries” are illustrated in Fig. 1.

On the left, cadmium yellow pale, cadmium red, and ultramarine are typical of the paints regarded as “primary” in traditional art school color theory. While these paints can deliver all the hues, the greens and purples are very dull. If more vivid greens and purples are required an artist would need to include more paints on the palette. A more extensive gamut can be achieved with the addition of a turquoise or cerulean blue and a vivid magenta such as Bengal rose. The paints used on the right were lemon yellow, Bengal rose, and turquoise blue. These paints appear similar to the yellow, magenta, and cyan process inks as used in printing. Their status as preferred primaries for subtractive mixing is clear from the

more extensive color gamut that they can deliver. Within the mixture lines are the colors (visual) that could be achieved with that set of paints. If a desired color (visual) would be located outside the mixture lines there may be a paint that could be added that would bring that color (visual) within reach [19].

The Role of Color Names in the Confusion of Different Kinds of Primaries

In the section “[Primaries for Mixing other Colors in Print and on Screen](#)” above it is argued that the difference between primaries for additive and subtractive mixing is masked by the way that the names “red” and “blue” are each used to describe a range of different hues. The only point of difference between the two sets of primaries seems to be that green is a primary for additive mixing and yellow for subtractive mixing; both sets include “red” and “blue.” It might seem that the unique hues, as visual primaries, resolve the conflict by including both yellow and green, but this only adds to the confusion. And further confusion comes from the common expectation of art students that they can “mix all the other colors” from the primaries red, yellow, and blue. The red, yellow, and blue of traditional art school color theory are typically defined in terms used for the visual



Primary Colors, Fig. 2 Four sets of primary colors identified loosely as “red,” “yellow,” “green,” and “blue” to show how use of color names contributes to the way the sets are confused

primaries, the unique hues. The four sets of primaries are compared in Fig. 2.

Conclusion

There has been a long history of singling out particular colors as being somehow special. The realization that a comprehensive range of colors (visual) can be mixed from a limited set of paints, inks, ► [dyes](#), or lights led to the development of ways to reproduce fully colored images in prints and projected photographs, techniques pioneered by Le Blon and Maxwell. This realization also inspired ideas about the physiology of the eye and how it is that humans can experience such a great variety of colors.

Concepts of primary colors can help people to work successfully with colors. These concepts can also be confused. The key is to recognize that there are two kinds of primary colors: physical primaries, which take the form of lights, paints, inks, and dyes to be used in mixing a comprehensive range of other colors (visual), and visual primaries as very specific visual experiences – pure white, black, and the four unique hues. By using the visual primaries for reference the physical processes involved in the production and reproduction of colors, which may involve the physical primaries, can be monitored and controlled.

Cross-References

- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [Color Circle](#)
- [Dye](#)
- [Newton, \(Sir\) Isaac](#)
- [Psychologically Pure Colors](#)
- [Unique Hues](#)

References

1. Green-Armytage, P.: Seven kinds of colour. In: Porter, T., Mikellides, B. (eds.) *Colour for Architecture Today*. Taylor and Francis, Abingdon (2009)
2. Thompson, D.E.: *The Concise Oxford Dictionary*. Oxford University Press, Oxford (1995)
3. Hope, A., Walch, M.: *The Color Compendium*. Van Nostrand Reinhold, New York (1990)
4. Beare, J.: *Greek Theories of Elementary Cognition*. The Clarendon Press, Oxford (1906)
5. Aristotle: *Sense and Sensibilia*. In: Barnes, J. (ed.) *The Complete Works of Aristotle*. Princeton University Press, Princeton (1984 [4th c. BC])
6. da Vinci, L.: *Leonardo on Painting*. Yale University Press, New Haven (1989 [1518])
7. Boyle, R.: *Experiments and Considerations Touching Colours*. Johnson Reprint Corporation, New York (1964 [1664])
8. Le Blon, J.C.: *Coloritto*. Van Nostrand Reinhold, New York (1980 [1756])
9. Newton, I.: *Opticks*, 3rd edn. Bell & Sons, London (1931 [1730])

10. McLaren, K.: Newton's indigo. *Color Res. Appl.* **10**(4), 225–229 (1985)
11. Young, T.: On the theory of light and colours. In: Peacock, G. (ed.) *Miscellaneous Works of the Late Thomas Young*. John Murray, London (1972 [1801])
12. Kemp, M.: *The Science of Art*. Yale University Press, New Haven (1990)
13. Hård, A., Sivik, L.: NCS – Natural Color System: a Swedish standard for color notation. *Color Res. Appl.* **6**(3), 129–138 (1981)
14. Hering, E.: *Outlines of a Theory of the Light Sense*, transl. L. M. Hurvich, D. Jameson. Harvard University Press, Cambridge, MA (1964 [1905])
15. Berns, R.: *Billmeyer and Saltzman's Principles of Color Technology*, 3rd edn. Wiley, New York (2000)
16. Green-Armytage, P.: Primary confusion. In: Arnkil, H., Hämäläinen, E. (eds.) *Aspects of Colour*. The University of Art and Design Helsinki UIAH, Helsinki (1995)
17. Itten, J.: *The Art of Color*. Van Nostrand Reinhold, New York (1961)
18. Itten, J.: *The Elements of Color*. Van Nostrand Reinhold, New York (1970)
19. Green-Armytage, P.: Paints as 'magnets' to guide the mixing process. In: Hansuebsai, A. (ed.) *Proceedings AIC 2003 Bangkok Color Communication and Management*, pp. 309–311. Chulalongkorn University, Bangkok (2003)

Primate Color Vision

► Comparative Color Categories

Primitive Colors

► Primary Colors

Principal Colors

► Primary Colors

Printing

► Coloration, Textile

Protanopia

Galina V. Paramei¹ and David L. Bimler²

¹Department of Psychology, Liverpool Hope University, Liverpool, UK

²School of Psychology, Massey University, Palmerston North, New Zealand

Synonyms

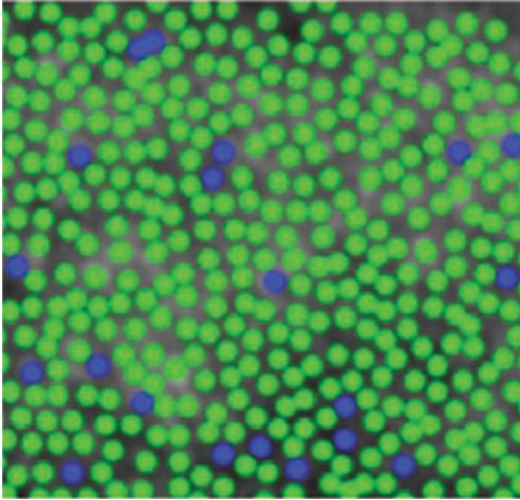
[Color blindness](#); [Congenital color vision abnormality](#); [Daltonism](#); [Dichromacy](#); [“Red blindness”](#); [Red-green deficiency](#); [X-linked inherited color vision deficiency](#)

Definition

Protanopia [from the Greek *protos* (first) + *an* (not) + *opia* (a visual condition)] is a congenital form of severe color deficiency (dichromacy) affecting the red-green opponent color system. Protanopes do not distinguish between colors along a specific direction (protanopic confusion lines) in color space and are able to match all colors using two primaries (unlike trichromats who require three primaries). Protanopia is linked to the X-chromosome and arises from loss or alteration of the gene encoding the opsin of the long-wavelength (L-) photopigment; this malstructure results in the absence of functioning L-cones in the retina. Protanopia follows a recessive pedigree pattern, with incidence of ca. 1 % in the Caucasian male population. It is distinct from protanomaly in which some degree of discrimination among such colors remains.

Retinal Mosaic

At the level of retinal function, protanopia is the absence of L-cones (Fig. 1) (“► [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)” by Stockman), i.e., cone cells expressing a photopigment containing an L-opsin component, tuned for selective sensitivity to longer wavelengths such as orange and red light [1].



Protanopia, Fig. 1 Adaptive optics retinal imaging: Pseudocolor image of the protanopic cone mosaic; all cones are of the M- (“green”) or S- (“blue”) type [Carroll, J., Neitz, M., Hofer, H., Neitz, J., Williams, D.R.: Functional photoreceptor loss revealed with adaptive optics: An alternate cause of color blindness. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8461–8466 (2004); Fig. 4b, p. 8465. Copyright (2004) National Academy of Sciences, U.S.A.]

Genetic Background

Functional loss of the L-cone photoreceptor cells is caused by a mutant cone photopigment gene on the X-chromosome (“► [Genetics of photoreceptors, genetics and color vision deficiencies, genes of cone photopigments, genes and cones](#)” by Rodríguez-Carmona). The gene coding for the L-opsin may be missing altogether from a protanope’s genome; it may be replaced by a second copy of the M-opsin (sensitive to slightly shorter wavelengths), or it may be present but damaged so that the protein is not expressed or cannot be transported to the ciliary segment of cone cells expressing that gene [2].

When present, the L-opsin gene is located on the X-chromosome, followed by one or more copies of the M-opsin gene (“► [Genetics of photoreceptors, genetics and color vision deficiencies, genes of cone photopigments, genes and cones](#)” by Rodríguez-Carmona). This architecture, and the high similarity of the two genes, facilitates mismatching of DNA during meiotic crossover, leading to deleterious mutations of the L-opsin encoding gene [3].

With twoX- chromosomes, females are far less prone to protanopia. Notably, mothers and daughters of protanopes can be heterozygous for protanopia (i.e., carriers of the condition) with one normal and one aberrant X-chromosome; the former is active in a portion of their cone cells, providing enough L-cones to sustain normal trichromacy. However, minor departures from normality are discernible [4], including reduced sensitivity to long wavelengths, known as Schmidt’s sign, and reduced subjective prominence of red-green dissimilarity [5].

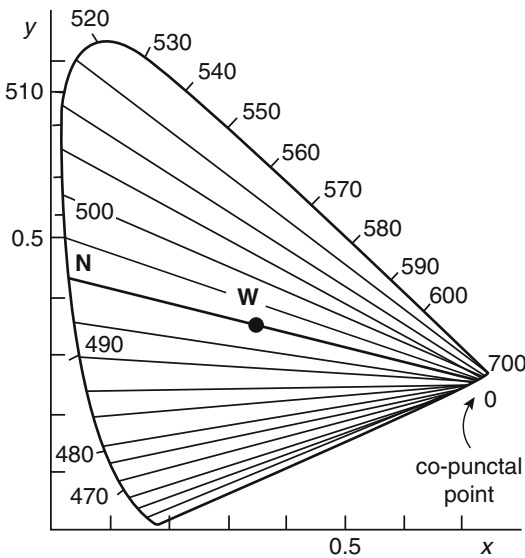
Confusion Lines

L-signal/M-signal comparison in bipolar ganglion color-comparison cells is not present in protanopes (“► [Color Vision, Opponent Theory](#)” by Wuerger, Xiao); hence there is no difference in the appearance of any colors which differ only in the L-M signal. That is, although protanopes lack the L- or “red” cones, they are not “red blind” in a literal sense but rather are red-green blind. Examples of colors confused by protanopes are red, yellow, and green or blue-green, gray, and purple. Plotted in a chromaticity diagram (► [CIE Chromaticity Diagrams, CIE purity, CIE dominant wavelength](#) and ► [CIE Chromaticity Coordinates; xyY; chromaticity coordinates to matrix](#)), the confused colors lie on straight “confusion lines” that converge at the protanope copunctal point (Fig. 2).

Luminous Efficiency, Luminance Effect, and Wavelength Discrimination

In the absence of L-cones tuned to longer wavelengths, protanopes display a distinctive insensitivity to 550–750 nm lights (orange and red) [6]. Their luminosity function has peak sensitivity around 540 nm compared to the maximum near 555 nm of normal trichromats, i.e., it is blue shifted toward the λ_{max} of the M-cone pigment [1, 7]. Phenomenally, reds appear relatively dark for protanopes and are confused with grays and bluish turquoise.

Thus, protanopes attend more to luminance changes than normal trichromats as a color cue, compensating for their poor hue discrimination [8]. Notably, at lower photopic and mesopic



Protanopia, Fig. 2 Confusion lines indicate colors that cannot be distinguished by a protanope. The line going through the white point (*W*) indicates neutral (achromatic) colors. The lines converge at the copunctal point defining the chromaticity of the missing L-cone fundamental. Drawn in the CIE 1931 chromaticity diagram [16]

luminance levels, protanopes' color discrimination slightly improves, perhaps reflecting input from rod cells [9].

Wavelength discrimination, as a function of wavelength, is considerably worse for protanopes than for normal trichromats. Their lowest discrimination thresholds ($\Delta\lambda$), approaching the trichromatic value, are around the protanopic "neutral point", ca. 490 nm (see below). $\Delta\lambda$ rises on both sides of this minimum, increasing rapidly on the short-wavelength side and on the long-wavelength side until wavelengths are not distinguished beyond approximately 550 nm [6].

Color Perception

A protanope recognizes far fewer distinct colors than a trichromat ("► [Color Naming](#)" by Bonnardel). Instead of all possible equiluminant colors being arranged in a two-dimensional plane ("► [Psychological Color Space and Color Terms](#)" by Bimler), for a protanope they form a single-dimensional line, running between the two extremes of saturated blue and saturated yellow, separated by a "neutral zone" of subjective gray,

including blue-green spectral light, with midpoint at ca. 490 nm [10]. Protanopes distinguish approximately 21 steps along this line, each corresponding to a separate band of interchangeable colors [1]. The usually separate chromatic qualities of "saturation" and "hue" are thus combined along this line. This protanopic linear representation of equiluminant colors is part of a broader reduction of a three-dimensional into a two-dimensional color space.

Whether the subjective qualities of the two extremes are identical to the normal trichromatic "blue" and "yellow" remains controversial and rests on reports from a handful of individuals with unilateral dichromacy. The blue-yellow convention is used in dichromacy simulation software, which converts color scenes to show how a protanope would see them [11].

Despite this model of a two-hue palette for self-luminous colors, the color percepts of protanopes remain in question. Their object-color palette was proved to include all six Hering component colors (red, green, blue, yellow, white, and black) [12]. However, in contrast to normal trichromats, protanopes perceive only a weak green component in object colors and do not experience "unique green," while their red and green components are confounded with an achromatic element ("► [Color categorization and naming in inherited color vision deficiencies](#)" by Bonnardel).

Diagnosis of Protanopia

A number of tests for protanopia are available, varying in sensitivity, specificity, and ease of use [13, 14] ("► [Color Vision Testing](#)" by Paramei, Bimler). Pseudoisochromatic plates and arrangement/panel tests, widely used in surveys/field studies, directly address a protanope's confusions between color pairs along the confusion axis.

In the pseudoisochromatic tests (Ishihara test, Hardy-Rand-Rittler test), each diagnostic plate shows a normal-observer design ("► [Marisa Rodriguez-Carmona Pseudoisochromatic Plates](#)" http://link.springer.com/referenceworkentry/10.1007/978-3-642-27851-8_93-1 digit or curved lines) defined by a color difference which disappears for protanopes, so that the design vanishes

or is supplanted by an alternative design demarcated by a different chromatic distinction.

In the arrangement tests “► [Color Vision Testing](#)” (Farnsworth-Munsell 100 Hue test, Farnsworth D-15 test), the one-dimensionality of a protanope’s color “plane” collapses a color circle into a line (“► [Color Circle](#)” by Green-Armytage), causing drastic departures from a normal observer’s sequence. Errors, or cap transposition “distances,” are summed to quantify the severity of protanopia. Plotted graphically, the transpositions in the D-15 manifest as diametrical circle crossings, indicating the protanopic angle of the confusion axis (“► [Color Perception and Environmentally Based Impairments](#)” by Paramei).

The “gold standard” of diagnosis of protanopia is the Rayleigh match on a small viewing field (2°) of the Nagel anomaloscope (“► [Color Vision Testing](#)”) [14]. Unlike normal trichromats who reproducibly choose a unique ratio of red and green to match the yellow intensity, a protanope can match *any* red/green ratio to the yellow light by adjusting the luminance of the latter, dimming it if the mixed light is mainly red, while a green-dominated mixture requires a more luminant yellow. Notably, some protanopes, when setting Rayleigh matches tested with a large viewing field (10°), perform as anomalous trichromats. Various factors related to retinal eccentricity are presumed, e.g., rod intrusion, S-cone intrusion, spatial variation in the macular pigment, etc. [1].

More recently developed computerized tests for color vision diagnosis, the Cambridge Colour Test and Colour Assessment and Diagnosis test (“► [Color Vision Testing](#)” by Paramei, Bimler), interactively quantify any loss of chromatic sensitivity along the protanopic confusion line. For a protanope, in the frame of the CIE ($u'v'$) 1976 chromaticity diagram (“► [CIE \$u',v'\$ uniform chromaticity scale diagram and CIELUV color space](#)” by Schanda), a MacAdam ellipse of indistinguishable colors displays a major axis lengthened to infinity and orientation pointing to the protanope copunctal point.

Incidences of Protanopia

The prevalence of inherited protanopia varies between human populations. For Caucasian

populations, the reported incidence of protanopia is 1.01 % in males and 0.02 % in females [1]. Recent surveys of red-green deficiency in populations of different racial origin found the incidence of protanopia in Caucasian males to vary between 0.96 % and 1.27 %. In a marked difference, the incidence is about half this rate in indigenous populations in Asia, Africa, Australia, and Central America but is observed to raise in geographic areas that have been settled by incoming migrants, in particular, in coastal areas visited by Europeans in historical times [15]. The increasing prevalence of congenital red-green deficiencies, including protanopia, has tentatively been explained by a relaxation in selection pressures as hunting and gathering cultures evolve toward industrialized societies [1]. More recently, however, it is rather attributed to genetic drift and founder events (variable sampling of the gene pool contingent on the population size) [15].

Cross-References

- [CIE Chromaticity Coordinates \(xyY\)](#)
- [CIE Chromaticity Diagrams, CIE purity, CIE Dominant Wavelength](#)
- [CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams](#)
- [CIE \$u', v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Color Circle](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision, Opponent Theory](#)
- [Color Vision Testing](#)
- [Cone Fundamentals](#)
- [Pseudoisochromatic Plates](#)
- [Psychological Color Space and Color Terms](#)

References

1. Sharpe, L.T., Stockman, A., Jägle, H., Nathans, J.: Opsin genes, cone photopigments, color vision and colorblindness. In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 3–52. Cambridge University Press, Cambridge (1999)

2. Carroll, J., Neitz, M., Hofer, H., Neitz, J., Williams, D. R.: Functional photoreceptor loss revealed with adaptive optics: an alternate cause of color blindness. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8461–8466 (2004)
3. Neitz, M., Neitz, M.: The genetics of normal and defective color vision. *Vision Res.* **51**, 633–651 (2011)
4. Jordan, G., Mollon, J.D.: A study of women heterozygous for colour deficiencies. *Vision Res.* **33**, 1495–1508 (1993)
5. Bimler, D., Kirkland, J.: Colour-space distortion in women who are heterozygous for colour deficiency. *Vision Res.* **49**, 536–543 (2009)
6. Hecht, S., Schlaer, S.: The color vision of dichromats. I. Wavelength discrimination, brightness distribution, and color mixture. *J. Gen. Physiol.* **20**, 57–82 (1936)
7. Pitt, F.H.G.: The nature of normal trichromatic and dichromatic vision. *Proc. R. Soc. Lond. B* **132**, 101–117 (1944)
8. Paramei, G.V., Bimler, D.L., Cavonius, C.R.: Effect of luminance on color perception of protanopes. *Vision Res.* **38**, 3397–3401 (1998)
9. Bimler, D.L., Paramei, G.V.: Bezold-Brücke effect in normal trichromats and protanopes. *J. Opt. Soc. Am. A* **22**, 2120–2136 (2005)
10. Massof, R.W., Bailey, J.E.: Achromatic points in protanopes and deutanopes. *Vision Res.* **16**, 53–58 (1976)
11. Viénot, F., Brettel, H., Ott, L., Ben M'Barek, A., Mollon, J.D.: What do colour-blind people see? *Nature* **376**, 127–128 (1995)
12. Logvinenko, A.: On the colours dichromats see. *Color. Res. Appl.* **39**, 112–124 (2014)
13. Birch, J.: *Diagnosis of Defective Colour Vision*. Oxford University Press, Oxford (1993)
14. Pokorny, J., Smith, V.C., Verriest, G.: Congenital color defects. In: Pokorny, J., Smith, V.C., Verriest, G., Pinckers, A.J.L.G. (eds.) *Congenital and Acquired Color Vision Defects*, pp. 183–241. Grune and Stratton, New York (1979)
15. Birch, J.: Worldwide prevalence of red-green color deficiency. *J. Opt. Soc. Am. A* **29**, 313–320 (2012)
16. Pitt, F.H.G.: *Characteristics of Dichromatic Vision*. Medical Research Council Special Report Series, No. 200. London: His Majesty's Stationery Office (1935)

Pseudoisochromatic Plates

Marisa Rodriguez-Carmona
Optometry and Visual Science, City University
London, London, UK

Synonyms

Color vision test; Isochromatic color confusions;
Screening plates

Definition

Printed pseudoisochromatic plates are the most widely used type of color vision test to screen for color vision deficiency. The principle is that the color of a target (digit or letter) embedded in a background of another color appears “falsely of the same color” to color-deficient people. Target and background chromacities are chosen carefully to be the ones confused by people with color vision deficiency.

Design of Pseudoisochromatic Plates

Pseudoisochromatic plates must be carefully designed in order to be effective. As a result of the printing process, minor misalignments between a figure of one chromaticity and the uniform background of another undistinguishable chromaticity may reveal the shape of the digits or letters. In addition, the relative luminous efficiency of the eye varies considerably within normal trichromats and even more so in color-deficient observers. Hence, a figure and background that are equally light for one color defective are not necessarily so for another. The first pseudoisochromatic plates, introduced in Germany by Stilling in 1877, addressed these issues by breaking up the test target and background into a number of discrete patches each with its own shape and contour, and the luminance of the individual patches was varied randomly rather than being equated. This ensured (as much as possible) that neither edge artifacts nor luminance differences could be used to discriminate the target from the background; thus, the target can be detected only by color discrimination.

Stilling's pseudoisochromatic plates are all of “vanishing” design, so called because the target is read correctly by the normal trichromatic observer and not seen by the color defective (i.e., the “vanishing” figure is pseudoisochromatic to the color defective). A later variant was the “transformation” plate, where two figures are embedded in the background, one that is read by the normal observer and another that has the appropriate chromatic and lightness contrast to be read by

the color-deficient observer. The Japanese ophthalmologist Ishihara published in 1917 the first edition of a set of plates that included vanishing and transformation plates, as well as hidden plates, in which the color defective can see a digit that is camouflaged for the normal by random color variation. Another design is the diagnostic or classification plate which is a more sophisticated version of the vanishing design plate that allows for differentiating between the type of red-green color deficiency, i.e., protan or deutan. The Ishihara plates are not designed for examining yellow-blue deficiencies, i.e., tritan deficiency. In 1954, the AO-HRR (American Optical – Hardy, Rand, and Rittler) plates were printed and comprise both screening and diagnostic plates for tritan defects, as well as diagnostic and grading plates for protan and deutan defects. The plates have vanishing designs containing geometric shapes (circle, cross, and triangle) that are printed in neutral colors on a background matrix of gray dots. The saturation of the neutral colors increases in successive plates to produce designs with progressively larger color difference steps identifying different levels of deficiency.

The Ishihara and the AO-HRR pseudoisochromatic plates are often used together because their functions are complementary. The Ishihara plates are used for screening for red-green color deficiency and the AO-HRR plates to confirm protan/deutan classification, estimate the severity of red-green deficiency, and identify tritans.

Some pseudoisochromatic tests have designs containing shapes or “pathways” for the examination of nonverbal subjects and young children.

Administration of Pseudoisochromatic Plates

The majority of pseudoisochromatic tests have been designed to be viewed at 60–70 cms and standardized for natural daylight illumination (or for CIE standard illuminant C). The examiner turns the pages and controls the viewing time. An introductory plate is often included to demonstrate the visual task.

Examples of Pseudoisochromatic Plates

A number of other pseudoisochromatic tests have been produced, but none have been used as widely as the Ishihara and the AO-HRR plates. Tests have been published in Japan, e.g., the Ohkuma plates (1973), Tokyo Medical College test (1957), and Standard Pseudoisochromatic Plates (1st and 2nd editions) and in the USA, e.g., the Dvorine plates; in Germany, e.g., the Velhagen-Broschmann plates (29th edition published in 1992); in Sweden, e.g., the Bostrom-Kugelberg plates (1972); and in France, e.g., the Lanthony Tritan Album, among many others. Despite many contenders, the pseudoisochromatic plates of Ishihara remain today the dominant instrument for routine screening of color vision.

Cross-References

- ▶ [CIE Standard Illuminants and Sources](#)
- ▶ [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- ▶ [Spectral Luminous Efficiency](#)

Psychological Color Effects

Nilgün Olguntürk

Department of Interior Architecture and Environmental Design, Faculty of Art, Design and Architecture, Bilkent University, Ankara, Turkey

Synonyms

[Emotional color effects](#); [Perceptual color effects](#)

Definition

Psychological color effects are the outcome of color sensations as determined by human perception and emotion, including color influences on feelings, thinking, and behavior.

Introduction and Classification

Colors may alter people's percepts and may affect how they feel. Colors with their psychological effects are most studied under three main areas: advancing and receding/retreating colors, in relation to temperature (warm vs. cool) and in relation to emotions. Advancing and receding/retreating colors are related with visual perception where people tend to see some colors as if they are closer than they actually are or vice versa. The effect of color on warmth perception is related with feeling warm, which could be a physical or sensorial effect. In this case, some colors may increase or decrease the feeling of warmth, although the actual physical temperature remains the same. In relation to emotion, people's emotions may alter when they are exposed to different colors. There may be even more ways that colors may effect percepts and emotions, but the most studied three topics will be expanded in this overview.

Advancing and Receding/Retreating Colors

Colors in Isolation

In studies treating colors in isolation, Luckiesh provided one of the earliest arguments on the "retiring" and "advancing" effects of color letters placed in the same plane [1]. In 1918, he used an apparatus with red and blue filters "of fairly high purity" with which he altered the color of the letters *X* and *E* viewed inside wooden boxes. The subjects moved the red *X* until it appeared to lie in the same plane as the blue *E*. He found a lot of inter-subject variability, but still in most of the cases it was necessary to move the red *X* further away in order to make it appear to be in the same plane with the blue *E* [1, also cited in 2]. One explanation on the advancing quality of red comes from the operational mechanism of the eye. The lens of the eye has to adjust to focus the red light wavelengths, as their natural focal point lies behind the retina. Thus, red advances, creating the illusion that red objects are closer than they actually are [3]. This phenomenon is about the monocular depth perception of the eye where

shortwave light refracts in the eye's optical media more than longwave light; thus, the equidistant sources of different colors cannot be simultaneously focused on the retina, which is also called chromatic aberration [4].

Successive studies on the advancing property of color found instances when blue was judged to be "nearer" than red. Pillsbury and Schaefer, in 1937, had subjects view either red neon or blue neon and argon lights [5]. When the lights were placed equidistantly, the blue light was judged the nearer. The explanation for this "conflict" in apparent nearness of red vs. blue is explained by subsequent research that suggests lightness (or brightness) to be the controlling quantity for apparent distance [2]. The Purkinje shift, a well-known physiological phenomenon, demonstrates that blue light appears lighter than red at low luminance levels.

Colors in Combination

In studies with color combinations, the relationship between perceived colors, e.g., contrast, becomes important. Mount et al. conducted the first outdoor research of color distance in 1956 [6]. They mounted four chromatic (yellow, green, red, blue) and four gray papers on plywood. The plywood with chromatic colors and grays attached on it was viewed outdoors against a dirt bank under full sun. Participants compared a chromatic color or a gray against a dark or light gray standard on each trial. It was found that each chromatic color was judged to be closer than its nearest matching gray. Each of the hues and the grays appeared closer when viewed against a dark standard rather than the light one. They found no difference in advancement for one hue over another. On the other hand, as saturation of a color was increased with respect to its background, its apparent position advanced (an effect equal at most to 1.5 % of the standard distance). Colors having high lightness contrast with their background appeared advancing (an effect equal at most 3 % of the standard distance). Thus, increasing relative contrast by increasing an object's lightness and/or saturation as compared to its background makes the object appear closer [2].

These findings are supported by the 1960 study of Oyama and Nanri [7]. They had participants compare standard and variable circular shapes in all combinations of achromatic and chromatic relations on varying backgrounds under laboratory conditions. They found that the apparent size of the figure increased as its lightness increased, while the lightness of the background decreased. They found no effect of hue on apparent size [2, 7].

Egusa confirmed the findings of the studies above [8]. In 1983 he found an effect of hue when hemifields of different hues were compared for perceived depth. The green-blue difference in perceived depth was smaller than the red-green difference, with red appearing nearer. He also noted that a higher saturated color was judged nearer when it was red or green, but there was no such effect for the blue [8].

In the studies above, it appears that there may be a combined effect of all attributes of color, hue, saturation, and lightness, on advancing or receding/retreating colors, especially when they are viewed in combination. Lightness seems to be the most dominant attribute. Colors having high lightness contrast with their backgrounds appear advancing. Saturation seems to be the second most important attribute of color in judgments of nearness. As saturation of a color increases with respect to its background, it seems closer in the visual field [2].

In Relation to Temperature (Warm/Cool)

The definition of warm is directly associated with color as “having the color or tone of something that imparts heat, specifically of a hue in the range yellow through orange to red” [9]. Warmth perception is a multisensory experience of the immediate environment including physical (e.g., thermal properties, surface properties, etc.) and sensory (e.g., visual, tactile, olfactory, etc.) aspects.

Color in Relation to Physical Aspects (Perceived Temperature) of Warmth

Hue affects warmth perception independent from saturation and lightness [10, 11]. Although no relationship was found between colored illumination and perceived temperature [12], the following studies establish a relationship between applied hue (as in wall finishing colors) and perceived temperature. Itten’s study reports that in a blue-green room,

participants felt cold at 15 °C, whereas they felt cold at 11.1 °C in a red-orange room [13]. In 1976 Porter and Mikellides found that occupants preferred 2.2° higher indoor environment temperature in a blue room than in a red room. Occupants felt cold at 22.2 °C in a blue room, while this temperature was preferable in a red room. In a red room, they felt warm at 24 °C [14].

Color in Relation to Sensory Aspects of Warmth (Warm Colors vs. Cool Colors)

Warm colors are perceived warmer than cool colors [11, 15]. In interiors, one study reports that red and yellow induce the perception of warmth, while white is regarded as the coolest color [16]. In applied surface colors, hue influences warmth perception, where warm colors (red, orange, yellow) are regarded as being warmer than cool colors (green, blue, purple).

In a study on “nasal thermal sensation,” participants felt cold in their nasal nostril when smelling green-colored water and felt warm in their nasal nostril while smelling red-colored water. Thus, the visual sense affects the warmth perception of the olfactory sense as well [17].

Color Emotion

Color as an effective design tool influences people’s emotions in interior spaces. Red is one of the powerful colors and creates the highest number of emotional responses [18]. It is associated with surprise, happiness and sadness [19], energy, vitality and power [20], fear, and anger [21]. Thus, red evokes emotions in a range from the negative (anger, fear, and sadness) to positive (happiness, energy, power) ones [22].

Green is associated with nature and trees, which creates a feeling of comfort with a number of positive emotions such as feelings of relaxation, happiness, comfort, peace, and hope [23].

Blue is connoted with infinity and serenity in relation to the sky and the sea, representing peacefulness and relaxation [24]. Blue was found also to be associated with happiness [19], calmness, peacefulness, relaxation, modernism, coldness and dullness [25], and sadness [26].

Users usually desire colors in order to feel something or in order to create a psychological

bond with an object or a space. It should also be noted that if people indicate any color as being their least favorite in their daily lives, in some instances they also associate an interior with that specific color with disgust [22].

Psychosomatic Effects of Color

Each color and color combination has its own sensation. Individually or in relation to each other, colors may become eye irritants or may cause headaches. They can effect brain waves, hearth rates, blood pressure, and respiratory rates [27]. There is an ancient and widespread faith in the healing power of color, some of which is studied under color healing or color therapy [28, 29].

Cross-References

- [Anchoring Theory of Lightness](#)
- [Color Combination](#)
- [Color Contrast](#)
- [Mixed Chromatic Adaptation](#)
- [Unique Hues](#)

References

1. Luckiesh, M.: On "retiring" and "advancing" colors. *Am. J. Psychol.* **29**, 182–186 (1918)
2. Camgöz, N.: Effects of hue, saturation, and brightness on attention and preference. Dissertation, Bilkent University, Ankara (2000) (Also available by UMI, Bell & Howell Co., Ann Arbor, 2001).
3. Mahnke, F.H., Mahnke, R.H.: *Color and Light in Man-Made Environments*. Van Nostrand Reinhold, New York (1987)
4. Sundet, J.M.: Effects of color on perceived depth: review of experiments and evaluation of theories. *Scand. J. Psychol.* **19**, 133–143 (1978)
5. Pillsbury, W.B., Schaefer, B.R.: A note on advancing-retreating colors. *Am. J. Psychol.* **49**, 126–130 (1937)
6. Mount, G.E., Case, H.W., Sanderson, J.W., Brenner, R.: Distance judgment of colored objects. *J. Gen. Psychol.* **55**, 207–214 (1956)
7. Oyama, T., Nanri, R.: The effects of hue and brightness on the size perception. *Jpn. Psychol. Res.* **2**(1), 13–20 (1960)
8. Egusa, H.: Effects of brightness, hue and saturation on perceived depth between adjacent regions in the visual field. *Perception* **12**(2), 167–175 (1983)
9. Merriam-Webster online dictionary. <http://www.merriam-webster.com/dictionary/warm>
10. Wright, B.: The influence of hue, lightness and saturation on apparent warmth and weight. *Am. J. Psychol.* **75**(2), 232–241 (1962)
11. Wastiels, L., Schifferstein, H.N.J., Heylighen, A., Wouters, I.: Red or rough, what makes materials warmer? *Mater. Design.* **42**, 441–449 (2012)
12. Berry, P.C.: Effects of colored illumination upon perceived temperature. *J. Appl. Psychol.* **45**, 248–250 (1961)
13. Itten, J.: *The Elements of Color*. In: Birren, F. (ed., trans: Van Hagen, E.). New York: Van Nostrand Reinhold (1970)
14. Porter, T., Mikellides, B.: *Colour for Architecture*. Studio Vista, London (1976)
15. Newhall, S.M.: Warmth and coolness of colors. *Psychol. Record.* **4**, 198–212 (1941)
16. Wastiels, L., Schifferstein, H.N.J., Heylighen, A., Wouters, I.: Relating material experience to technical parameters. *Build. Environ.* **49**, 359–367 (2012)
17. Michael, G.A., Rolhion, P.: Cool colors: Color-induced nasal thermal sensations. *Neurosci. Lett.* **436**, 141–144 (2008)
18. Boyatzis, C.J., Varghese, R.: Children's emotional associations with colors. *J. Genetic Psychol.* **155**(1), 77–85 (1994)
19. Terwogt, M.M., Hoeksma, J.B.: Colors and emotions. *J. General Psychol.* **122**(1), 5–17 (1995)
20. Kaya, N., Crosby, M.: Color associations with different building types: An experimental study on American college students. *Color Res. Appl.* **31**(1), 67–71 (2006)
21. da Pos, O., Green-Armytage, P.: Facial expressions, colors and basic emotions. *Color Design Creat.* **1**(1), 1–20 (2007)
22. Helvacioğlu, E.: *Color-emotion associations in interior spaces*. Dissertation, Bilkent University, Ankara (2011)
23. Kaya, N., Epps, H.N.: Relationship between color and emotion: A study of college students. *Coll. Stud. J.* **38**(3), 396–405 (2004)
24. Fehrman, K.R., Fehrman, C.: *Color: The Secret Influence*. Prentice-Hall, New Jersey (2000)
25. Manav, B.: Color-emotion associations and color preferences: A case study for residences. *Color Res. Appl.* **32**(2), 144–150 (2007)
26. Zentner, M.R.: Preferences for colors and color-emotion combinations in early childhood. *Dev. Sci.* **4**(4), 389–398 (2001)
27. Kaiser, P.K.: Physiological response to color: A critical review. *Color Res. Appl.* **9**(1), 29–36 (1984)
28. Helen, V.: *Colour*. Marshall Editions, London (1983)
29. O'Connor, Z.: Color psychology and color therapy: Caveat emptor. *Color Res. Appl.* **36**, 229–234f (2011)

Psychological Color Space and Color Terms

David L. Bimler

School of Psychology, Massey University,
Palmerston North, New Zealand

Synonyms

Color appearance solid; Color similarity space;
Relational color space

Definition

Psychological color space is the relational structure among color stimuli that can be found using empirical tasks that assess color similarities. Color terms are the lexical categories (which can vary across different ethnolinguistic groups) that are used to label, or describe, color appearances organized as meaningful partitions of the psychological color space.

Psychological Color Space

A color-ordering system requires three independent modes of variation to accommodate the full gamut of color experience and arrange all possible hues within a coherent continuum, in which similar pairs of colors are adjacent neighbors. This requirement, so familiar now, only emerged after a long history of trials with one- or two-dimensional schemes [1]. Here colors are considered in isolation as opposed to in the context of other colors where induced simultaneous-contrast effects complicate color appearance.

“Psychological color space” is synonymous with a “color solid” containing all physically attainable hues. Numerous psychological qualities or attributes have been nominated as its underlying dimensions ([1], Chap. 5). The long delay before the acceptance of the three-dimensionality of color space suggests that most of these attributes are not immediately evident and require a

trained observer. Perhaps the most intuitive of these qualities, the circular sequence of hues around a color wheel, only became an organizing principle after Newton’s description of the prismatic spectrum. Later theorists refined the Newtonian color wheel by inserting nonspectral purples between the red and violet spectral extremes.

One tradition portrays the color solid as a sphere or as a pair of cones or pyramids joined at their bases. The “equator” is the color wheel, containing the purest, most saturated form of each hue. A *hue angle* locates each hue within the rainbow sequence. Perpendicular to the wheel and through its center runs an achromatic axis of gray tones, connecting the polar extremes of white and black. Derived forms of a given hue share the same angular coordinate, occupying space between that point on the wheel and the axis. In effect they comprise one page in a “book of color” radiating out from a central spine. In the Natural Color System – one example of a color space – these intermediate hues are derived from the pure hue by dilution with various amounts of black or white or both.

Clearly this is a cylindrical coordinate system. As well as the angular coordinate, it implies a radial coordinate (the distance out from the axis), variously labeled as “saturation” or “colorfulness” or “purity.” The third coordinate, running parallel to the axis, is some version of “lightness.”

Hering [2] argued that in addition to lightness (defined by its extremes black and white), “red” and “green” are the extremes of a second opposition. Each cancels out the other so the two cannot coexist. “Blue” and “yellow” are a third opposition. These three oppositions form an alternative way of spanning color space, with rectilinear axes which specify a hue by its content of “redness vs. greenness” and “blueness vs. yellowness.” In terms of radial coordinates, the “cardinal hues” of yellow, red, blue, and green might be located at 0°, 90°, 180°, and 270°, respectively.

In practice the purest yellow hues are relatively light, and conversely the purest purples are relatively dark. Thus, all saturated hues can only be located on the equator of the color solid by treating the lightness scale differently on each of

the radial color “pages.” The alternative is to treat “lightness” as an *objective* property that can be measured with a photometer and slice color space into planes of equal lightness, with the most saturated example of each hue located empirically in the appropriate plane (in some depictions of color space, lightness is replaced with a “brightness” axis, a more subjective quality which also increases with saturation). The effect is to tilt the color wheel to form an oblique angle with the achromatic axis, distorting the geometry of double-cone and double-pyramid models of color space.

In addition, “maximum saturation” is not a constant quality around the wheel. The most saturated “reds,” for instance, are perceptually more intense than the most saturated blues, i.e., more dissimilar from the corresponding gray tone. Incorporating these aspects of color perception produces a less regular color solid. In the widely used “Munsell color space” [3], the radial coordinate (chroma) applies the same metric across hues, reaching its highest values for the intense red hues of intermediate lightness so that they bulge out on the side of the solid. In explanations of how color names are assigned to color space, the geometrical salience of this region marks the red hues as particularly “namable,” along with “black” and “white” [4].

Munsell space emphasizes the principle of *uniformity*, that the distance between two hues in the space should reflect the perceptual dissimilarity between them. That is, a given distance at any region of color space, such as a unit increment in saturation level, should correspond to the same inter-hue dissimilarity. This empirical approach also led Munsell to expand the blue/red quadrant of the color wheel and insert “purple” as a fifth cardinal hue, contrary to the Hering polarities.

In another application of the uniformity principle, hue samples were chosen to form a square grid spanning each equal-lightness plane [5]. That is, dissimilarities were intended to be the same between each pair of adjacent samples, as far as departures from Euclidean geometry would permit. The hue samples are also specified by their coordinates on three rectilinear axes *L* (lightness), *g*, and *y*, but the *g* and *y* scales were determined

afterwards rather than fundamental to this uniform color scale scheme (the OSA-UCS). Again, the resulting color solid is irregular.

An extension of the uniformity principle takes advantage of multidimensional scaling methods (MDS) to explore color ordering. Here ratings of dissimilarity among pairs of hue samples are elicited from observers and analyzed with MDS algorithms, yielding geometrical solutions where points representing the samples are arranged so that distances among them reflect the dissimilarities. In a series of empirical studies, the samples were selected from the Munsell system to vary across all three coordinates [6]. The results generally support that system’s uniformity.

Numerous other models of color space exist, emphasizing different aspects of color experience. There is a tension between the uniformity principle and other features that might be desirable, in particular the “linearity” principle that mixtures of two colored lights should be represented by points along a straight line. CIE tristimulus space is an example of a physiological model where distances are strongly nonuniform.

Why Three Dimensions?

In the final analysis, three dimensions are necessary – though they can be conceptualized and spanned in different ways – because human color vision rests on the existence of three distinct populations of photoreceptor cells in the retina (*S*, *M*, and *L* cones), reducing the spectral information within a given stimulus to three levels of cone outputs. In the first stage of visual processing, neurons in the retina create combinations of the raw cone outputs that are more informative than single values in isolation: a sum (*L*+*M* or “luminance”) and two cone-output differences (*L*-*M* and *S*-(*L*+*M*)). These combinations provide the three axes of *cone excitation space*, widely used in vision research to define experimental stimuli and results.

However, physical light can vary along far more than 3 degrees of freedom. Thus, two stimuli with physically different light spectra may appear perceptually identical (“metamers”) and be shown as the same point in color space because they provide the same stimulation to each of the cone

classes. That is, trichromatic vision captures only a fraction of the information contained within the spectra of stimuli in the natural visual world. Factor analysis of spectra from a range of objects under a range of lighting conditions indicates that to capture most of the spectral variation would require six or seven distinct photopigments and additional color dimensions [7].

Notably, higher-dimensional spaces are observed in species of reptiles and birds. The three-dimensionality of human color vision suggests that additional photopigments bestowed no advantage over the course of human evolution, i.e., the differences among metamers were not significant in the ecological niches occupied by prehuman primates.

The majority of mammals make do with dichromatic color vision, having only two cone classes. This in turn implies that trichromatic vision *was* adaptive for some primates, with evolution selecting for the divergence of L- and M-photopigment genes. Several adaptive advantages have been advanced to account for this feature of human color space. The difference between M and L cones allows normal trichromats to detect red objects against green backgrounds. In particular, ripe fruit stands out against foliage, as do older, more digestible leaves [8]. Further speculative benefits from the red-green dimension of color space include social signaling, a sexual selection role, and preserving color constancy under changes in lighting conditions [7].

Abnormal Color Space

Color space is convenient for representing the effects of color vision deficiencies (CVDs). Depending on the specific form of deficiency, two colors may be harder to discriminate (i.e., perceptually more similar) when they are separated along a particular direction in the space. This characteristic direction – running between red and green, for the common forms of CVD – is the *confusion axis*. Thresholds for distinguishing between a reference hue and small variations of it are increased in this direction (e.g., slightly redder and slightly greener versions). The locus of hues that are just noticeably different

from the reference, which would form a circle when plotted in a suitable color plane, instead becomes an ellipse elongated along the confusion axis.

Alternatively, one could say that the *personal* color space for a CVD observer is compressed, reducing distances along the confusion axis. The degree of compression corresponds to the severity of the deficit. This interpretation underlies the design of widely used panel tests for CVDs. In the extreme case where an entire class of photoreceptor is absent, color space becomes two- rather than three-dimensional, and the color circle collapses to a line. The principle is supported by numerous studies where MDS reconstructed personal color spaces for individuals with CVDs.

Color Terms

As seen above, color space provides a framework for analyzing color language. For a given *Basic Color Term* (BCT), the range of hues to which that term can be ascribed (its extension) comprises the corresponding *Basic Color Category* (BCC). The BCC can be treated as a geometrical domain by mapping that range into color space.

Each BCC is circumscribed by its boundaries with adjacent color categories. “Red,” for instance, has boundaries with the domains of “pink,” “orange,” “purple,” and “black.” On one side of the “watershed,” a stimulus might be labeled more often than not as “red,” whereas a second stimulus, nearby in color space, might attract “pink” as the majority name. Thus, the continuity of color is *partitioned* into color-category domains. There are four points to note here:

1. The vicinity of a category boundary creates *categorical perception* effects in perceptual tasks. These effects are vitiated by verbal interference, indicating that language plays some role in them.
2. For English, the partition is *complete* in the sense that the category domains cover color space in its entirety; there is an applicable term for any realizable color. This is not necessarily the case in every language.

3. Empirically, the domains of English color categories are *connected* and *convex*. That is, there are no situations where two hues receive the same color term but have some hue between them in color space which is labeled with a different term. In particular, there is no term with a domain separated into two non-contiguous parts.
4. The boundaries of a BCC are part of the BCT's definition, between them circumscribing the hues that it can label. They are not *all* of its definition, however. A color term is also characterized by its *focal hues*, i.e., the hues picked out from the color gamut as prototypal best examples. One may also define the BCC "centroid" hue, the "center of mass" or mean location of its domain in color space.

So far the requirements for a color term to be "basic" have not been spelled out. One criterion have been assumed implicitly: a BCT must be understood by almost all speakers of a language, with consensus about its meaning, i.e., few disagreements or misunderstandings resulting from divergent usage. A second criterion is that to be a color term at all, a word must generalize across semantic domains and be applicable to any arbitrary object, as opposed to words like "roan," "palomino," and "dun," specific to animal coats.

In English, 11 color words are BCTs, 3 achromatic terms, and 8 chromatic terms: black, white, grey, red, pink, orange, yellow, brown, green, blue, and purple.

English also provides a rich vocabulary of *subordinate* color terms. Some of these are compounds, generated from a BCT and made more specific by modifying it with a qualifier, e.g., "dark blue" or "bluish green," or with a noun as exemplar, e.g., "sky blue." Others are derived directly from specific objects or pigments (e.g., "lilac" or "magenta"). That is, to use them is to use an implicit metaphor ("lilac colored"). A third criterion for BCTs is that they must not be compounds or still linked to a noun and carrying its connotations. This third criterion requires some care. The BCT "orange" arose from the name of the fruit, but the metaphor has become so routine

that it is now a separate word with no citric associations in the minds of speakers.

However, the distinction between basic and subordinate terms is not clear-cut. Rather, "basicness" is a continuous gradient along which can be located by examining their usage, generative properties, etc. [9].

Color terms tend to form a nested, hierarchical pattern, with a broader basic term (*hypernym*) divisible into qualified, more specific *hyponyms*. There are exceptions, color being a continuum rather than a structure of nested boxes, and a number of subordinate terms straddle the borders between BCTs. "Turquoise" and "teal" lie between "blue" and "green" and are more specific than either, though speakers may argue about their exact usage.

Primary and Secondary Terms

Among the BCTs, a distinction is sometimes drawn between *primary* and *secondary* terms. The former are the six "cardinal hues" comprising the three oppositions of the Hering opponent-hue scheme. Primary terms lend themselves to "unique hue" judgments (see entry ► [Unique Hues](#)). "Unique green," for instance, is a wavelength that is seen as "green" on its own with no admixture of "blue" or "yellow." Secondary terms are derived from overlaps or intersections of primaries. Three of them – orange, purple, and pink – are the intersections between "red" and its primary neighbors "yellow," "blue," and "white," respectively. "Brown" is the overlap between "yellow" and "black," while "grey" occupies a middle ground between the achromatic extremes of "white" and "black." However, the putative primary/secondary distinction does not appear on the basicness scale as an obvious jump. In addition, it leaves unexplained why some of the terms for primary-term intersections are basic while others (e.g., turquoise) are subordinate.

Primary terms are featured in the *color-naming task*. Here observers characterize a series of briefly presented stimuli using a constrained list of options. A 590-nm light, for instance, might be described as "red with some yellow" (or "70 % red, 30 % yellow"). Repeated exposures build up the precision of the description. The procedure

allows subtle nonlinearities in color vision to be quantified, such as the Bezold-Brücke effect (the changes in chromatic content of a colored light as it becomes more intense). In this context, observers are comfortable with describing hues as combinations of “red” and “yellow” when a secondary term such as “orange” is not an available response. This contrasts with the situation when a primary term such as “blue” is absent from the options; observers prefer to leave that component of the stimulus appearance unlabeled (“none of the above”).

Color vocabulary becomes more finely divided in the course of language acquisition from childhood to maturity. One might expect primary terms to be learned earlier than secondaries, but the evidence for a standard order of acquisition is equivocal [10]. Cultural factors are involved, with a historical drift toward acquiring terms at an earlier age.

Ontological Status

The ontological status of the BCTs and the categories of experience they represent are still unclear. Are the qualities of “redness” and “blueness” part of the basic genetic patrimony of every human, hardwired into human brains (as the *universalist* position has it)? To what extent is the genetic endowment modified by formative visual environment and cultural factors? (specifically, the color lexicon of one’s first language).

A number of MDS studies have elicited judgments of dissimilarities among pairs of color *terms* rather than pairs of hues. The resulting solutions, in which the BCTs are embedded in a *conceptual* color space, show no substantial differences between languages (e.g., English and Mandarin) or from perceptual MDS models [11]. Notably, color-deficient informants internalize the normal relationships within color space such as the diametrical opposition between “red” and “green” [12]. That is, they are intellectually aware of color relationships that they have not detected from observation. MDS has also been applied to *empirical* relationships among BCTs, using the number of samples on the boundary between two BCCs as the index of their similarity.

The issue has inspired comparisons with non-English and non-European languages. It early became evident that many cultures do not emphasize hue as a topic of discussion to the same extent as English. Many languages do not distinguish “color” from other qualities of surface appearance such as lightness, sheen, or texture: depending on context, words may refer to color or some other visual effect. A notable example is the difficulty of translating putative color adjectives in the *Iliad* from Homeric Greek into consistent English equivalents. In other languages, terms may have chromatic applications but are enmeshed in a web of cultural connotations.

Systematic cross-cultural research requires a standardized palette of hue samples. Informants are presented with each sample in term and asked to identify its color. One such palette, the Munsell array, has been featured in the *World Color Survey* and the *Mesoamerican Color Survey* and in numerous studies of individual cultures (see entry ► [World Color Survey](#)).

The methodology is open to critiques, by presupposing for each culture an interest in color as a concept and the presence in each language of color terms per se – neither domain specific nor carrying primarily non-chromatic meanings [13]. That is, it presupposes what it purports to discover, imposing a modern English/European conceptual framework and only capturing those aspects of color discourse that fit. Thus, results from this research tradition may underestimate the prevalence of “non-partition” languages where some colors go unlabeled.

Bearing this caveat in mind, the WCS found the number of BCTs to vary widely across languages, making finer or coarser chromatic distinctions. Thus, many languages subsume “blue” and “green” within a broad “grue” term (some “grue languages” possess a word equivalent to “purple,” in further evidence against the primary/secondary distinction). There is, however, a cross-cultural consensus about the focal hues and “centroid hues” of the BCTs. A language may have fewer BCTs than English, but the prototypical examples of their BCCs are generally close to an English equivalent [14]. Moreover, these linguistic distinctions are *prioritized*, in the sense that some

are never made unless other “earlier” distinctions have been made already (for instance, if a language marks only three color categories, these will be “dark colors,” “light colors,” and “red”). These sequence relationships define a “Berlin-Kay sequence” (see entry ► [Berlin and Kay Theory](#)).

There is clearly a special status for BCTs. Several explanations for their cross-cultural nature have been advanced [4, 15]. At the same time, CP effects on color-similarity judgments vary between languages, depending on the boundaries between BCCs. These boundaries necessarily shift with the size of the color lexicon but can differ even between languages with the same number of color terms (see entry ► [Comparative Color Categories](#)).

Cross-References

- [Berlin and Kay Theory](#)
- [Comparative Color Categories](#)
- [Unique Hues](#)
- [World Color Survey](#)

References

1. Kuehni, R.G., Schwarz, A.: *Color Ordered: A Survey of Color Systems from Antiquity to the Present*. Oxford University Press, New York (2008)
2. Hering, E.: *Outlines of a Theory of the Light Sense*. Harvard University Press, Cambridge (1957). Translated by L. M. Hurvich & D. Jameson
3. Munsell, A.H.: *Atlas of the Munsell Color System*, 2nd edn. Wadsworth-Howland, Malden (1915)
4. Regier, T., Kay, P., Khetarpal, N.: Color naming and the shape of color space. *Language* **85**, 884–892 (2009)
5. MacAdam, D.L.: Uniform color scales. *J. Opt. Soc. Am.* **64**, 1691–1702 (1974)
6. Indow, T.: Multidimensional studies of Munsell color solid. *Psych. Rev.* **95**, 456–170 (1988)
7. Maloney, L.T.: Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A*, **3**, 1673–1683 (1986)
8. Sumner, P., Mollon, J.D.: Did primate trichromacy evolve for frugivory or folivory? In: Mollon, J.D., Pokorny, J., Knoblauch, K. (eds.) *Normal and Defective Colour Vision*, pp. 21–30. Oxford University Press, Oxford (2003)
9. Kerttula, S.: Relative basicness of color terms: modeling and measurement. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color*, pp. 151–169. John Benjamins, Philadelphia (2007)
10. Pitchford, N.J., Mullen, K.T.: The developmental acquisition of basic colour terms. In: Pitchford, N.J., Biggam, C.P. (eds.) *Progress in Colour Studies: Volume II. Psychological Aspects*, pp. 139–158. John Benjamins, Philadelphia (2006)
11. Moore, C.C., Romney, A.K., Hsia, T.-L.: Cultural, gender, and individual differences in perceptual and semantic structures of basic colors in Chinese and English. *J. Cogn. Culture* **2**, 1–28 (2002)
12. Shepard, R.N., Cooper, L.A.: Representation of colors in the blind, color-blind, and normally sighted. *Psych. Sci.* **3**, 97–104 (1992)
13. Saunders, B.: Revisiting basic color terms. *J. R. Anthropol. Inst.* **6**, 81–99 (2000)
14. Regier, T., Kay, P., Cook, R.S.: Universal foci and varying boundaries in linguistic color categories. In: Bara, B.G., Barsalou, L., Bucciarelli, M. (eds.) *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* pp. 1827–1835. Lawrence Erlbaum Associates, Mahwah, NJ (2005)
15. Bimler, D.: Are color categories innate or internalized? Hypotheses and implications. *J. Cogn. Culture* **5**, 265–292 (2005)

Psychological Impact of color

- [Color Psychology](#)

Psychological Primary Colors

- [Psychologically Pure Colors](#)

Psychologically Pure Colors

Wayne Wright
Department of Philosophy, California State
University, Long Beach, Long Beach, CA, USA

Synonyms

[Psychological primary colors](#)

Definition

Psychologically pure colors are colors that exhibit a homogeneous appearance on some dimension of perceived color. Purity of appearance is closely tied up with theoretical understandings of the three canonical color attributes of hue, saturation, and lightness. According to the dominant Hering opponent color framework, the hues red, green, blue, and yellow are distinctive and fundamental to all hue perception, in part due to the fact that they admit of pure or unique variants and all other hues (such as orange) appear as mixtures of them. Saturation and related notions are characterized essentially in terms of purity, involving judgments regarding differences in the appearance of chromatic stimuli and achromatic stimuli. Highly saturated colors appear to have low achromatic content or to be more intensely chromatic. In the achromatic domain, the lightness dimension runs from light to dark. The corresponding extreme ends of that dimension, white and black, have an appearance that wholly excludes the other, while intermediate steps between them are grays that appear to be mixtures of black and white. The concept of purity plausibly could also be usefully applied in connection with other ways of categorizing perceived color, e.g., warm versus cool and positive feeling versus negative feeling.

Overview

The concept of subjective purity figures in discussions of the three attributes of perceived color around which the bulk of contemporary theorizing is organized: hue, lightness (value), and saturation (chroma). For hue, the apparent purity of certain hues has played a central role in the development of the dominant perspective on the structural relations among the hues: Ewald Hering's opponent colors theory [1]. In Hering's theory, hues are denominated by terms such as "red," "blue," "green," and "yellow"; note that the terms used to refer to hues from the standpoint of theorizing about color experience need not reflect the color terms or concepts employed by particular perceivers or groups of perceivers. According to

Hering's conception of the hues, the four hues just named have a special status as *primary* hues. They are supposed to be psychologically basic elements out of which all hue sensations are constructed. The primary hues are hypothesized to be arranged along two independent axes, red/green and blue/yellow. This opponent structure is thought to explain phenomenological observations about the purity of some hues and the mixed appearance of others, as it allows for *unique* percepts of each primary hue and *binary* combinations of primary hues not on the same axis. Purity is the essential criterion for being a unique hue, as unique hue percepts are those that consist entirely of a single primary hue. For example, while most greens are either bluish or yellowish, unique green is the shade of green that is perceived as having no tinge of either blueness or yellowness to it. By contrast, orange is a binary hue, always perceived as some blend of redness and yellowness, while turquoise is seen as a combination of greenness and blueness. It is important to emphasize that the purity of the unique hues and the impure nature of binary hues are thought to be phenomenologically present in experience.

Regarding the unique hues, their apparent purity goes hand in hand with their perceptual salience and the privileged status of the Hering primaries as elemental hue sensations. As for binary hues, they are said to look to contain proportions of their component primaries and to not be perceived as unanalyzable, integral wholes formed from the primaries. The experimental literature on hue scaling fits with the preceding phenomenological observations and Hering's opponent colors theory. Subjects in these studies have no little or no difficulty understanding the task instructions and can produce judgments about hue content for all hue stimuli using only the Hering primaries (in permissible combinations for nonunique hues) and which total to 100 %, with unique instances of the Hering primaries excluding all other hues [2]. To be clear, there is considerable variability across subjects when it comes to which particular stimuli are designated as instances of the unique hues [3]. The binary hues seem completely unsuitable for use in a hue scaling task, and the subjective purity of, say,

unique blue makes it absurd to ask subjects to judge what proportions of purple and turquoise (or red and green, for that matter) look to be in it.

The color appearance of achromatic stimuli (i.e., those with no hue content) varies along the lightness dimension between the extremes of black and white, with a gradient of shades of gray in between. Subjective purity goes along with those achromatic extrema in much the same way it does with the unique hues. The apparent purity of whiteness has long been of cultural significance, and distinctions are commonly made between pure and impure whites. The concept of absolute or pure black has also aroused much interest, and artists for centuries have consciously sought to avoid pure blacks when rendering natural scenes. That black and white have a perceptual salience akin to that of the unique hues is evident in the fact that samples of them used in the World Color Survey received by far the most selections as “best examples” of basic color terms across all participants from the 110 languages surveyed [4]. Hering took white and black to be psychologically basic (and opposed) color sensations on a par with the primary hues. He also proposed a measure of color purity stated in terms of proportions of whiteness, blackness, and primary hues perceived in a stimulus [5].

Saturation and related notions inherently are purity concepts that involve the perceived difference between a chromatic (i.e., hue-bearing) stimulus and an achromatic stimulus. Highly saturated colors appear quite vivid and more “full” of their component hue than do less saturated colors. As with hue, scaling procedures are available for saturation, with subjects making judgments about the chromatic purity of stimuli in terms of the percentages of their total color sensations that are chromatic and achromatic [2].

While not an essential commitment of views like Hering’s that account for color phenomenology in terms of red/green, blue/yellow, and white/black axes, throughout the last 100 years, it has been common to believe or hope that those axes have a direct neurophysiological realization. Although caution is in order when allowing

phenomenology to influence neurophysiological theorizing, the apparent purity and perceptual salience of the unique hues, white, and black make it natural to suppose that they must have some correspondingly distinctive neurophysiological underpinning. With respect to perceived hue, a familiar idea is that there are separate, independently operating red/green and blue/yellow neural “systems” or “channels,” the combined outputs of which account for the particular unique or binary hue appearance encountered in experience. For example, unique green is seen when the blue/yellow system is nulled and the red/green system points to green, while a purple sensation arises when the red/green and blue/yellow systems are tipped toward red and blue, respectively. It should be pointed out that this theoretical perspective need not endorse a necessary connection between any particular kind of neural activity and any particular kind of color phenomenology across all perceivers. Rather, what matters is that each individual perceiver has her own personal profile of neural responses such that the hypothesized red/green and blue/yellow systems can exhibit the suggested null and non-null activity. In any event, the idea that subjective purity has a direct neurophysiological grounding is complicated by a range of empirical findings regarding *inter alia* the characteristics of color-opponent processing cells in the visual system [2, 6, 7].

Despite purity being so important to the understanding of key attributes of color experience, appeals to it have not gone unquestioned. For example, in the early part of the twentieth century, Carl Stumpf rejected saturation as a basic attribute of perceived color on the grounds that it is, rather than a perceptual characteristic pertaining to the chromatic purity of a sensation, a cognitive abstraction that involves a comparison between a perceived color and some ideal of it [5]. Saturation and its brethren clearly have had a vexed history. However, perhaps the most interesting concerns voiced about appeals to subjective color purity concern the possibility that its application has been overly restricted in treatments of hue. Such worries ultimately touch on the basic motivations

for the deeply entrenched Hering opponent colors theory.

In the philosophical and empirical literatures on color, statements that echo the remarks above about the essentially mixed appearance of binary hues are widespread. However, expressions such as “pure purple” and “pure orange” can be heard in everyday conversations and are invoked in artistic and commercial settings. It is certainly easy to find paints and other products with such labels. Some scientists and philosophers have also disputed claims about the necessarily mixed appearance of all hues other than the unique Hering primaries. They contend that their own phenomenological reflection on hues such as orange and purple reveals that there are also pure variants of them, e.g., a pure purple that does not look to contain any blue or red [8, 9]. Others have objected that empirical results concerning unique hue judgments that seem to support the thesis that the binary hues have a subordinate status to the Hering primaries are an artifact of the instructions given to participants. The alternative suggestion is that a minor modification of those instructions would lead subjects to find hues such as “unique orange” and “unique turquoise” that are just as subjectively salient as the unique Hering primaries [10].

As was observed above, the form of purity relevant to the unique/binary hue distinction is perceptual (psychological), not physical or otherwise objectively assessable. This, together with the heterodoxies just noted, invites the possibility that different perceivers may have either (i) different criteria for what counts as a pure hue percept or (ii) fundamentally different kinds of structure to their color experiences that either allow or forbid the possibility of hues such as pure orange. Both are potential sources of experimental or theoretical confusion, as well as interesting topics of study in their own right. Another idea worth considering, likely in connection with (i), is that the frequency of pronouncements regarding the mixed appearance of all purples (oranges, etc.) is due to the pervasive influence of the Hering opponent colors theory on scientific

and philosophical discussions of color. The attractively tidy picture the Hering theory offers of both psychological and neurophysiological aspects of color vision certainly could affect not only how experiments are constructed and their results are interpreted but also how individuals construe their own phenomenology. It has already been pointed out that empirical difficulties confront attempts to find neurophysiological mechanisms mirroring the hypothesized opponent hue axes of Hering’s model of color phenomenology. Further research open to the possibility that purity in hue perception might not be restricted to red, green, blue, and yellow could present a serious challenge to the phenomenological tenets that make Hering’s theory such an appealing account of color perception.

At a minimum, it appears that there is a need to further sharpen the criteria for what should count as a pure color percept, if matters such as those just noted are to be resolved. Of course, what is to be made of the construct of “psychologically pure colors” depends on how psychological color is understood. That, in turn, may have important consequences for the role played by the notion of purity. For example, certain (lexical, practical, emotional, etc.) categories that are associated with color plausibly could affect or facilitate judgments of purity in various ways. Potentially in play here are distinctions that go beyond the sort of structural relations among the hues that figure in Hering’s theory, such as “dessicated” versus “moist,” “hot” versus “cold,” or “positive feeling” versus “negative feeling” colors; see Schloss and Palmer’s entry on color preference. It may therefore be appropriate to consider that subjective color purity is not univocal but rather is highly dependent on the task used to assess it and thus should be defined in each circumstance of use in terms of the particular parameters of its empirical assessment.

Cross-References

- [Berlin and Kay Theory](#)
- [Color Preference](#)
- [Unique Hues](#)

References

1. Hering, E.: *Outlines of a Theory of the Light Sense*. Harvard University Press, Cambridge, MA (1920/1964)
2. Abramov, L., Gordon, J.: Seeing unique hues. *J. Opt. Soc. Am. A* **22**, 2143–2153 (2005)
3. Kuehni, R.: Variability in unique hue selection: a surprising phenomenon. *Color Res. Appl.* **29**, 158–162 (2004)
4. Cook, R., Kay, P., Regier, T.: The world color survey database: history and use. In: Cohen, H., Lefebvre, C. (eds.) *Handbook of Categorisation in the Cognitive Sciences*, pp. 224–241. Elsevier, Amsterdam (2005)
5. Kuehni, R.: *Color Space and Its Divisions*. Wiley, Hoboken (2003)
6. MacLeod, D.: Into the neural maze. In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, pp. 151–178. MIT Press, Cambridge, MA (2010)
7. Malkoc, G., Kay, P., Webster, M.: Variations in normal color vision IV. Binary hues and hue scaling. *J. Opt. Soc. Am. A* **22**, 2154–2168 (2005)
8. Koenderink, J.: *Color for the Sciences*. MIT Press, Cambridge, MA (2010)
9. Wright, W.: Reply to broackes. *Rev. Philos. Psychol.* **2**, 629–641 (2011)
10. Jameson, K.: Where in the world color survey is the support for the Hering primaries as the basis for color categorization? In: Cohen, J., Matthen, M. (eds.) *Color Ontology and Color Science*, pp. 179–202. MIT Press, Cambridge, MA (2010)

Public Lighting

► [Road Lighting](#)

Punctum Cecum

► [Blind Spot](#)

Pure Colors

► [Primary Colors](#)

Purkyně, Jan Evangelista

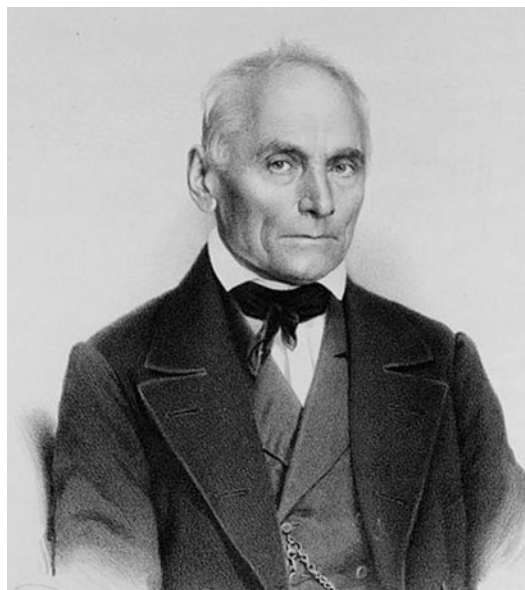
Michal Vik¹ and Renzo Shamey²

¹Faculty of Textile Engineering, Laboratory Color and Appearance Measurement, Technical University of Liberec, Liberec, Czech Republic
²Color Science and Imaging Laboratory, College of Textiles, North Carolina State University, Raleigh, NC, USA

Biography

Jan Evangelista Purkyně (Fig. 1) was born sometime between 17 December and 19 December 1787 in the castle in Libochovice near Litomeřice, Bohemia (then part of the Austrian monarchy) and now Czech Republic [1]. Purkyně died on 28 July 1869.

He was the oldest son of Josef and Rozálie (born Šafránková) Purkyně. His father worked as the chamberlain of Libochovice manor. Jan learned to observe nature and people initially from his father, who used to take him on business trips. He inherited diligence and discipline and



Purkyně, Jan Evangelista, Fig. 1 Purkyně, Jan Evangelista (1787–1869)

also a sense of humor from his mother. Jan began to read early and was introduced to Komenský's publication *Orbis Pictus*. He also learned to read Latin and Greek texts from a priest called Schiffner in Libochovice.

In 1793, Purkyně's father died and Rozálie was left with her two sons Jan and Josef. With assistance from Rozálie's friends, Jan was enrolled at the Piarist Grammar school. After graduation from the grammar school (1804) and upon recommendation of his teachers, Jan went to a monastic house in Dobrá Voda. Due to his excellent performance, he managed to complete the 3-year program in 1 year. He then started to teach at the lower grammar school in Strážnice, and from there, he went to teach at the Piaristic Institute in Litomyšl. There, he devoted himself to the study of philosophy and history and poetry and became a devoted fan of the classical German philosopher F. W. Schelling.

In 1807, Jan returned home, but in the fall of the same year, he started to study philosophy in Prague University. He had to leave his studies after the third year because of financial reasons and accept to teach Baron Hildprandt's son Ferdinand in Blatná. With Hildprandt's financial support, Jan went back to Prague in 1812 to study medicine.

Upon completion of his medical program, he unsuccessfully applied to several faculty positions including professor of pharmacology in Prague, professor of anatomy and physiology in Graz, and professor of anatomy in Ljubljana.

In 1822, he secretly traveled to Germany and was introduced to Schultz and subsequently met with Goethe. In 1823, and upon the recommendation of Goethe, the Prussian king signed a decree which nominated Purkyně a professor in Physician School of Wroclaw (Breslau in Prussia) where he continued his research. There he lectured physiology, eye pathology, and psychology, but his lectures were not well attended. He continued his study of subjective visual perception and published a comprehensive work on this topic which was dedicated to Goethe. In 1826, he was accepted as a Masonic Trainee in Berlin, 1 year later as Journeyman and subsequently as Master. In 1825, he published his work "Neue Beiträge

zur Kenntnis des Sehens in Subjektiver Hinsicht" (Reimer, Berlin, pp. 109–110), which earned him the eponymy Purkinje effect (or Purkinje shift) [2].

In 1827, Jan Evangelista Purkyně (a protestant) got married to Julie Anežka Rudolphi (a catholic) in Berlin, and they kept their respective denominations. With Julie, he had two daughters and two sons. Both of his daughters died of cholera. Purkyně became a member of the Leopoldine science academy in 1829. In 1830, he became a professor of botany at Wroclaw University. In 1834, his wife died, leaving Purkyně with two young sons. He did not remarry.

Purkyně turned his focus to botany and was awarded the Montyon Prize in France for his monograph "De cellulis antherarum fibrosis nec non de granorum pollinarium formis" in 1833. In 1836, the State Secretary accepted Purkyně's suggestion for the establishment of a physiological institute (the first in Central Europe). In 1849, Purkyně was called back to Prague as a professor by the emperors' decree. Purkyně was later nominated as a member of London's King Society. Purkyně was also honored as the guardian of Maticе česká – a committee for edification of language and literature from 1852 to 1858. Thanks to Purkyně, the first Czech industry school (1857) was also opened in Prague where he became the principal from 1857 to 1859. His house was also a place where many Czech artists met and discussed work.

Major Accomplishments/Contributions

Subjective Visual Phenomena (1818–1825)

Jan Purkyně was one of the best known anatomists and physiologists of his time. He was probably one of the early pioneers of what we now call vision science. As a medic, Purkyně studied differences in subjective visual phenomena including lightshade, galvanic and vassal patterns (embranchment of vassals in his own eye), glare patterns, subjective feelings in darkness (phosphenes), blind spot, unity of both eyes' visual fields, double sight, indirect sight,



Purkyně, Jan Evangelista, Fig. 2 Observing *red* and *blue* flowers under average (*left*), dim (*middle*), and dark (*right*) illumination conditions. The simulation shows that

under low levels of illumination levels, the *blue* colors appear much brighter compared to *red* ones

colorblindness in peripheral retina, light patterns, and afterimages. Probably, one of the most well-known contributions of Purkyně to color and vision science is that which bears his name, the Purkyně effect.

As a medical student during spring walks through the flowering countryside, Jan observed that after the sunset, the color of flowers appeared to change; red blossoms appeared darker, yellows faded, and the blues seemed brighter. Purkyně studied this phenomenon systematically and found out the same results with other object colors. That is, under decreased illumination, blue colors appeared lighter than reds. He focused his studies on this topic, and in 1818, he defended his dissertation in the faculty of medicine on “Contributions to the knowledge of sight from a subjective sense” [3]. The effect is shown in the picture of blue and red flowers simulated under three illumination conditions: average (when cones are active), dim (when rods intrude), and dark (when cones become mostly deactivated). This phenomenon is due to the contribution of rods on the perceived color of the scene. Rods require low levels of illumination to become activated (and are associated with night vision), whereas cones are activated at much higher levels of illumination (and are associated with day and color vision). Under certain conditions, e.g., during dawn or dusk, there is just sufficient illumination to activate cones, but illumination is low enough to activate rods also. If the contribution of rod signals to the overall image and thus the perceived color exceeds about 10 %, the effect

becomes known as rod intrusion. Since rods’ peak sensitivity is around 496 nm, which is much closer to the peak sensitivity of the short-wavelength-sensitive cones (responsible for blue colors at 419 nm) compared to that of long-wavelength-sensitive cones (responsible for red colors at 558 nm), the overall perception of the observed scene is shifted toward blue at low levels of illumination when rods become activated. Therefore, blue colors appear brighter, whereas red colors appear very dull and almost black (Fig. 2).

Objective Eye Examination Techniques (1823)

Purkyně recommended a technique for systematic objective examination of eyes using reflective pictures. In this process, candle flame mirrors on the front and back of the cornea and then on the front and back of the retina. Purkyně examined the possibility of using reflective pictures to measure the curvature of cornea (which became the principle of keratometry and ophthalmometry) and its use in diagnosis of eye diseases and defects [4]. After many years, Purkyně obtained an achromatic Plossl microscope, which he placed in his own apartment since he did not have a suitable place within the university. He was deeply interested in the stomach’s mucosa structure and discovered stomach glandule. He also observed eye luminescence and the possibility of observing eyes’ background in vivo, a principle of ophthalmometry which was later established by Helmholtz in 1850 [5].

References

1. Bhattacharyya, K.B.: *Eminent Neuroscientists: Their Lives and Works*, p. 182. Bimal Kuman Dhur of Academic Publishers, Kolkata (2011)
2. J. Purkinje: *Neue Beiträge zur Kenntniss des Sehens in subjectiver Hinsicht* [Observations and Experiments on the Physiology of the Senses: New Contributions to the Knowledge of Vision in Its Subjective Aspect] 2 vols., 192 pp. Georg Reimer, Berlin (1825) (in German)
3. Purkinje, J.: *Beobachtungen und Versuche zur Physiologie der Sinne*. Erstes Bändchen, Beiträge zur Kenntniss des Sehens in subjectiver Hinsicht. Doctoral thesis (1819) (in German)
4. Purkinje, J.: *Commentatio de examine physiologico organi visus et systematis cutanei* [Contributions to physiological research of sight and skin system]. Doctoral thesis, University of Wrocław, Wrocław (1823) (in Latin)
5. Wade, N.J., Brožek, J.: *Purkinje's Vision: The Dawning of Neuroscience*. Lawrence Erlbaum Associates, Mahwah (2001)

Q

Quartz Halogen Lamps

► [Halogen Lamp](#)

R

Rare Earths

► [Lanthanoid Ion Color](#)

Rare-Earth Elements

► [Lanthanoid Ion Color](#)

Rare-Earth Ions

► [Lanthanoid Ion Color](#)

Rational Models of Color Category Learning

► [Bayesian Approaches to Color Category Learning](#)

Rayleigh and Mie Scattering

David J. Lockwood
Measurement Science and Standards, National
Research Council Canada, Ottawa, ON, Canada

Synonyms

[Elastic light scattering](#); [Elastic scattering](#); [Lorenz-Mie theory](#); [Lorenz-Mie-Debye theory](#); [Mie solution](#); [Mie theory](#)

Definitions

Rayleigh scattering refers primarily to the elastic scattering of light from atomic and molecular particles whose diameter is less than about one-tenth the wavelength of the incident light.

Rayleigh line refers to the unshifted central peak observed in the spectroscopic analysis of scattered light.

Mie scattering refers primarily to the elastic scattering of light from atomic and molecular particles whose diameter is larger than about the wavelength of the incident light.

Thomson scattering is elastic scattering of light from free electrons.

Raman scattering is inelastic scattering of light from objects whereby the scattered photon has a lower (Raman Stokes scattering) or higher (Raman anti-Stokes scattering) energy than the incident photon.

Introduction

From ancient times, people have gazed up at the sky in daylight and asked the perennial question “Why is the sky blue?” [1]. Other similar and related questions are “Why is the night sky black?,” “Why are sunrises and sunsets red?,” and “Why are the clouds white?” Rayleigh [2–5] and Mie scattering [6] lie behind the long-sought answers to all such questions about the colors seen in the sky.

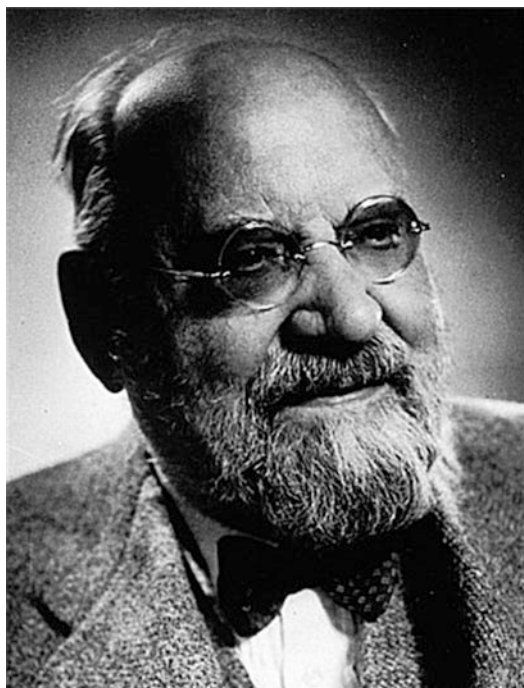


Rayleigh and Mie Scattering, Fig. 1 Lord Rayleigh (1842–1919) ca. 1870

The phenomenon called Rayleigh scattering is named after its discoverer, John William Strutt, 3rd Baron Rayleigh (Fig. 1), a physicist who is known for a number of significant scientific discoveries that he made toward the end of the nineteenth century including the discovery of argon, an achievement for which he earned the Nobel Prize for Physics in 1904. However, the use of the term “Rayleigh scattering,” in general, and also “Rayleigh line,” in specific use in light scattering spectroscopy, has been controversial in the past and has led to some confusion [7]. These nomenclature issues will be clarified in the following discussion.

It is also noteworthy that Rayleigh contributed widely to scattering theory in eight different categories; for an overview of these different contributions, see Reference [8]. However, this entry is limited to a description of classical Rayleigh scattering of light from single small objects such as molecules and atoms. Thus, it is ignoring coherence effects that arise in solids, liquids, and gases at atmospheric pressure or even free electrons where it is known as Thomson scattering.

The scientific history behind the development of the term “Mie scattering” or “Mie theory” is

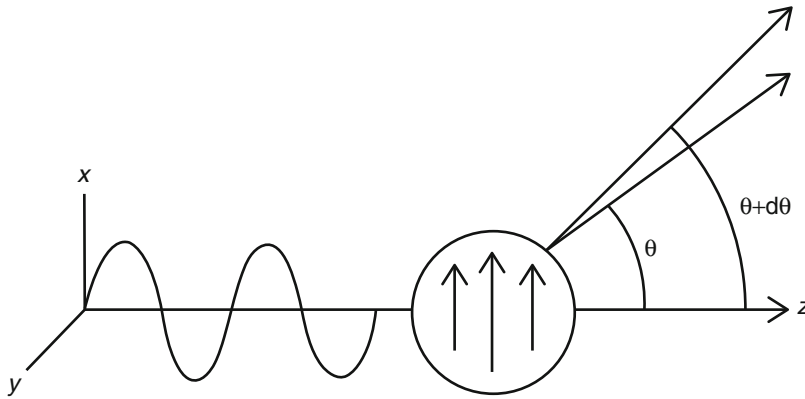


Rayleigh and Mie Scattering, Fig. 2 Gustav Mie (1868–1957)

convoluted [9]. Mie theory solves one of the most important problems in optics – the absorption and scattering of light by a small sphere of arbitrary size and refractive index. In 1908, Gustav Mie (Fig. 2) published his now famous paper on the simulation of the color effects connected with colloidal gold particles [6] wherein he applied Maxwell’s electromagnetic theory to calculate light scattering from small spherical particles. He was not the first to perform this calculation, as similar investigations were also carried out and published earlier in 1863 by Alfred Clebsch and in 1890 by Ludvig Lorenz and soon afterwards in 1909 by Peter Debye [9]. However, Mie was the first to apply Maxwell’s equations to find a solution and his name is now firmly attached to the theory, which is sometimes also called Lorenz-Mie theory and Lorenz-Mie-Debye theory.

Classical Description

Rayleigh treated the scattering of light by a spherical particle, whose diameter is much smaller than



Rayleigh and Mie Scattering, Fig. 3 A plane wave polarized in the x - z plane is incident on the dielectric sphere from the left along the z direction. The arrows on the sphere denote the polarization of the dielectric material. Part of the scattered wave is shown scattered between angle

θ and $\theta + d\theta$ (Reproduced with permission from A.J. Cox, A.J. DeWeerd and J. Linden, *Am. J. Phys.* 70, 620–625 (2002). Copyright 2002, American Association of Physics Teachers)

that of the wavelength of light. It is generally accepted that this size regime applies for particles (or molecules) of size less than $1/10$ th the vacuum wavelength, λ , of the incident light. Rayleigh considered a plane wave incident on a dielectric sphere of radius r and of relative permittivity (dielectric constant) ϵ , as illustrated in Fig. 3 [10]. The probability that the sphere scatters light at angle θ is proportional to the differential scattering cross section, $d\sigma(\theta)/d\Omega$, which is defined as the ratio of the power scattered into the solid angle $d\Omega$ between θ and $d\theta$ (see Fig. 3) to the incident power per unit area. For unpolarized incident light, the differential cross section becomes [10]

$$\frac{d\sigma(\theta)}{d\Omega} = (1/2) (2\pi n_0/\lambda)^4 r^6 \frac{[(\epsilon - \epsilon_0)/(\epsilon + 2\epsilon_0)]^2 (1 + \cos^2\theta)}{(\epsilon - \epsilon_0)/(\epsilon + 2\epsilon_0)}, \quad (1)$$

where ϵ_0 is the relative permittivity and n_0 the refractive index of the medium surrounding the sphere of refractive index n . Integrating this equation over the entire solid angle yields the total cross section:

$$\sigma_R = (8\pi/3) (2\pi n_0/\lambda)^4 r^6 \frac{[(\epsilon - \epsilon_0)/(\epsilon + 2\epsilon_0)]^2}{(\epsilon - \epsilon_0)/(\epsilon + 2\epsilon_0)}. \quad (2)$$

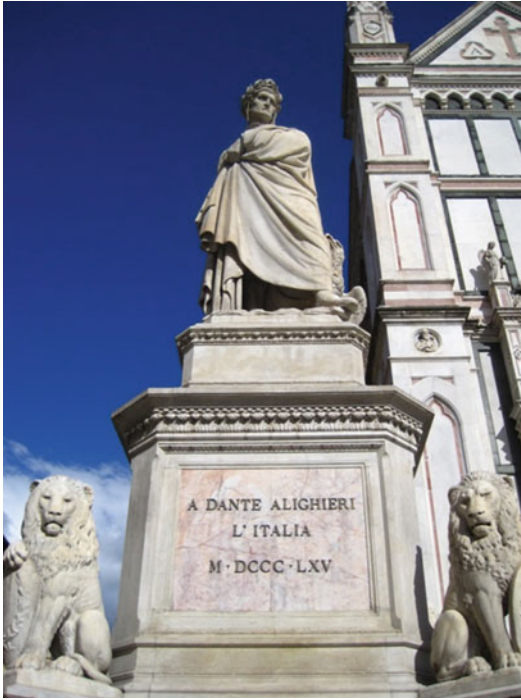
The intensity I_R of the Rayleigh scattered light is given by [11]

$$I_R = I (2\pi n_0/\lambda)^4 (r^6/2D^2) \frac{[(\epsilon - \epsilon_0)/(\epsilon + 2\epsilon_0)]^2 (1 + \cos^2\theta)}{(\epsilon - \epsilon_0)/(\epsilon + 2\epsilon_0)}, \quad (3)$$

where I is the intensity of the unpolarized incident light and D is the distance between the particle and the observer.

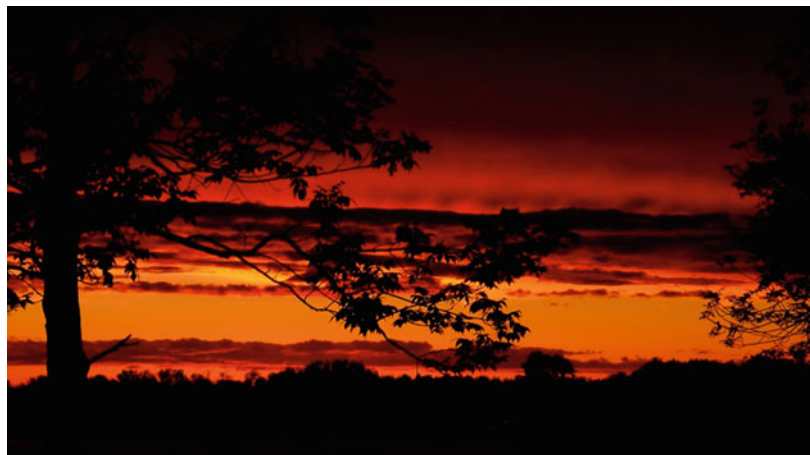
It is now possible to explain why the sky overhead looks bright blue as illustrated perfectly in Fig. 4 that shows Rayleigh scattering of sunlight giving rise to the intense blue color of the sky in a photograph of Dante Alighieri's effigy in Florence, Italy. Equation 4 indicates that the Rayleigh scattering intensity varies as the fourth power of the inverse wavelength of the incident light. This is known as Rayleigh's law. For example, the intensity of Rayleigh scattering at 400 nm in the blue part of the visible spectrum is higher by a factor of 9.38 compared with that at 700 nm in the red portion of the spectrum. So, when considering Rayleigh scattering of visible sunlight from molecules in the air, it is clear that blue light will be brighter than red light. Conversely, at sunset, when the sunlight travels further through the Earth's atmosphere, more blue light is scattered out of the beam path than red light, so the sunset

glows red (Fig. 5). The $(1 + \cos^2\theta)$ term in Eq. 3 contains the angular dependence of the scattering and indicates that Rayleigh scattering intensity at 90° is one-half that in the forward or backward directions.



Rayleigh and Mie Scattering, Fig. 4 Dark blue skies overhead in Florence, Italy: Statue of Dante in the Piazza di Santa Croce in Florence (Enrico Pazzi, 1865) (Copyright Lilia R. Lockwood)

Rayleigh and Mie Scattering, Fig. 5 This photograph of the sunset seen through clouds illustrates both Rayleigh and Mie scattering (Copyright Terry D. Cyr)



For particle sizes larger than λ , Mie scattering predominates [6, 12] and for particles that are much larger than λ , a third type of atmospheric scattering, known as nonselective scattering, occurs [13]. A description of this last type of scattering, which can be considered as comprising a combination of Mie scattering, absorption, and multiple scattering, is outside the scope of this entry. Nonselective scattering is not wavelength dependent and is the primary cause of haze in the lower atmosphere. Water droplets and large dust particles can cause this type of scattering.

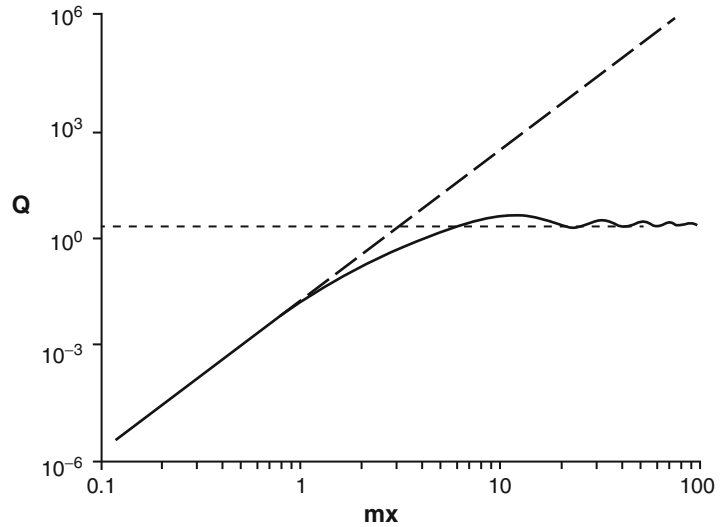
A mathematical description of the physics of Mie scattering is the result of a considerably more complicated calculation than that needed for Rayleigh scattering [12]. The calculation involves considering a plane wave incident on the sphere and a spherical scattered wave. The total scattering cross section, σ_M , can be expressed in the form of an infinite series as [10]

$$\sigma_M = (\lambda^2/2\pi n_0^2) \sum (2m + 1) (|a_i|^2 + |b_i|^2), \quad (4)$$

where the summation is from $i = 1$ to infinity and the coefficients a_i and b_i are expressed in terms of spherical Hankel functions and spherical Bessel functions of the first kind; these functions are dependent on the magnetic permeabilities of the sphere and surrounding medium and the values of parameters $m = n/n_0$ and $x = 2\pi n_0 r/\lambda$, which is termed the size parameter.

Rayleigh and Mie Scattering,

Fig. 6 Scattering efficiency $Q = \sigma/\pi r^2$ versus mx ($x = 2\pi n_0 r/\lambda$ and $m = n/n_0 = 1.59/1.33$) for Rayleigh (*dashed curve*) and Mie (*solid curve*) scattering. The *dotted line* indicates the limiting value of $Q = 2$ (Reproduced with permission from A.J. Cox, A.J. DeWeerd and J. Linden, *Am. J. Phys.* 70, 620–625 (2002). Copyright 2002, American Association of Physics Teachers)



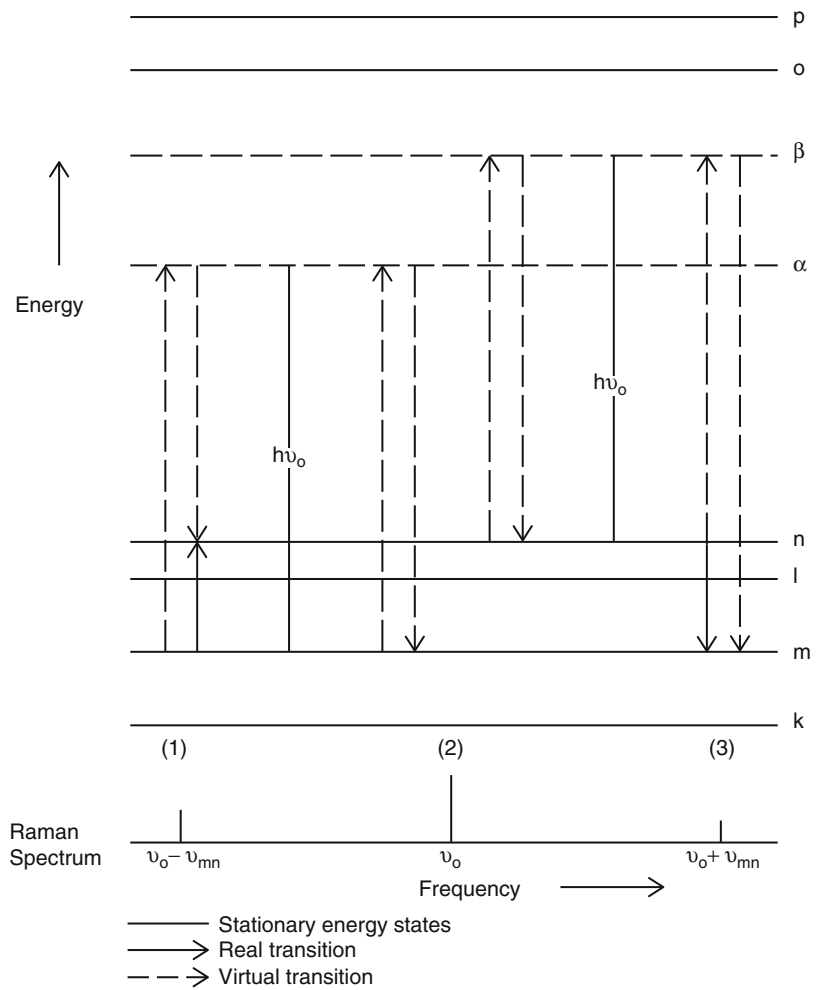
Representative results from calculations of the Rayleigh and Mie scattering cross sections using Eqs. 2 and 4, respectively, are shown in Fig. 6 in terms of the scattering efficiency, $Q = \sigma/\pi r^2$. This figure demonstrates that for small spheres with $mx \ll 1$, ($x = 2\pi n_0 r/\lambda$ and $m = n/n_0 = 1.59/1.33$), the scattering efficiencies of Rayleigh and Mie scattering are very nearly the same, whereas for larger spheres with high mx values, the behavior is completely different. The Mie scattering efficiency departs from the Rayleigh λ^{-4} behavior and approaches a limiting value of $Q = 2$, as a consequence of the “extinction paradox” related to equal geometrical and diffraction contributions to the cross section [10]. The diffraction contribution is most readily discerned at a distance from the sphere and is strongly peaked in the forward direction. Thus Mie scattering produces a scattering pattern like an antenna lobe with a forward lobe that becomes more intense and sharper with increasing particle size. For larger particle sizes, as Fig. 6 shows, Mie scattering is not strongly wavelength dependent. This is the reason that such scattering from water droplets in clouds, mist, or fog produces white light, as can be seen in Fig. 4 in the lower part of the sky behind the lion. Likewise, due to its larger forward lobe, Mie scattering produces the whitish glare that is seen surrounding the sun when large particulate matter is present in the air.

Quantum Description

In describing Rayleigh scattering, only the classical interaction of light with small particles has been considered so far. However, for very small particles, a proper account of Rayleigh scattering requires a quantum-mechanical description. This is given, for example, by Placzek [14] in the early days of quantum theory and, in a more modern setting, by Loudon [15], which is followed here. The scattering of light by an atom or molecule is a second-order radiative process: a quantum (or photon) of energy $\hbar\omega$ (ω is the angular frequency and $\hbar = h/2\pi$, where h is Planck’s constant) of the incident light beam is destroyed and a quantum of scattered light of energy $\hbar\omega_s$ is created. When $\omega_s = \omega$, the process is called elastic scattering or Rayleigh scattering in the nomenclature of this field of research, which is where the use of terminology leads to naming confusions for different scattering processes [7]. When $\omega_s \neq \omega$, the process is called inelastic scattering or Raman scattering; the energy difference thus created is accommodated by the atom or molecule that is involved in the scattering process. Energy may be transferred to, or taken from, the atom or molecule via an electronic transition, in which case it is called Raman Stokes, or Raman anti-Stokes, scattering. These three different electronic energy transition processes are illustrated in Fig. 7. Electronic

Rayleigh and Mie Scattering,

Fig. 7 Quantum energy level transition representation of (1) Raman Stokes scattering, (2) Rayleigh scattering, and (3) Raman anti-Stokes scattering. The corresponding peaks created in the light scattering spectrum are shown below the three sets of different virtual transitions (After David J. Lockwood, Ph.D. Thesis, Canterbury University, 1969)



transitions can occur between any states but usually arise from or terminate at the ground state or thermally populated higher states, and the excited states are usually virtual states, that is, they are not stationary states of the system. However, if the energy of the incident or scattered light is close to or at a stationary energy state, electronic resonant scattering occurs.

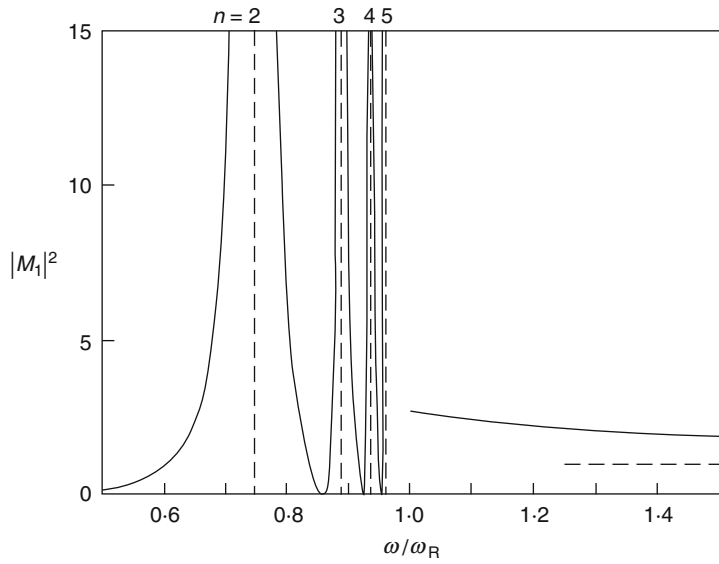
After a full quantum-mechanical derivation, the differential cross section of the elastic scattering of light by an atom, $d\sigma/d\Omega$, can be written as [15]

$$d\sigma/d\Omega = (e^4 \omega^4 / 16\pi^2 \epsilon_0^2 \hbar^2 c^4) \left| \sum \{ [\epsilon_s \cdot \mathbf{D}_{li} \epsilon \cdot \mathbf{D}_{il}] / (\omega_i - \omega) + [\epsilon \cdot \mathbf{D}_{li} \epsilon_s \cdot \mathbf{D}_{il}] / (\omega_i + \omega) \} \right|^2, \quad (5)$$

for the case of the atom returning to its ground state $i = 1$ at the conclusion of the scattering event. The summation index i runs over all the eigenstates of the atomic Hamiltonian. Note that the cross section varies as the fourth power of the frequency of the incident light, as in the classical result. This equation is a simplified form of the general Kramers-Heisenberg formula for the differential cross section that also covers inelastic Raman scattering [15] and forms the basis for quantum-mechanical scattering theory. In Eq. 5, the summation is carried out over all stationary states i ; c is the velocity of light, e the electron charge, ϵ and ϵ_s are the unit polarization vectors for the incident and scattered light, and \mathbf{D} is the sum of the electron coordinates and is

Rayleigh and Mie Scattering,

Fig. 8 Frequency dependence of the elastic scattering matrix element $|M_1|^2$ for a hydrogen atom. The *vertical dashed lines* show the positions of resonances with the $n = 2, 3, 4$, and 5 electronic excited states. The *horizontal dashed line* shows the limiting value of $|M_1|^2$ for $\omega \gg \omega_R$ in the Thomson scattering region (From Ref. [15] after the work of M. Gavrilu [16])



proportional to the atomic-dipole moment. Evaluating the matrix elements that determine \mathbf{D} is not trivial and has only been performed for simple systems such as atomic hydrogen [16].

Equation 5 may be simplified under two limiting conditions. The first is where ω is much larger than an atomic excitation frequency ω_i , but much less than mc^2/\hbar where m is the electron mass. In this high-frequency limit (corresponding to Thomson scattering), the cross section reduces simply to [15]

$$d\sigma/d\Omega = Z^2 r_e^2 (\mathbf{\epsilon} \cdot \mathbf{\epsilon}_s)^2 \quad (\text{for } \omega \gg \omega_i) \quad (6)$$

where Z is the number of electrons in the atom and $r_e = e^2/4\pi\epsilon_0 mc^2$ is the classical electron radius. Interestingly, this result from the quantum theory for the elastic scattering cross section in the high-frequency limit is identical to that obtained from the classical theory for Thomson scattering [15].

The second limiting case is where ω is much smaller than all atomic excitation frequencies ω_i , in which case Eq. 5 becomes [15]

$$d\sigma/d\Omega = (e^4 \omega^4 / 16\pi^2 \epsilon_0^2 \hbar^2 c^4) \left| \sum (\mathbf{\epsilon}_s \cdot \mathbf{D}_{1i} \mathbf{\epsilon} \cdot \mathbf{D}_{i1} + \mathbf{\epsilon} \cdot \mathbf{D}_{1i} \mathbf{\epsilon}_s \cdot \mathbf{D}_{i1}) / \omega_i \right|^2. \quad (7)$$

This expression for the cross section in the low-frequency limit can readily be evaluated for atomic hydrogen, yielding [15, 16]

$$d\sigma/d\Omega = (81 r_e^2 / 64) (\omega / \omega_R)^4 (\mathbf{\epsilon} \cdot \mathbf{\epsilon}_s)^2 \quad (\text{for } \omega \ll \omega_i), \quad (8)$$

where $\hbar\omega_R = me^4/32\pi^2\epsilon_0^2\hbar^2$ is the ground state binding energy of hydrogen.

The full expression given in Eq. 5, for the elastic scattering cross section has been determined for hydrogen by Gavrilu [16]. His tables of results for the magnitude of the scattering matrix element as a function of the frequency of the incident light have been graphically depicted by Loudon [15] and some results are illustrated in Fig. 8. For this case, the differential cross section can be written as [15]

$$d\sigma/d\Omega = r_e^2 |M_1|^2 (\mathbf{\epsilon} \cdot \mathbf{\epsilon}_s)^2, \quad (9)$$

where

$$M_1 = (m\omega^2/\hbar) \sum \left(2\omega_i |X_{1i}|^2 \right) / (\omega_i^2 - \omega^2) \quad (10)$$

and X_{1i} are the appropriate electric-dipole matrix elements. Note that in the frequency region below

Rayleigh and Mie

Scattering, Fig. 9 A composite photograph of the Earth viewed from space showing white light scattering from water droplets in clouds and the surrounding blackness of space (Copyright NASA Goddard Space Flight Center)



ω_R , there are several strong electronic transition resonances evident (only those up to the fifth excited state were calculated by Gavrilu), separated by zeros in the differential cross section. Such resonances could not be predicted from the classical model used by Rayleigh. The inclusion of radiative damping in the theory removes the infinite values from the cross section and the accompanying zeroes evident in Fig. 8. Note also that at high frequencies, the cross section value approaches that of Thomson scattering, as given by Eq. 6 with $Z = 1$, and eventually reaches it when $\omega \gg \omega_i$.

Representative Scattering Media

Here we review Rayleigh and Mie scattering processes at work in different media of interest, including solids, liquids, and gases. Some examples of practical applications of Rayleigh and Mie scattering in these representative media are also provided.

In fluids, conventional Rayleigh scattering is most commonly observed. Although proof of Rayleigh's law in gases was obtained quite early

on [2, 17], at just a few isolated wavelengths, it was not until later that similar information was obtained over a wide wavelength range. In 1973, Stone [18] reported on measurements of the Rayleigh scattering from CCl_4 and C_2Cl_4 as a function of wavelength between 600 and 1,060 nm by placing the liquid sample in a hollow fused-quartz fiber and measuring the light scattered by the liquid through the fiber wall. Stone determined that the scattering loss rate was 25 dB/km for CCl_4 and 68 dB/km for C_2Cl_4 at 632.8 nm and that the scattering loss rate followed a λ^{-4} dependence over the entire spectral range.

As mentioned in the Introduction, Rayleigh scattering is most commonly encountered in nature with the spectacular variations in color of the sky during the day and at sunset. In the absence of these scattering effects from atmospheric gases, the sky would be black, just like in outer space (Fig. 9). In practical applications, Rayleigh scattering from gaseous molecules is becoming an increasingly useful tool for diagnostic purposes. For example, Rayleigh imaging is particularly advantageous for visualizing the complex flow fields in fluid dynamics and combustion phenomena associated with shock waves and



Rayleigh and Mie Scattering, Fig. 10 Rayleigh scattering of white light from an Australian opal (Copyright David J. Lockwood)

boundary layer structures having large density fluctuations [19]. In these measurements, quantitative information on these complex structures is generated by using high-power short-pulse lasers that freeze the density field in time.

In solids, there are many examples of Rayleigh scattering at work. For example, Rayleigh scattering is mainly responsible for the blue appearance of opalescent materials when viewed under white light (Fig. 10), although photonic band gap effects related to their microscopic structure frequently introduce striking color variations in their appearance, as can be seen in Fig. 10. The transmission of optical signals in optical fibers [20] and within microspheres or microdisks [21] is strongly affected by optical scattering processes including significant Rayleigh scattering. Glass fibers are disordered materials that exhibit microscopic variations in their density and refractive index. Rayleigh scattering from these fluctuations results in unwanted energy losses in optical transmission through the fiber with an attenuation coefficient, α_R , given by [20, 21]

$$\alpha_R = (8\pi^3/3\lambda^4)n^8p^2kT\beta_T \quad (11)$$

where n is the fiber refractive index, p the photoelastic coefficient of the glass, k the

Boltzmann constant, β_T the isothermal compressibility, and T is a fictitious temperature at which the density fluctuations are “frozen” into the glass ($\sim 1,500$ K for fused silica). Note that as the contribution of Rayleigh scattering to the attenuation coefficient scales with the inverse fourth power of the wavelength, the Rayleigh scattering losses predominate at shorter wavelengths. A more unusual example is the observation of recoilless Rayleigh scattering by atoms in solids. This has been observed, for example, by employing equipment designed for Mossbauer effect studies (i.e., photon sources and analyzers with extreme selectivity in energy) to observe Rayleigh scattering in Pt, Al, graphite, and paraffin using an incident light energy of 23.8 keV [22].

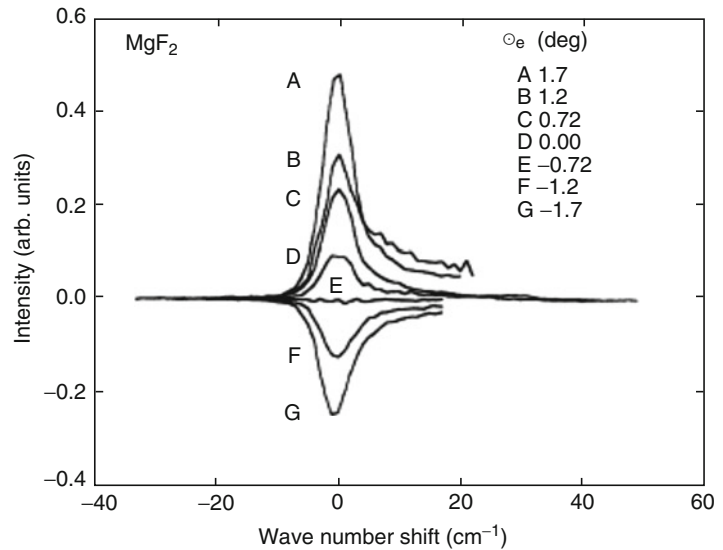
Unusual Rayleigh scattering effects can also be seen in solids. For example, optical studies of gases contained in nanoporous materials with 70 nm diameter pores show a λ^{-4} Rayleigh-type decay of the scattering coefficient with increasing wavelength [23]. Measurements of light scattering spectra from rutile structure fluorides using circularly polarized incident light have revealed Rayleigh optical activity. This occurs when the incident light is tilted away from the crystal c -axis direction of the tetragonal crystal structure [24]. Rayleigh optical activity is observed experimentally through a difference in the intensity of the Rayleigh scattering when excited with right and left circularly polarized incident light. The Rayleigh circular intensity differential (CID), Δ_α , is defined by

$$\Delta_\alpha = (I_\alpha^R - I_\alpha^L)/(I_\alpha^R + I_\alpha^L) \quad (12)$$

where I_α^R and I_α^L are the scattered light intensities with α polarization due to right and left circularly polarized incident light, respectively. The dependence of the CID spectrum on small rotations of the crystal about the laboratory y -axis was measured in optically transparent MgF_2 and in crown glass for reference purposes. For the case of the crystal $c(a)$ axis nearly aligned along the direction of the monochromatic incident(scattered) laser light, the strong dependence of the Rayleigh CID spectrum on the crystal angle is shown in Fig. 11.

Rayleigh and Mie

Scattering, Fig. 11 CID spectrum of the Rayleigh scattering in MgF_2 with the crystal rotated by angle θ_e around the laboratory y-axis (crystal a -axis). The crystal c -axis is in near alignment with the laboratory z-axis (After Ref. [24])



Most importantly for correct identification purposes, the CID changes sign as the crystal is rotated through the true c -axis alignment at $\theta_e = -0.7^\circ$. Notably, no CID is observed for small rotations of the crystal about the orientation where the incident light and scattered light propagate perpendicular to the optic (c) axis, nor is any CID observed from the crown glass. These results indicate that the Rayleigh scattering CID observed in the MgF_2 crystal is due to depolarization effects associated with its birefringence and not polarization-dependent scattering cross sections, as confirmed theoretically [24].

Mie scattering is also found everywhere in nature: in the lower atmosphere, as noted above, in fluids like milk and latex paint, and even in biological tissue. In the latter case, Mie theory has been applied to determine if scattered light from appropriately treated tissue can be used to diagnose cancerous from healthy cells [25, 26]. Mie scattering is used in particle size determination for particles in non-absorbing media [27], in the determination of the oil concentration in polluted water [28], in parasitology [29], and in the design of metamaterials [30].

In summary, it is evident that Rayleigh scattering and Mie scattering are ubiquitous, being found in the everyday and colorful optical wonders that surround us.

Cross-References

- [Glare](#)
- [Light, Electromagnetic Spectrum](#)
- [Maxwell, James Clerk](#)

References

1. Lilienfeld, P.: A blue sky history. *Opt. Photonics News* **15**(6), 32–39 (2004). doi:10.1364/OPN.15.6.000032
2. Strutt, J.W.: On the light from the sky, its polarization and colour. *Philos. Mag. Ser. 4* **41**, 107–120 (1871)
3. Strutt, J.W.: On the light from the sky, its polarization and colour. *Philos. Mag. Ser. 4* **41**, 274–279 (1871)
4. Strutt, J.W.: On the scattering of light by small particles. *Philos. Mag. Ser. 4* **41**, 447–454 (1871)
5. Strutt, J.W.: On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky. *Philos. Mag. Ser. 5* **47**, 375–384 (1899)
6. Mie, G.: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.* **330**(3), 377–445 (1908)
7. Young, A.T.: Rayleigh scattering. *Phys. Today* **35**(1), 42–48 (1982)
8. Twersky, V.: Rayleigh scattering. *Appl. Optics* **3**, 1150–1162 (1964)
9. Hergert, W., Wriedt, T.: *The Mie Theory*. Springer-Verlag, Berlin (2012). doi:10.1007/978-3-642-28738-1_2
10. Cox, A.J., DeWeerd, A.J., Linden, J.: An experiment to measure Mie and Rayleigh total scattering cross sections. *Am. J. Phys.* **70**, 620–625 (2002)
11. Hayes, W., Loudon, R.: *Scattering of Light by Crystals*, p. 2. Wiley-Interscience, New York (1978)

12. Bohren, C.F., Huffman, D.R.: Absorption and Scattering of Light by Small Particles. Wiley-Interscience, New York (1983)
13. Wallace, J.M., Hobbs, P.V.: Atmospheric Science: An Introductory Survey. Academic, Orlando (1977)
14. Placzek, G.: The Rayleigh and Raman scattering. In: Marx, E. (ed.) Handbuch der Radiologie, vol. 6, Part 2, pp. 209–374. Akademische Verlagsgesellschaft, Leipzig (1934)
15. Loudon, R.: The Quantum Theory of Light. Oxford University Press, London (1973), Chapt. 11
16. Gavrilu, M.: Elastic scattering of photons by a hydrogen atom. Phys. Rev. **163**, 147–155 (1967)
17. Cabannes, J.: Sur la diffusion de la lumière par l'air. C. R. Acad. Sci. **160**, 62–63 (1915)
18. Stone, J.: Measurement of Rayleigh scattering in liquids using optical fibers. Appl. Optics **12**, 1824–1827 (1973). doi:10.1364/AO.12.001824
19. Miles, R.B., Lempert, W.R., Forkey, J.N.: Laser Rayleigh scattering. Meas. Sci. Technol. **12**, R33–R51 (2001)
20. Lines, M.E.: Scattering losses in optic fiber materials. I. A new parameterization. J. Appl. Phys. **55**, 4052–4057 (1984). doi:10.1063/1.332994
21. Gorodetsky, M.L., Pryamikov, A.D., Ilchenko, V.S.: Rayleigh scattering in high-Q microspheres. J. Opt. Soc. B **17**, 1051–1057 (2000)
22. Tzara, C., Barloutaud, R.: Recoilless Rayleigh scattering in solids. Phys. Rev. Lett. **4**, 405–406 (1960) and “Erratum” 539
23. Svensson, T., Shen, Z.: Laser spectroscopy of gas confined in nanoporous materials. Appl. Phys. Lett. **96**, 021107 (2010)
24. Hoffman, K.R., Yen, W.M., Lockwood, D.J., Sulewski, P.E.: Birefringence-induced vibrational Raman and Rayleigh optical activity in uniaxial crystals. Phys. Rev. B **49**, 182 (1994)
25. Tsai, M.C., Tsai, T.L., Shieh, D.B., Chiu, H.T., Lee, C. Y.: Detecting HER2 on cancer cells by TiO₂ spheres Mie scattering. Anal. Chem. **81**(18), 7590–7596 (2009). doi:10.1021/ac900916s
26. Wang, M., Cao, M., Guo, Z.R., Gu, N.: Generalized multiparticle Mie modeling of light scattering by cells. Chin. Sci. Bull. **58**(21), 2663–2666 (2013)
27. Gompf, B., Pecha, R.: Mie scattering from a sonoluminescing bubble with high spatial and temporal resolution. Phys. Rev. E **61**(5), 5253–5256 (2000)
28. Lindner, H., Fritz, G., Glatter, O.: Measurements on concentrated oil in water emulsions using static light scattering. J. Colloid Interface Sci. **242**, 239–246 (2001)
29. Serebrennikova, Y.M., Patel, J., Garcia-Rubio, L.H.: Interpretation of the ultraviolet–visible spectra of malaria parasite *Plasmodium falciparum*. Appl. Optics **49**(2), 180–188 (2010)
30. Zhao, Q., Zhou, J., Zhang, F.L., Lippens, D.: Mie resonance-based dielectric metamaterials. Mater. Today **12**(12), 60–69 (2009). doi:10.1016/S1369-7021(09)70318-9

Red Blindness

- [Protanopia](#)

Red-Green Deficiency

- [Deutanopia](#)
- [Protanopia](#)

Reference Standard Lamp

- [Standard Lamp](#)

Reflectance Standards

Bor-Jiunn Wen

Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

Definition

A reflectance standard is a physical reference sample that includes ratio values between the total amount of radiation, as of light, reflected by a surface, and the total amount of radiation incident on the surface across the visible spectrum. Reflectance standards are used for the calibration and verification of spectrometers. There are two kinds of reflectance standards: diffuse and specular. In practice, white and black reflectance standards are utilized for calibration. They require a similar specular or diffuse surface property to the device under test (DUT).

Applications

Diffuse and specular reflectance standards are used to calibrate colorimeters, reflectometers,

spectroradiometers, bidirectional reflectance distribution function (BRDF) scatterometers [1], ultraviolet–visible (UV–VIS) spectrophotometers, and Fourier transform infrared (FTIR) spectrometers.

Diffuse white reflectance standard samples can be obtained with diffuse reflectance of 98 % or more. Some materials can be carefully sanded (some require water with the sanding) or cleaned to refresh the surface to its maximum reflectance as the surface becomes soiled or contaminated. Such reflectance standards can be used for making illuminance from a luminance measurement of the standard. Moreover, with the proper measurement geometry, luminance factor can also be obtained. Therefore, the measurement geometry is used to calibrate the standard. If the reflectance (or diffuse reflectance) of the standard is around 98 % or 99 %, it usually does refer to the reflectance; that value can then only be used for a uniform hemispherical illumination. For industrial applications, white ceramic tiles are often used for a reflectance of 90 %. If we use an isolated source at a particular angle, there is no reason to expect that the 99 % value is even close to the proper value of the luminance factor for that geometrical configuration [1–3].

For calibration of the spectral reflection, one reference white standard and one reference black standard of calibrated spectral reflectance factor $R_{WS}(\lambda)$ and $R_{BS}(\lambda)$ respectively for the same measurement geometry as in the test configuration can be used [4]. Figure 1 depicts the white and black specular reflectance standards. In many

commercial systems, black wedges are widely used for black specular reflectance standards of optical measurements.

To measure the spectral reflection of the device under test (DUT), set up the measurement geometry and warm up the light source. First, measure the spectral reflectance factors of the reference white standard, $R_{WS}'(\lambda)$, with a spectrophotometer. Second, measure the spectral reflectance factors of the reference black standard, $R_{BS}'(\lambda)$. Third, measure the spectral reflectance factor of the DUT, $R'(\lambda)$. The spectral reflectance factor, $R(\lambda)$, is calculated in Eq. 1.

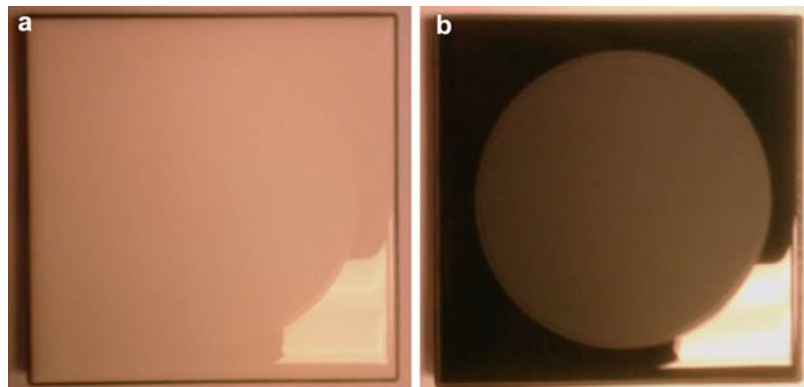
$$R(\lambda) = [R'(\lambda) - R_{BS}'(\lambda)] \cdot \left[\frac{R_{WS}(\lambda) - R_{BS}(\lambda)}{R_{WS}'(\lambda) - R_{BS}'(\lambda)} \right] + R_{BS}(\lambda) \quad (1)$$

where $R_{WS}(\lambda)$ and $R_{BS}(\lambda)$ are the calibrated spectral reflectance factors of the reference white standard and the reference black standard respectively. Then, International Commission on Illumination (CIE) tristimulus values, X , Y , and Z are obtained as

$$\begin{cases} X = k \cdot \sum_{\lambda} R(\lambda) \cdot S(\lambda) \cdot \bar{x}(\lambda) \cdot \Delta\lambda \\ Y = k \cdot \sum_{\lambda} R(\lambda) \cdot S(\lambda) \cdot \bar{y}(\lambda) \cdot \Delta\lambda \\ Z = k \cdot \sum_{\lambda} R(\lambda) \cdot S(\lambda) \cdot \bar{z}(\lambda) \cdot \Delta\lambda \end{cases} \quad (2)$$

where

Reflectance Standards,
Fig. 1 (a) White and (b) Black specular reflectance standards



$$k = \frac{100}{\sum_{\lambda} S(\lambda) \cdot \bar{y}(\lambda) \cdot \Delta\lambda}$$

$S(\lambda)$: the relative spectral power distribution of the illuminant considered

$\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$: color-matching functions of CIE 1931 standard colorimetric observer

$R(\lambda)$: spectral reflectance factor of the DUT

$\Delta\lambda$: wavelength interval

λ : wavelength range from 380 to 780 nm

For the calibration of the filter photometric reflection, one reference white standard and one reference black standard of calibrated CIE tristimulus values (X_{WS} , Y_{WS} , Z_{WS}) and (X_{BS} , Y_{BS} , Z_{BS}) respectively for the same measurement geometry as in the test configuration shall be used [4]. The filter photometric reflection procedure of the DUT is as follows.

First, set up the measurement geometry and warm up the light source. Second, measure the CIE tristimulus values of the reference white standard, X_{WS}' , Y_{WS}' , and Z_{WS}' using a colorimeter. Third, measure the CIE tristimulus values of the reference black standard, X_{BS}' , Y_{BS}' , and Z_{BS}' . Finally, measure the CIE tristimulus values of the DUT, X' , Y' , and Z' . CIE tristimulus values, X , Y , Z are obtained using Eq. 3.

$$\begin{cases} X = [X' - X_{BS}'] \cdot \frac{X_{WS} - X_{BS}}{X_{WS}' - X_{BS}'} + X_{BS} \\ Y = [Y' - Y_{BS}'] \cdot \frac{Y_{WS} - Y_{BS}}{Y_{WS}' - Y_{BS}'} + Y_{BS} \\ Z = [Z' - Z_{BS}'] \cdot \frac{Z_{WS} - Z_{BS}}{Z_{WS}' - Z_{BS}'} + Z_{BS} \end{cases} \quad (3)$$

where X_{WS} , Y_{WS} , Z_{WS} and X_{BS} , Y_{BS} , Z_{BS} are the calibrated CIE tristimulus values of the reference white standard and the reference black standard respectively.

The measurement method of spectral or filtered photometric reflection is performed for the reflection of the DUT by using a spectrophotometer or a colorimeter. Compared with the filter photometric reflection, spectral reflection measurement results include wavelength information. However, a spectrophotometer is more expensive than a

colorimeter. According to practical applications, the measurement method of spectral or filter photometric reflection is appropriately utilized. Typically, color, fluorescence, grayscale, and wavelength reflectance standards are usually used in the color industry. They are used to check the reliability of the instruments used. Various types of standard according to their application are introduced here.

Color Reflectance Standards

Diffuse color standards are used to calibrate colorimeters and spectrophotometers within the range 360–830 nm. Compared with standards for gloss surfaces, the diffuse nature of the standards simplifies measurements by removing the effects of viewing or illumination geometries.

Fluorescence Reflectance Standards

Fluorescence standards aid in the development of optically brightened materials, such as paper and textiles, and are widely used in the cosmetics industry. The appearance of an illuminated object is dependent upon the illumination and viewing conditions. For example, the total radiance factor of paper and board products containing fluorescent whitening agents (FWAs) is the sum of the reflected radiance factor and the luminescent radiance factor [5]. As the luminescent radiance factor originates from fluorescence of the FWAs, its magnitude depends on the amount of UV radiation of the illumination. It is therefore essential to calibrate not only the radiance (reflectance) factor scale, but also the UV content of the illumination used in the measurement apparatus. The adjustment of the UV content is achieved by altering the position of the UV filter in the apparatus so that the light incident upon the sample has an effective UV content corresponding to that in the CIE illuminant C. The UV filter adjustment is based on the ISO brightness ($C/2^\circ$) value [5].

Grayscale Reflectance Standards

Grayscale diffuse reflectance standards are used to establish the linearity and accuracy of reflectance spectrophotometers and colorimeters. The standards are suitable for users who need a wide dynamic range of reflectance values that are

abbreviated in range, but not in number of calibration steps. Industries such as clay and processed minerals, paper, and paints typically require such a range, which accurately reproduces the range of reflectance values exhibited by those products.

Wavelength Reflectance Standards

Wavelength calibration standards provide stable absorption spectra for validating the wavelength calibration of spectrophotometers in the ultraviolet–visible–near infrared (UV–VIS–NIR) region. Complete absorption spectral data are supplied with each standard. A holmium oxide standard is available for UV–VIS–NIR calibrations, while a dysprosium oxide standard is offered for NIR calibrations. Meanwhile, an erbium oxide standard is provided for VIS–NIR calibrations. Alternatively, a mixed oxide standard is available, which is doped with all three rare earth oxides for use over the entire UV–VIS–NIR region.

Cross-References

- [CIE Tristimulus Values](#)
- [Illuminance Meter](#)
- [Instrument: Colorimeter](#)
- [Instrument: Spectrophotometer](#)
- [Standard Measurement Geometries](#)

References

1. VESA 2.0: Flat Panel Display Measurements Standard (2001)
2. CIE 15: Colorimetry. 3rd edn (2004)
3. SID IDMS Information Display Measurements Standard, v1.03, 1 June 2012
4. SEMI D68-0512: Test Methods for Optical Properties of Electronic Paper Displays (2012)
5. ISO 2470–1:2009: Paper, board and pulps – measurement of diffuse blue reflectance factor – Part 1: indoor daylight conditions (ISO brightness) (2009)

Relational Color Space

- [Psychological Color Space and Color Terms](#)

Relative Visibility Function

- [Spectral Luminous Efficiency](#)

Rendering Natural Phenomena

Oskar Elek

Computer Graphics Department, Max-Planck-Institut Informatik, Saarbrücken, Germany

Synonyms

[Simulation and image synthesis of natural phenomena](#)

Definition

In computer graphics, rendering is a process of synthetically generating an image – or a sequence of images – of an object, based on its mathematical and possibly physical description. Natural phenomena are inherently very diverse, and therefore, their rendering is a very heterogeneous field with different paradigms and approaches.

As the name implies the objects and phenomena of interest will mainly have a natural origin. However, due to similar characteristics of the underlying problems, rendering of artificial objects made of natural materials is often considered a part of the field as well.

Although the main target of rendering is the creation of images, it is usually not trivial to obtain the mathematical and physical description of the simulated entities (i.e., the input data). Because of this the rendering algorithms can potentially require coupling with simulation methods, which provide means to computationally generate the required data. Alternatively, acquired or even hand-modeled data can be used, in case the simulation proves to be infeasible for some reason (e.g., too time-consuming).



Rendering Natural Phenomena, Fig. 1 Examples of sparse, fluid, and solid phenomena

Categories of Natural Phenomena

The richness of the natural environment of course implies the existence of a tremendous amount of phenomena and objects. The structure of this system is hardly apparent, but can roughly be classified according to the physical characteristics of the phenomena as follows (also see examples in Fig. 1):

- **Sparse phenomena**, involving gases, aerosols, and vapors but also effects of electromagnetism or high-energy particles. These include various well-known meteorological phenomena like atmospheric scattering, clouds, fog, rainbows, or lightning; astronomical phenomena such as auroras, stars, nebulae, and other stellar bodies; and also smaller-scale phenomena like fire, smoke, and dust.
- **Fluid phenomena**, most notably oceans and other large water bodies and effects associated with their surfaces, such as waves and streams. Also medium- and small-scale liquid substances, especially beverages like milk, fruit juices, and coffee; suspensions like blood, paints, and inks; and volcanic phenomena such as lava flow. In addition, especially from the perspective of simulation methodology, it is possible to regard fine-grained solid materials like sand and partly solid substances such as gels, as fluid phenomena.
- **Solid objects and phenomena**. On a large scale, primarily geological formations ranging from mountains to entire planets. On a medium scale, organic entities like vegetation, biological tissues, hair, and fur and inorganic objects such as ice and rock formations, crystals, metals and their alloys, and man-made objects manufactured from these. Among small-scale solid objects, most effort has been focused on rendering precious gems. Additionally, rendering of natural and artificial solid objects



Rendering Natural Phenomena, Fig. 2 A small quartz crystal and a large piece of sandstone

exhibiting a layered structure has attracted research attention, for example, coated or painted objects, oxidized and patinated metals, composite materials, gemstones, and many others.

From the perspective of rendering and also modeling, another important distinction can be made between phenomena and objects with well-definable geometry and opaque surfaces and those with a dominating volumetric character (either from the spatial or the optical point of view).

The first group will likely be modeled using geometric primitives and rendered with algorithms suitable for distinguishing between discrete parts of the phenomenon, e.g., the object-air interface. The light interaction will primarily be taking place on these interfaces, mainly as reflection and refraction. Most opaque solids belong into this category but also some fluid substances, such as clear liquids.

On the other hand, the second group will likely utilize a volumetric representation (or a combination with a geometric one) and naturally also volume-rendering methods. Materials with these properties are called participating media, and the dominating optical interactions here will be scattering and absorption. These characteristics are inherent to virtually all sparse phenomena but also to most fluid and many solid ones.

This distinction is naturally relative and without a sharp transition. As an example, one can imagine that a small stone grain will be translucent and have an apparent volumetric structure, while a large piece will appear opaque and will interact with light mostly by diffuse surface reflection (Fig. 2).

Causes of Color

Another important aspect of natural phenomena from the perspective of color science is how they interact with light and produce color and the overall appearance [1]:

- **Emission**, due to *incandescence* (e.g., stellar radiation, volcanic activity, lightning), *gas excitation* (auroras, artificial gas lamps), or a combination of these (fire). The variation of color is primarily caused by varying energy density. Virtually all natural light originates in these processes.
- **Geometric causes**, such as *scattering* (atmosphere, clouds, smoke, milk, and generally almost all substances exhibiting volumetric properties or diffuse reflection), *dispersion* (rainbows, sundogs, snow), *diffraction* (opals, thin filament-like objects such as spider webs and hair, carapaces of certain bugs), *interference* (single- or multilayered structures such as bubbles and certain gemstones and insects),

and *polarization* (specular reflection from smooth surfaces). These interactions are usually elastic, i.e., they conserve energy, and the color variation is mainly caused by geometric configurations and relations.

- **Absorption and reemission**, mostly in *organic compounds* (plant and animal tissues, natural and artificial pigments, dyes, and inks) and all *metals* but also in certain minerals and semiconductors and occasionally in light liquids, most notably in *pure water*. These interactions are usually inelastic, leading to energy loss described by the Beer-Bouguer law. Color variation is caused by relative efficiency of the reemission in different wavelengths.

Historical Overview

Synthesizing physically plausible images of natural phenomena has arguably been one of the primary foci of rendering from the beginning [2]. However, the complexity of most natural phenomena prevented their plausible rendering until the early 1980s, mostly because of the lack of theoretical understanding and computational power.

A significant milestone was reached by introducing radiosity [3] and stochastic ray tracing [4] algorithms in 1984, enabling physically based simulation of global illumination effects. A more general solution was then presented in 1986 [5] by James Kajiya in the form of the rendering equation – a unified mathematical framework that enables the simulation of all effects that conform to geometric optics. These approaches were then extended to rendering of volumetric phenomena in the late 1980s and early 1990s.

Algorithmic improvements have gone hand in hand with the evolution of computing hardware, making rendering of image sequences and even entire movies feasible by the end of the 1980s. However, probably the biggest breakthrough came with the introduction of parallel programmable graphics processor units (GPUs) to the consumer market in 2001. Their programmability enables researchers and developers to produce specialized algorithms focused on simulating

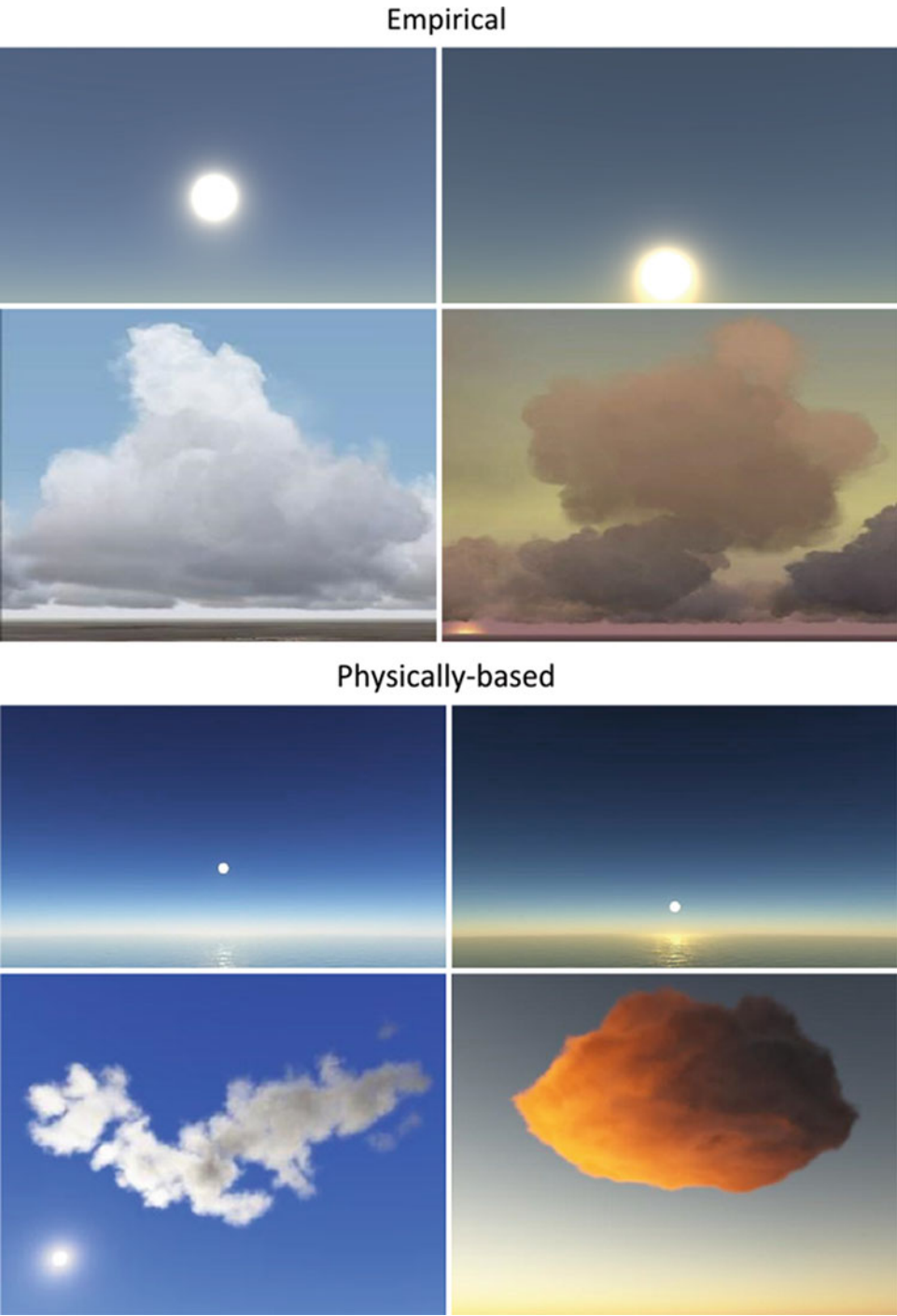
and rendering effects of diverse nature. Programmable GPUs quickly became widespread and enabled the development of algorithms and applications capable of simulating numerous natural phenomena, including video games and other interactive applications.

Rendering Methodologies

The traditional view is to split (realistic) rendering as a field into noninteractive (“offline”) rendering [2] and interactive rendering [6], with the former putting emphasis on quality and the latter on speed. This applies to rendering of natural phenomena just as well. The spectrum of possible approaches is however much more continuous than that and rather depends on the *paradigm* chosen for designing a particular rendering method (refer to Figs. 3 and 4).

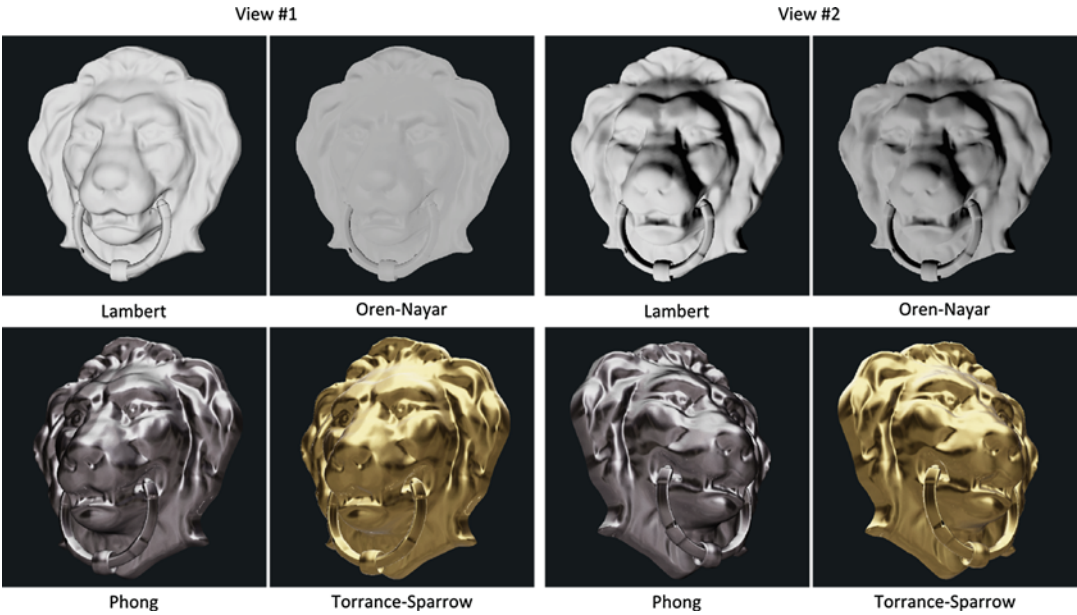
Empirical or phenomenological methods are generally *top-down*; their design starts by collecting observations of a phenomenon to be simulated and then proceeds to invent an arbitrary algorithm which best reproduces the phenomenon while keeping computational costs as low as possible. As such, empirical methods are used mainly in interactive applications where the important factors are visual plausibility, robustness, and speed, while physical plausibility and radiometric accuracy are secondary. Notable examples include:

- **The Lambert reflectance model** (1760) states that reflection from a sufficiently rough surface is perfectly diffuse, i.e., isotropic with regard to observation direction. Despite building on physically meaningful assumptions (multiple scattering underneath the surface) and the fact that many materials (such as uncoated paper) behave very closely to the model, there is no ideal diffuse reflector.
- **The Phong reflectance model** (1975) approximates reflection from glossy surfaces by the cosine function raised to an exponent proportional to the surface smoothness. This leads to reflection in the form of a blurry circular patch which gets brighter and smaller with increasing



Rendering Natural Phenomena, Fig. 3 A comparison of interactive (empirical versus physically based) methods for rendering atmospheric scattering and clouds. The empirical approaches are typically faster, but produce less

dynamic and believable results and in some cases require additional artistic input (Image credit: Petr Kellnhofer using the model of Sean O’Neil, (*top ‘sunset’ pair*), Niniane Wang (*top ‘clouds’ pair*))



Rendering Natural Phenomena, Fig. 4 A comparison of empirical and physically based reflectance models. *Top row:* The Lambert and the (physically based) Oren-Nayar models on a diffuse gray clay material. The Oren-Nayar model correctly produces a “flatter” distribution, which is caused by backscattering on the rough clay surface. *Bottom row:* A glossy object made of gold, reflecting an environment map. The Torrance-Sparrow model produces a more

plausible reflection, which is mainly visible in grazing angles where the Phong model incorrectly produces darker reflections in disagreement with the Fresnel law. In addition, the Phong model fails to reproduce the yellow hue of gold, since it does not handle conductors properly. This could only be obtained by additional manual tweaking

- smoothness and in the limit case corresponds to the Dirac function for a perfect mirror. Despite having no physical basis, the model has been widely adopted and is still used, mainly thanks to its simplicity.
- **Particle systems** (1983) are ubiquitously used in both interactive and offline applications to model volumetric phenomena (such as clouds). In this sense, particles are small semitransparent entities often modeled by mapping a texture onto a rectangular geometric primitive that always faces the observer. If used in sufficient numbers, they are able to mask their discrete nature, while still being cheaper than the corresponding full volumetric representation (e.g., a 3D voxel grid). However, simulating global illumination effects in conjunction with particle systems requires additional effort, which might hinder their utilization. As a result it is often necessary to design another empirical

model to compute illumination in an intended way, increasing the necessary amount of work.

The main downsides of empirical methods are that they often require a lot of artistic supervision during the content creation process and that they seldom work outside the range of phenomena they were designed for.

Physically based methods proceed in the opposite, bottom-up fashion. Using physically meaningful data, these methods apply the laws of physics to render the target phenomenon. Consequently, rather than being centered around a particular simulated phenomenon, these methods share more principal similarities and are usually useful for rendering a whole class of phenomena to which the implemented laws and assumptions apply. On the downside, their computational cost is often much higher than that in empirical approaches. This is usually treated by applying

Occam's razor, simplifying the rendering algorithm in ways that decrease the simulation time but do not negatively affect the result. Examples of such methods include:

- **The Torrance-Sparrow reflectance model** (1992) simulates reflection from rough glossy surfaces (and as such being an alternative to the Phong model). The method considers the simulated surface to consist of microscopic facets oriented according to a certain statistical distribution (i.e., Gaussian distribution). Each facet is assumed to reflect light according to the Fresnel law, and additionally, probabilistic shadowing of neighboring facets is taken into account. Although in some situations the model produces results similar to the Phong model, it is energy conserving, correctly handles objects made of conductors, and behaves plausibly in grazing angles, albeit being somewhat more difficult to understand.
- **Photon mapping** (1996) [7] is a general framework for rendering global illumination effects. It is applicable to both surface and volume illumination, which makes it especially suitable for rendering natural phenomena. Its main idea lies in distributing the light energy by shooting and tracing small energy particles, photons. Every interaction of a photon with the simulated environment is recorded, and when all illumination energy is distributed, the algorithm calculates the energy density at observed locations by performing local photon density estimation. Much work has been invested into improving the original technique, resulting in one of the most versatile global illumination algorithms to date.

Originally, physically based methods were only used in offline applications. Recently, however, with the increasing performance and flexibility of the programmable hardware, the boundary between offline and interactive methods has gotten weaker. Many physically based rendering algorithms (such as ray tracing or photon mapping) can today be implemented to run interactively, albeit with decreased rendering quality.

Finally, **predictive methods** represent a step further than physically based methods. Opposite to empirical approaches, their primary focus is physical correctness and radiometric accuracy. The utilized algorithms must be spectral, unbiased, and support all significant physical phenomena that occur in the simulated environment. They require measured and acquired input data, and the results usually need to be viewed under controlled conditions. The resulting methods typically need validation and are generally very slow, even compared to physically based approaches. As such they are useful mainly in virtual prototyping applications such as in automotive industry, architecture, and gem processing.

Phenomena Simulation

As mentioned in the "Introduction," in addition to rendering a phenomenon (and thereby producing an image of its momentary appearance), it is often necessary to obtain data about its spatial and temporal development prior to the rendering. Virtually all natural phenomena are dynamic in a certain sense, and this dynamicity can be regarded in the short and the long term.

Short-term development usually stems from the character of the phenomenon itself and its internal dynamics. Its nature can be *synthetic* (e.g., condensation of water vapor leading to cloud formation), *evolutionary* (for instance, flow of liquid particles in a stream), or *destructive* (such as cracking or shattering of an iceberg). Capturing this behavior will often entail using discrete particle or continuous dynamics.

On the other hand, **long-term development** is usually related to interactions of a phenomenon with its surrounding environment [8], sometimes referred to as *weathering* (in case the process is destructive). Many forms of this behavior exist, for instance, terrain formation, soil erosion and cracking, oxidation and patination of metals, organic tissue decomposition, and others.

From the methodological point of view, approaches for simulating natural phenomena can again be divided into empirical and physically based. Very often, however, a combination of



Rendering Natural Phenomena, Fig. 5 Multiple superposed frequencies (octaves) of Perlin noise (*top left*) and examples of procedurally generated content – terrain,

forest, and fire (Image credit: Franck Doassans (forest), Sergei Bolisov (fire))

these two shows to be the best compromise. The cause of this is arguably the fact that while it is often very difficult to devise an empirical method that captures the entire complexity of a given phenomenon well, a complete physically based simulation of any larger system can easily become intractable. Consequently, many approaches simulate the global behavior of the target phenomenon in a physical, possibly simplified way (for instance, using a reduced resolution of the simulated structure) and then add the remaining details empirically. The most frequent way to perform the latter is using *fractal* functions or systems:

- **Terrain rendering** methods usually generate the overall terrain morphology by simulating the orogenetic and erosive processes (or simply use satellite data) and add more detailed features by random fractal perturbations.
- **Rendering of plants** frequently employs the so-called L-systems to generate the plants and trees. L-systems are iterated functions described by grammars that imitate branching in real plants. Adding random perturbations

into the system can produce plausibly looking plants with unlimited number of variants that retain the overall character defined by the L-system.

- **Rendering of ocean waves** can generate larger-scale waves with a fluid dynamics simulation or synthesizing them from trochoidal wave theory using measured wave frequency spectra and then again add random fractal perturbations to break disturbing repetitive patterns.

Despite the convenience of fractal functions, their inherent *iterative* nature makes them difficult to use in some cases. Therefore, efforts have been made to design regular functions with fractal properties. The most successful method to do so is the *Perlin noise* [9], which has become a cornerstone for generating many diverse natural phenomena and also fuelling future development in this area (see examples in Fig. 5). Since Perlin noise behaves as a regular function, it enables on-the-fly evaluation without requiring additional storage, making it feasible for use even in interactive applications.

It might of course be possible to acquire the data needed for rendering a given phenomenon, although in some cases this can prove difficult or impractical for quantitative reasons:

- The phenomenon of interest might be too large and as a result produce overwhelming amounts of data. A good example is acquisition of terrains, which, although today possible via satellite scanning, still does not produce data with resolution sufficient for some applications.
- Dynamic phenomena are generally difficult to acquire, especially in cases when the scanning process is slower than the rate of the phenomenon's significant change. Acquisition of flames and smoke is a good example here.
- Scanning processes yield limited amounts of instances of the target phenomenon. If many such instances are needed (e.g., in cloud rendering), the cost of the acquisition process might become prohibitive.

Similar to physically based simulation, however, these problems can in some cases be overcome by applying procedural perturbation techniques to the scanned data.

Future

The present knowledge in physics theoretically allows us to explain and hence simulate virtually all observable natural phenomena. The limiting factors in doing so are therefore always computational resources. In the future, the increasing memory density will enable us to work with larger natural systems, and the growing parallel computing power of CPUs and GPUs will allow simulation and rendering of more complex phenomena. Especially in interactive applications, the current trend of using physically based approaches over empirical ones will most likely continue.

Cross-References

- [Blackbody and Blackbody Radiation](#)
- [Global Illumination](#)

- [High Dynamic Range Imaging](#)
- [Iridescence \(Goniochromism\)](#)
- [Kubelka, Paul](#)
- [Rayleigh and Mie Scattering](#)
- [Spectral Power Distribution](#)

References

1. Nassau, K.: *The Physics and Chemistry of Color*, 2nd edn. Wiley-Interscience, ISBN 0471391069 Hoboken, New Jersey, USA (2001)
2. Pharr, M., Humphreys, G.: *Physically Based Rendering*, 2nd edn. Morgan Kaufmann, ISBN 0123750792 Burlington, Massachusetts, USA (2010)
3. Goral, C.M., Torrance, K.E., Greenberg, D.P., Battaile, B.: Modeling the interaction of light between diffuse surfaces. *SIGGRAPH Comput. Graph.* **18**, 213–222 (1984)
4. Cook, R.L., Porter, T., Carpenter, L.: Distributed ray tracing. *SIGGRAPH Comput. Graph.* **18**, 137–145 (1984)
5. Kajiya, J.T.: The rendering equation. *SIGGRAPH Comput. Graph.* **20**, 143–150 (1986)
6. Akenine-Moller, T., Haines, E., Hoffman, N.: *Real-Time Rendering*, 3rd edn. AK Peters, ISBN 1568814240 Natick, Massachusetts, USA (2008)
7. Jensen, H. W.: Global illumination using photon maps. In: *Proceedings of EGWR*, pp. 91–100 Porto, Portugal (1996)
8. Dorsey, J., Rushmeier, H., Sillion, F.: *Digital Modeling of Material Appearance*. Morgan Kaufmann, ISBN 0122211812 Burlington, Massachusetts, USA (2007)
9. Perlin, K.: An image synthesizer. *SIGGRAPH Comput. Graph.* **19**, 287–296 (1985)

Retinal Art

- [Optical Art](#)

Retinex Theory

John McCann

McCann Imaging, Arlington, MA, USA

Synonyms

[Color and Lightness constancy](#); [Land Retinex Theory](#)

Definition

Retinex is the theory of human color vision proposed by Edwin Land to account for color sensations in real scenes. Color constancy experiments showed that color does not correlate with receptor responses. In real scenes, the content of the entire image controls appearances. A triplet of L, M, S cone responses can appear any color. Land coined the word “Retinex” (the contraction of retina and cortex) to identify the spatial image processing responsible for color constancy. Further, he showed that color sensations are predicted by three lightnesses observed in long-, middle-, and short-wave illumination. Retinex is also used as the name of computer algorithms that mimic vision’s spatial interactions to calculate the lightnesses observed in complex scenes.

Overview

Edwin H. Land, the inventor of hundreds of film patents, was struck by experiments showing that color sensations in real complex images depend on scene content. Film responds to the light falling on each tiny local region. Land realized that vision’s mechanisms were very different from film. His early experiments studied the colors observed in red and white projections [1]. He realized color appearance required both the cone responses to a local region and the neural spatial processing of the rest of the scene. He proposed the Retinex Theory.

Land coined the word Retinex to describe three independent spatial channels. In 1964 he wrote: “We would propose that all of the receptors with maximum sensitivity to the long-waves in the spectrum, for example, operate as a unit to form a complete record of long-wave stimuli from objects being observed. (For convenience of reference, let us call this suggested retinal-cerebral system a “retinex.”)” [2–5]. It is the word that describes the mechanism that performs the comparison of scene information to create the array of sensations of lightness in three channels.

Cone Quanta Catch

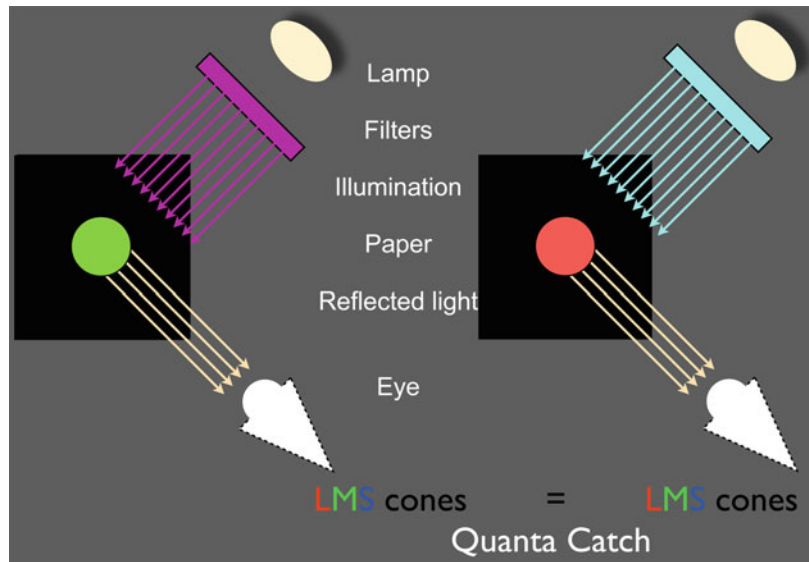
Visible light falls on objects that reflect some of it to the eye. Color vision depends on the spectrum of the illumination falling on an object and the spectrum of its reflectance. The product of these spectra describes the light coming to the eye. There are three types of cones in normal observers that are called L for long-wave-, M for middle-wave-, and S for short-wave-sensitive cones. The receptors’ spectral sensitivities multiplied by the light falling on the retina determines the L, M, S cone responses, namely, the “quanta catch” of the cones. The cones convert the quanta catch to nerve signals that pass through many spatial comparisons in the visual system. The cone quanta catch is the important transition from the physics of light to the physiology of vision. However, it is just the first step in the process.

Figure 1 illustrates a laboratory experiment that generates equal L, M, S quanta catches from different reflectance papers.

Light passes through filters that determine the spectrum of each illumination falling on circular pieces of paper. The light coming to the eye from the papers is modified again by the reflectance spectra of the papers. Further, in this experiment there is no light coming from the black surrounds.

On the left, there is a tungsten light source at the top. There is a filter that absorbs more middle-wave than long- and short-wave visible light (magenta arrows). The green circular paper reflects more middle-wave than long- and short-wave light. The light coming to the eye is the product of these spectra. The integrals of that light using the three cone spectral sensitivities determine the L, M, S quanta catch values.

On the right, there is the same tungsten light source at the top. There is a different filter that absorbs more long-wave than middle- and short-wave light (cyan arrows). The right-red paper reflects more long-wave than middle- and short-wave light. In this experiment, the spectra of the two illuminants and the two reflectances were adjusted to generate the same triplet of LMS cone quanta catches. Under these conditions, the left-green and right-red papers are identical retinal

Retinex Theory,**Fig. 1** Equal quanta catches from different papers

stimuli. They appear equal to each other, but that color is neither red nor green.

Mondrian Experiments

Land's color Mondrian experiment is similar, with the exception that he used a complex array of papers to simulate real-world scenes. The important difference is that in complex scenes, a particular quanta catch can appear in any color: red, green, blue, yellow, white, or black.

Figure 2 shows the double color Mondrian experiment [6]. It used two identical Mondrians made of color papers, and three different, non-overlapping spectral illuminants (long-, middle-, and short-wave visible light). In this experiment observers reported the colors of papers in the Mondrians. In this illustration, we will look at the circular red and green papers. Land adjusted the illuminant mixtures of light from the two sets of three projectors. The same amounts of L, M, S light came from the green circle on the left as from the red circle on the right. First, he turned on just the long-wave lights. He adjusted the amounts of illumination on the left-green and right-red circles so the meter readings were equal. Then, he did the same for middle- and short-wave light. The left-green and right-red

circles had equal quanta catches by the L, M, S cones.

In this complex scene, observers reported that equal quanta catches appeared green on the left and red on the right. Observers reported color constancy, namely, that the red paper looked red and the green paper looked green despite the identical cone quanta catch.

Land repeated this experiment with all the Mondrian papers. A constant L, M, S quanta catch could generate any color sensation. The presence of the complex scene introduced more information to the visual system. The red and green papers appeared equal in Fig. 1. The red paper looked red and the green paper looked green in Fig. 2. The scene's spatial content stimulated vision's spatial image processing mechanisms to generate color constancy. The post-receptor visual processing plays a dominant role in color appearance in real scenes. Land's word "Retinex" gave this spatial process a name. As well, he proposed a theoretical mechanism.

Retinex Mechanism

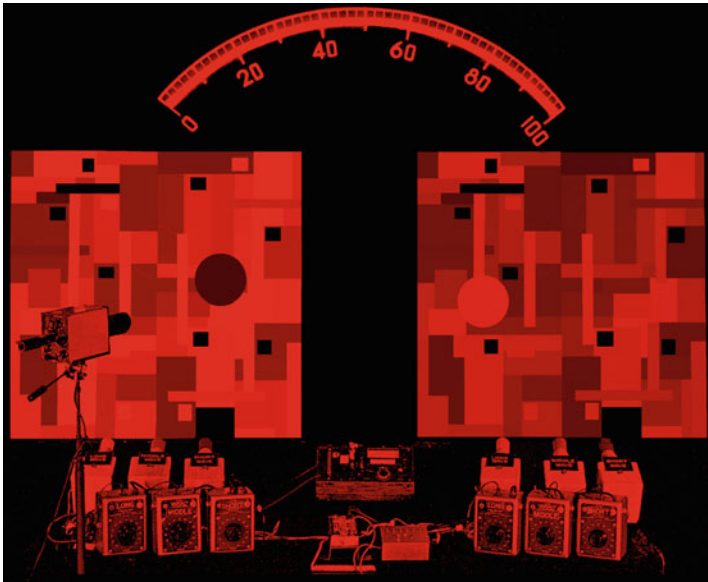
Figure 3 illustrates the pair of Mondrians in only long-wave light with more light on the left than on the right. Their appearances are nearly



Retinex Theory, Fig. 2 Land’s double Mondrian experiment. Two identical sets of matte colored papers, with separate L, M, S illuminating projectors with voltage transformers for control of the amount of light. Telephotometer readings were projected above the Mondrians. The

experimenter separately measured the L, M, S radiances from a *green circle* in the left Mondrian. Then, he adjusted the L, M, S radiances from a *red circle* on the *right* Mondrian to be the same. Observers reported different *red* and *green* colors produced by identical light stimuli

Retinex Theory, Fig. 3 Pair of Mondrians in only long-wave light. The *left* Mondrian has more illumination than the *right*. Observers report that the *left* set is slightly lighter than the *right*. Each corresponding area is nearly the same lightness in the *left* and *right* Mondrians



constant. This is a common observation: humans are insensitive to large changes in uniform illumination.

As illustrated here, the left-green circle looked dark when it generated the same L cone quanta

catch as the lighter right-red circle. Vision’s spatial image processing rendered the red and green papers with different lightnesses in long-wave light. The lightnesses are stable with large changes in overall illumination.

Retinex Theory,

Fig. 4 Pair of Mondrians in only middle-wave light. Now, the left Mondrian has less illumination than the right. In this wave band the pattern of lightnesses differs from that in long-wave light. That lightness pattern is indifferent to the amount of uniform illumination. Land adjusted the *left side* and *right side* illumination so that the *circles* had equal meter readings

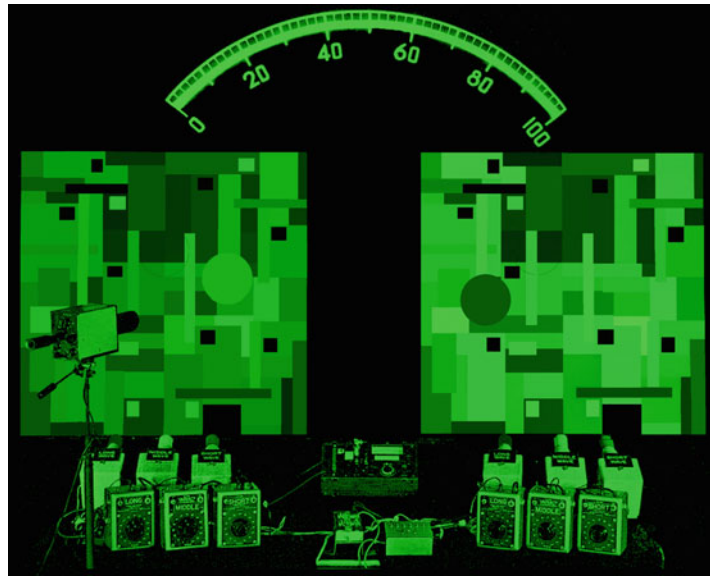


Figure 4 illustrates the Mondrian viewed in only middle-wave light. The left-green circle now looks light and the right-red looks dark. In the experiment they had the same M cone quanta catch. Vision's spatial image processing rendered the red and green papers with different and opposite lightnesses in Fig. 4.

In summary, if the two Mondrians are side by side in the same band of wavelengths, but different overall intensities, the observers report nearly the same set of lightnesses at corresponding locations in the left and right Mondrians. However, one side is detectably lighter than the other. With large uniform changes in illumination, observers report nearly constant lightnesses of the individual papers.

The left-green paper has more long-wave and less middle-wave illumination. The right-red paper has less long-wave and more middle-wave illumination. When adjusted, those adjustments in amount of illumination make the red and green papers have identical radiances. Those adjustments do not significantly alter the lightnesses of the areas in separate illumination. When viewing the Mondrian in combined illumination, in color, those changes in illumination do not change the color appearances of the red and green papers.

These observations led Land to propose the Retinex theory. The triplet of apparent lightnesses,

not cone quanta catches, determines the color appearance. Constant LMS lightnesses generate constant colors. That hypothesis led to a study of color appearances in L, M, S bands of light. Do all red colors have the same triplet of lightness appearances? Does a red color always look [light, dark, dark] in L, M, S light? Does a green always look [dark, light, dark]? Does color appearance always correlate with the triplet of L, M, S lightnesses?

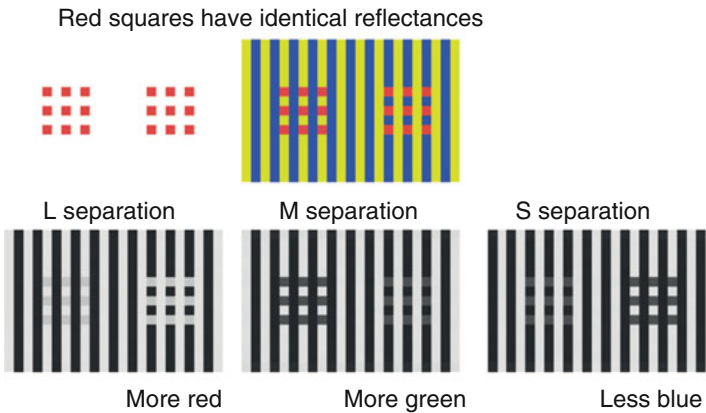
The experiment is easy. Find a red, a green, and a blue filter. Be sure that the filters exclude the other two-thirds of the spectra. With a green filter you should just see greens with different lightnesses. You should not see a mixture of greens and yellows and blues. If you do, you need a filter with a narrower band of transmission.

Identify a group of red objects. Look at them sequentially through the L, M, S filters. Look at them in different ambient illuminations. Look at them at different times of the day. Look at them in sunlight and shadows. Red colors are always (light, dark, dark) in L, M, S light. The same dependence of the triplet of L, M, S lightnesses holds for all colors (Table 1). Lightness is the output of spatial image processing. It is the result of post-receptor spatial processing. That is why lightness does not correlate with cone quanta catch. However, color does correlate with three

Retinex Theory, Table 1 Correlation table of color appearances and the apparent lightnesses in L, M, S illumination

Color Appearance	Appearance in L- light	Appearance in M- light	Appearance in S- light
Red	light	dark	dark
Yellow	light	light	dark
Green	dark	light	dark
Cyan	dark	light	light
Blue	dark	dark	light
Magenta	light	dark	light
White	light	light	light
Black	dark	dark	dark

Retinex Theory, Fig. 5 (Top-left) Color squares and (top-right) color assimilation change their appearance; (bottom) L, M, S separation images. Color appearances of red squares correlate with L, M, S lightnesses



lightnesses in long-, middle-, and short-wave light.

Retinex theory predicts that the triplet of L, M, S lightnesses determines color. Colors are constant with changes in illumination because the triplet of lightnesses is nearly constant.

Land’s observation still stands: The triplet of lightnesses correlates with color. The observation is important because a variety of different phenomena can influence lightness, such as simultaneous contrast, the Cornsweet effect, and assimilation. Regardless of the cause of the lightness changes, when two identical physical objects look different, color appearances correlate with their L, M, S lightnesses ([7], Section E).

In the color assimilation display (Fig. 5), there are two sets of nine red squares that have the same reflectance and appear the same (top left).

However, if these red squares are surrounded by yellow and blue stripes, they look different (top center): the left red squares fall on top of the yellow stripes, and the right ones on the blue stripes. The left squares appear a purple red, while the right ones appear a yellow orange. In other words, the left squares appear more blue and the right ones more yellow.

In the L separation the corresponding squares are lighter on the right of the separation; in the M separation these patches are lighter; in the S separation they are darker on the right. Land’s Retinex predicts that whenever L and M separations are lighter and S separation is darker, then that patch will appear more yellow. Whenever S separation is lighter and L and M separations are darker, then that patch will appear more blue. Colors correlate with L, M, S lightnesses ([7], pp. 221–281).

Retinex Image Processing

Land described that the fundamental challenge of color vision shifted to the ability to predict lightness; that is, the spatial interactions found in post-receptor neural processes. In 1967 Land and McCann proposed a computational model for calculating lightness from the array of all scene radiances [6]. The model compared each pixel with every other pixel in an image. The goal was to calculate the sensation of image segments that equaled what observers saw. In the past 50 years, there have been many implementations and variations of this process. They are called Retinex algorithms. It is curious that Land reserved the use of the term “Retinex” to describe three independent lightness channels. Today’s usage of the word includes a much wider range of computer algorithms that build calculated appearances out of arrays of radiances.

To calculate lightnesses in complex scenes, one must:

- Capture scene radiances.
- Convert scene radiances to cone and rod quanta catches.
- Calculate lightness using all pixels in the scene.
- Compare calculated lightness with observer matches.

The Land and McCann model [6, 9, 11] used:

- Edge ratios
- Gradient threshold (found to be unnecessary in later studies)
- Multiplication of edge ratios (made long-distance interactions)
- Reset to maxima (scaled the output)
(introduced dependence on scene content, e.g., simultaneous contrast)
- Average of many spatial comparisons

The first computer implementation of the model used an array of 20 by 24 pixels. McCann, McKee, and Taylor showed that long-, middle-, and short-wave computed lightnesses predicted

observer matches of color Mondrians in color constancy experiments [8].

Since the late 1960s, computer imaging has shown remarkable advances. Digital images have replaced film in most of photography. Computer graphics has made image synthesis ubiquitous. Retinex image processing has grown with the advances in digital imaging [9–11]. In the early 1980s Frankle and McCann introduced a multi-resolution algorithm that allowed efficient comparison of all pixels in the image [12]. Jobson and Kotera with their colleagues have studied the NASA Retinex. Rizzi and colleagues have developed the Milan Retinex ([7], pp. 324–328). Sobol extended that Retinex algorithm was used in the design of commercial cameras [13]. Other algorithms have used Retinex spatial processing in color gamut-mapping applications [14].

The important feature of real complex scenes is that the illumination is rarely uniform. Shadows and multiple reflections increase the dynamic range of light coming to our eyes and to cameras. The application of Retinex algorithms to high dynamic range (HDR) scenes has become a major topic of research and engineering applications. The limits of HDR scene capture and reproduction are controlled by optics, namely, optical veiling glare. Camera glare limits the range of light on the sensor, just as intraocular glare limits the range of light on the retina. The scene content controls the range of light in images. Vision’s post-receptor neural processes compensate for veiling glare. That explains humans’ high dynamic range of appearances from low-dynamic-range retinal images. The spatial mechanisms modeled by Retinex algorithms play a major role in compensating for glare and generating our range of color and lightness sensations.

Over the years many variations of spatial processing mimicking human vision have been called Retinex algorithms.

Cross-References

- [Color Constancy](#)
- [Glare](#)

References

1. McCann, J.J., Benton, J., McKee, S.: Red/white projections and rod/long-wave cone color: an annotated bibliography. *J. Electron. Imaging* **13**, 8–14 (2004)
2. Land, E.H.: The retinex. *Am. Sci.* **52**, 247–264 (1964)
3. Land, E.H.: The retinex theory of colour vision. *Proc. R. Inst. Gr. Brit.* **47**, 23–58 (1974)
4. Land, E.H.: Smitty Stevens' test of retinex theory. In: Moscovitz, H., et al. (eds.) *Sensation and Measurement, Papers in Honor of S. S. Stevens*, pp. 363–368. Applied Science Publishers, London (1974)
5. Land, E.H.: The retinex theory of color vision. *Sci. Am.* **237**, 108–128 (1977)
6. Land, E.H., McCann, J.J.: Lightness and retinex theory. *J. Opt. Soc. Am.* **61**, 1–11 (1971)
7. McCann, J., Rizzi, A.: *The Art and Science of HDR Imaging*, pp. 221–375. Wiley, Chichester (2012)
8. McCann, J.J., McKee, S., Taylor, T.: Quantitative studies in retinex theory, a comparison between theoretical predictions and observer responses to color Mondrian experiments. *Vision Res.* **16**, 445–458 (1976)
9. McCann J.J.: Lessons learned from Mondrians applied to real images and color gamuts. In: *Proceedings of IS&T/SID Color Imaging Conference*, vol. 7, pp. 1–8. Scottsdale (1999)
10. McCann, J.J.: Simultaneous Contrast and Color Constancy: Signatures of Human Image Processing, Chapter 6 in *Color Perception: Philosophical, Psychological, Artistic, and Computational Perspectives*, Vancouver Studies in Cognitive Studies: Davis, A., Davis, S. (eds.) pp. 87–101. Oxford University Press, Vancouver 2000
11. McCann, J.J.: Capturing a black cat in shade: past and present of retinex color appearance models. *J. Electron. Imaging* **13**, 36–47 (2004)
12. Frankle, J., McCann, J.: Method and apparatus of lightness imaging. US Patent 4,384,336 (1983)
13. Sobol, R.: Improving the retinex algorithm for rendering wide dynamic range photographs. *J. Electron. Imaging* **13**, 65–74 (2004)
14. McCann, J.J.: A spatial color-gamut calculation to optimize color appearance. In: MacDonald, L., Luo, R. (eds.) *Colour Image Science: Exploiting Digital Media*, pp. 213–233. Wiley, Chichester (2002)

Retinogeniculate Pathways

► [Magno-, Parvo-, Koniocellular Pathways](#)

Retinogeniculocortical Pathways

► [Magno-, Parvo-, Koniocellular Pathways](#)

Richter, Manfred

Rolf G. Kuehni
Charlotte, NC, USA

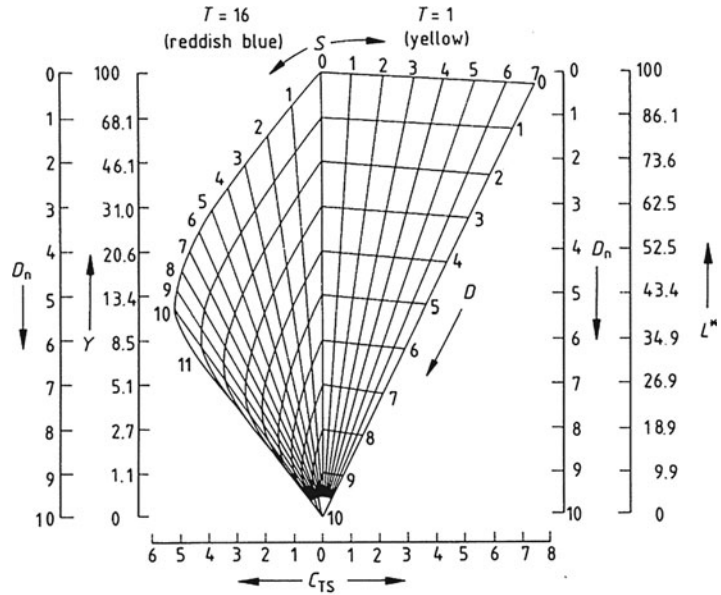


Biography

Richter was born on August 7, 1905, in Dresden, Germany, where he studied technical physics under Robert Luther at the Technical University from 1924 to 1933. The subject of his doctoral dissertation was Goethe's *Farbenlehre* as related to scientific problems. [1] In 1927, as an assistant in the department of color research of the German Institute of Textile Research and following Helmholtz' assistant Arthur König, he developed an international bibliography of publications in color science, an effort he continued until the

Richter, Manfred,

Fig. 1 Schematic representation of the samples of hues T1 and T16 in the S, D diagram (Ref. [3])

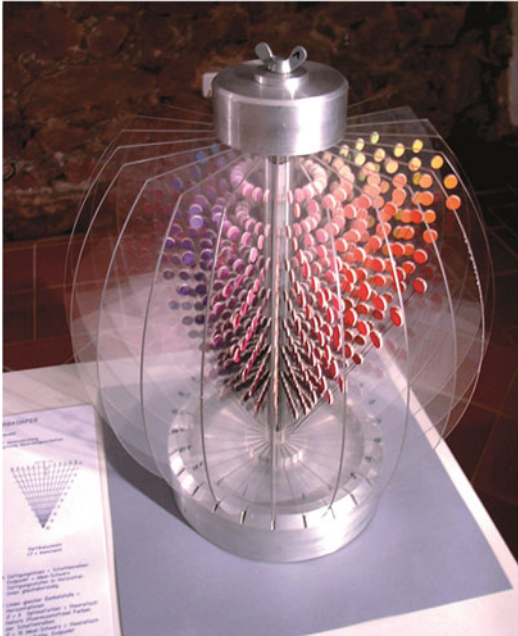


mid-1950s [2]. In 1934, he began work in the laboratories of the lamp manufacturer OSRAM in Berlin. In 1943, he transitioned to the *Materialprüfungsanstalt* (Office for testing of materials), later named *Bundesanstalt für Materialprüfung* (BAM), where he remained until 1962 and where he organized a color research laboratory. In 1941, he was asked by *Deutsches Institut für Normung* (DIN, German institution for industrial standards) to develop a standard color system and atlas, an effort that kept him occupied for an extended period of time. It is known as DIN6164 today. He was also a professor at the *Institut für Lichttechnik* (Institute for lighting technology) of the Berlin Technical University. In 1949, he was a founding member of the *Fachnormenausschuss Farbe* (FNF, Color standards committee). He was a leading member of *Deutsche farbwissenschaftliche Gesellschaft* (DfwG, German society for color science, founded in 1974) and a founding member of the journal *Die Farbe* (1951–2003). He was a member of the directorial board of the International Association of Colour (AIC) and active in several research committees of the International Association of Illumination (CIE). His passing on April 20, 1990, was the result of a traffic accident.

Major Accomplishments/Contributions

DIN6164: Richter's plan for a standard object color system and atlas was that it needed to be based on well-supported colorimetric data and the latest insights into color order. A starting point was Ostwald's color atlas and the Luther-Nyberg optimal object color solid. The system was to be perceptually uniform. As perceptual parameters he selected hue (T), saturation (S), and degree of darkness (D). Lacking a satisfactory colorimetric model of hue scaling he proceeded to experimentally determine a constant saturation contour in the CIE chromaticity diagram separated into 24 perceptually equal hue differences. He then scaled saturation from the neutral point to the spectral limit into up to 16 levels. The darkness scale D is based on logarithmic scaling of the relative brightness value scale proposed in 1928 by S. Rösch. It has a value of 0 for white and 10 for black. A schematic cross section of the system is shown in Fig. 1.

Atlases were published in 1960/62 with matte samples and 1978/83 with glossy samples. Figure 2 shows a 3D model of the system. As its name indicates, DIN6164 is a German industrial standard system.



Richter, Manfred, Fig. 2 3D representation of the samples of the DIN6164 system. © Eckhard Bendin

Publications: In 1940 Richter published a book on color science, with the cooperation of I. Schmidt and A. Dresler, that presented the subject in at that time likely the most comprehensive and detailed form. [4] Given the Second World War it was never translated into English. In 1976, he published *Einführung in die Farbmatrik* (Introduction to color metrics) [5].

References

1. Richter, M.: Das Schrifttum über Goethes Farbenlehre, mit besonderer Berücksichtigung der naturwissenschaftlichen Probleme. Pfau, Berlin (1938)
2. Richter, M.: Internationale Bibliographie der Farbenlehre und ihrer Grenzgebiete, 2 vols, Musterschmidt, Göttingen (1940–1949 and 1950–1954)
3. Richter, M., Witt, K.: The story of the DIN color system. *Color Research and Application* **11**, 138–148 (1986)
4. Richter, M. (with collaboration of I. Schmidt and A. Dresler). *Grundriss der Farbenlehre der Gegenwart* (Basics of the color science of the present). Steinkopff, Dresden (1940)
5. Richter, M.: *Einführung in die Farbmatrik*. de Gruyter, Berlin (1976)

Road Lighting

John D. Bullough

Lighting Research Center, Rensselaer Polytechnic Institute, Troy, NY, USA

Synonyms

Highway lighting; Public lighting; Roadway lighting; Street lighting

Definition

Road lighting is the application of illumination systems along roadways, primarily for the purpose of improving safety by increasing visibility of roadside hazards and by reducing the effects of glare from other light sources in the visual environment, such as vehicle headlamps.

Introduction

Road lighting systems are an important part of the highway safety infrastructure across the world. Two important purposes of road lighting are to allow drivers to see the roadway further ahead than their own vehicle's headlamps allow and to reduce glare from other vehicles' headlamps [1]. Indeed, road lighting is generally associated with reduced nighttime crash rates [2, 3], and the presence of illumination that results in increased luminances of potential road hazards can reduce the impact veiling luminances produced by bright sources of light in the driver's field of view. In the present article, existing practices and standards for road lighting are described, as well as the light source and luminaire technologies used to provide road lighting. For further details on road lighting and interactions with other visual information systems, consult references [4, 5].

Practices and Standards

CIE 115

In most of the world, the primary road lighting standard is the Commission Internationale de

l'Éclairage (CIE) *Lighting of Roads for Motor and Pedestrian Traffic*, CIE 115 [6]. In CIE 115, three main lighting classes are specified: for motorized traffic (M), for conflict areas between vehicles or between vehicles and pedestrians (C), and for pedestrian traffic (P). Each main class is subdivided into numerical classes based upon factors such as traffic volume and speed, road access control, ambient environment, and the mix of motorized and pedestrian traffic. Lower-numbered classes refer to more complex road situations requiring higher light levels.

Specifications for M classes M1 through M6 are given in average road surface luminance and range from 2 to 0.3 cd/m^2 . For C classes C0 through C5, illuminance specifications range from 50 to 7.5 lx. For P classes P1 through P6, illuminance specifications range from 15 to 2 lx. CIE 115 specifications also include limits for characteristics such as uniformity, glare, and, when facial recognition of pedestrians is important, vertical illuminance [6].

CIE 115 also provides a method for the specification of adaptive lighting to account for the possibility of reduced vehicle or pedestrian traffic at certain times throughout the night [6]. Light levels can be reduced to different classes, resulting in decreased energy use and light pollution.

Road lighting calculations are often performed using software that accepts a photometric file containing tabulated luminous intensity data for a luminaire. For a set of road geometric characteristics, pole locations and heights, the software calculates the illuminances or luminances, uniformities, and veiling luminances. Such calculations should account for lamp output reductions and dirt accumulation over time. The light output from a road lighting system could be 20–30 % lower after several years than when it was new and clean.

RP-8

In North America, the primary standard for most road lighting is the Illuminating Engineering Society (IES) *American National Standard Practice for Roadway Lighting*, RP-8 [1]. RP-8 serves as the basis for continuous road lighting system by the American Association of State Highway and

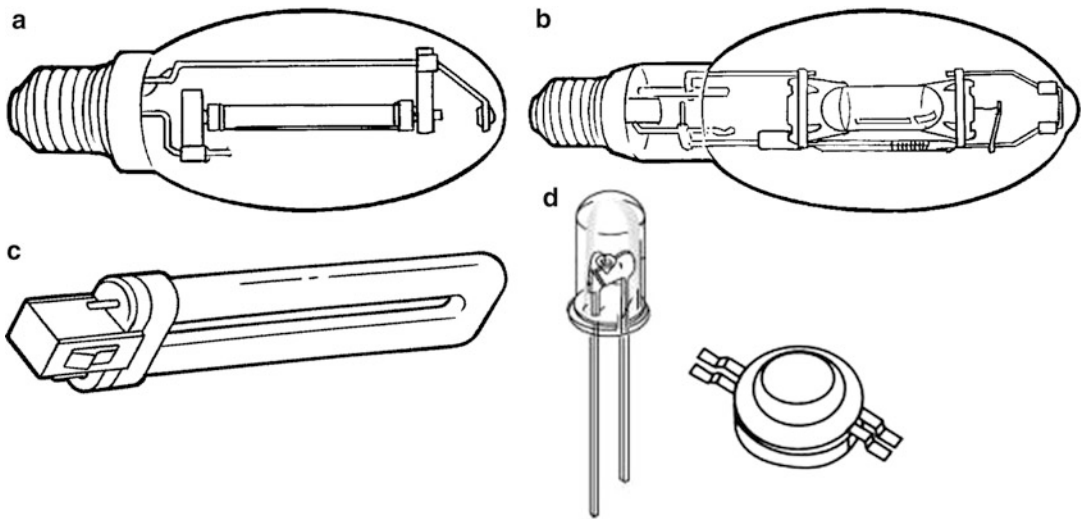
Transportation Officials [AASHTO [7]], which provides warrants for road lighting based on traffic volume and crash frequency at night. RP-8 includes three specification methods based on provision of illuminance, luminance, and small target visibility (STV), the latter of which is rarely used. Each method also includes requirements for uniformity of lighting and limits to control for glare in terms of veiling luminance. In future editions of RP-8, it is planned (as of June 2012) that only the luminance method will be recommended for most road lighting specifications.

The target light levels depend upon the road type (i.e., local, collector, major, or freeway) and the pedestrian conflict level. Specific illuminance criteria exist for different road surfaces, higher for darker than for lighter pavement. (The luminance method uses the pavement type as an input to the calculation.) For example, for a local road with asphalt pavement and a medium pedestrian conflict level, the average illuminance should be at least 7 lx, the minimum no less than 1/6 the average, and the veiling luminance should not exceed 40 % of the average road surface luminance [1]. For a major roadway with the same asphalt and pedestrian conflict level, the average illuminance should be at least 13 lx, the minimum should be no less than 1/3 the average, and the veiling luminance should not exceed 30 % of the average road surface luminance [1]. Calculations for RP-8 road lighting criteria are performed similarly as for CIE 115 criteria.

RP-8 criteria are for continuous lighting along roads excluding intersections and interchanges. Intersection lighting criteria are determined by adding the recommended levels for the intersecting roads.

Light Sources

Road lighting systems can use a wide variety of light sources, including discharge (high- and low-pressure sodium, metal halide, fluorescent, and induction lamps) and solid-state (light-emitting diodes) types [8]; several of these are illustrated in Fig. 1.



Road Lighting, Fig. 1 (a) High-pressure sodium (HPS) lamp (Courtesy of the Lighting Research Center). (b) Metal halide (MH) lamp (Courtesy of the Lighting Research Center). (c) Fluorescent lamp (Courtesy of the

Lighting Research Center). (d) Two light-emitting diode (LED) packages (Courtesy of the Lighting Research Center)

High-Pressure Sodium

High-pressure sodium (HPS) lamps emit light when a current is applied to sodium vapor. The inner envelope of an HPS lamp contains sodium, and the outer glass bulb absorbs ultraviolet (UV) energy and stabilizes the arc tube temperature. HPS lamp efficacies are high, and they have long operating lives and very high lumen maintenance. HPS lamp illumination is yellowish in color. “Whiter” HPS lamps can be developed by increasing the sodium vapor pressure or operating HPS lamps at high frequencies. HPS lamps are the most common sources used in road lighting. For additional details about the performance of HPS lamps, consult the entry “► [High- and Low-Pressure Sodium Lamp](#)” in this Encyclopedia.

Metal Halide

Metal halide (MH) lamps have an arc tube containing various metallic halide compounds in addition to mercury, which emit light across the visible spectrum. MH lamps have relatively high efficacy and are available in a range of white colors ranging from “warm” to “cool” in appearance. Lamp life is typically not as high as HPS lamps but has been improving substantially in recent years. Lumen maintenance values range

from fair to good. MH lamps with ceramic arc tubes and improved starting gear are available, which have much-improved life and lumen maintenance. For additional details about the performance of MH lamps, consult the entry “► [Metal Halide Lamp](#)” in this Encyclopedia.

Low-Pressure Sodium

Like HPS, low-pressure sodium (LPS) lamps use sodium vapor but at a much lower vapor pressure. LPS lamps are linear in shape. They produce nearly monochromatic yellow light near 589 nm. They have relatively long operating lives, excellent lumen maintenance, and very high luminous efficacy, although color rendering is essentially nonexistent with these lamps. LPS lamps are sometimes used near astronomical observatories because it is fairly easy to filter out the wavelengths emitted this source. LPS is more common in Europe than in North America for road lighting. For additional details about the performance of LPS lamps, consult the entry “► [High- and Low-Pressure Sodium Lamp](#)” in this Encyclopedia.

Fluorescent and Induction

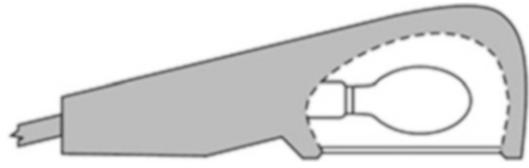
Fluorescent lamps are low-pressure gas discharge sources. Light is produced by fluorescent

phosphors activated by ultraviolet energy from a mercury arc. Fluorescent lamps have reasonably long operating lives, limited primarily by the life of the lamp end electrodes that generate the electrical discharge. Fluorescent lamps require ballasts to provide the starting and operating voltages and currents needed for proper lamp operation. Fluorescent lamp color of fluorescent lamps is determined by the phosphors used to coat the envelope. Fluorescent lamps have relatively higher luminous efficacies and lumen maintenance values.

A special fluorescent lamp type, the induction lamp, is being used increasingly in road lighting. Induction lamps have somewhat more compact shapes than tubular fluorescent lamps because they use a magnetic induction coil to provide the current that stimulates the mercury vapor inside the lamp. Typical operating lives are double or triple those of fluorescent lamps, with luminous efficacy comparable in value. For additional details about the performance of fluorescent and induction lamps, consult the entries “► [Tubular and Compact Fluorescent Lamp](#)” and “► [Induction Lamp](#)” in this Encyclopedia.

Light-Emitting Diodes

Light-emitting diodes (LEDs) are solid-state semiconductor junction devices that emit light when a current is passed through the junction. White light can be created by mixing light from red, green, and blue LEDs, or by using blue LEDs with phosphors that convert some of the blue light into yellow light, resulting in a mixture that is perceived as white. New LED packages exist in addition to 5-mm diameter epoxy capsule LEDs used for indicator lights. Some packages contain metal heat sinks to dissipate internal temperatures, which can reduce LED light output and shorten operating life. LED lighting technology is increasing in luminous efficacy and LEDs also have very long operating lives with gradual reductions in light output over time. For additional details about the performance of LEDs, consult the



Road Lighting, Fig. 2 Cobrahead luminaire for road lighting (Courtesy of the Lighting Research Center)

entry “► [Light-Emitting Diode, LED](#)” in this Encyclopedia.

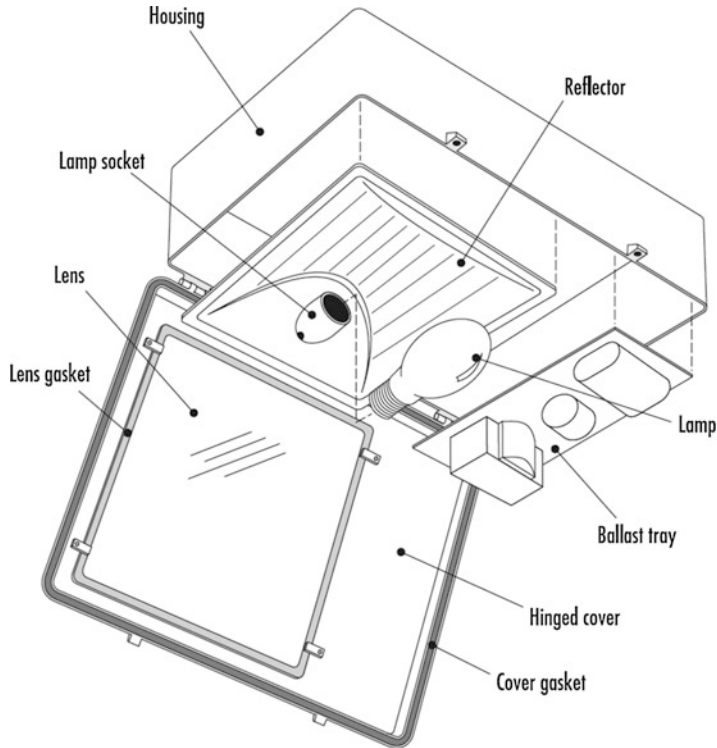
Luminaires

Road lighting systems generally consist of pole-mounted luminaires, usually with individual photocell or timer control. The most common luminaire type used for road lighting is the so-called cobrahead luminaire (Fig. 2), named for its distinctive shape. Figure 3 shows typical road luminaire construction. Most luminaires are designed for discharge lamps such as HPS or MH [9], but there is increasing interest in induction and LED systems [10, 11].

Luminaires for road lighting can be classified in several ways. A cutoff classification system [1], used by many North American road lighting specifiers, classifies luminaires according to their luminous intensity distributions in different angular regions (relative to nadir, defined as 0° directly below the luminaire) where light output could contribute to glare or light pollution. Table 1 summarizes the luminous intensity limits (in cd) for different cutoff classifications, relative to the light output (in lm) of the lamp inside the luminaire. The IES more recently adopted a system that simply uses the light output (in lm) emitted within various angular regions to serve as the basis for classification [12].

As mentioned previously, the most common control system for roadway lighting is a photocell mounted to an individual luminaire, which will switch the luminaire on and off at a specified ambient light level. Some systems will use

Road Lighting,
Fig. 3 Components within
a typical road lighting
luminaire (Courtesy of the
Lighting Research Center)



Road Lighting, Table 1 Light output limits for luminaire cutoff classifications [8], given in luminous intensity values (in cd) as a percentage of the luminaire’s lamp lumens (in lm)

Cutoff type	Light between 80° and 90°	Light above 90°
Full cutoff	100 cd per 1,000 lm	None
Cutoff	100 cd per 1,000 lm	25 cd per 1,000 lm
Semicutoff	200 cd per 1,000 lm	50 cd per 1,000 lm
Noncutoff	No limitation	No limitation

centralized control via a single photocell or time clocks. Technological innovations are making centralized control systems attractive, because these systems can also monitor performance of individual luminaires and alert the system operator when a lamp or ballast failure has occurred or is close to occurring [13]. They can also be useful in controlling adaptive road lighting systems.

Road Lighting, Table 2 Scotopic/photopic (S/P) ratios of common light sources [14, 15]

Light source	S/P ratio
HPS 250 W clear	0.63
MH 400 W clear	1.51
LPS	0.25
Fluorescent/induction (“cool white”)	1.48
LED (4,300 K)	2.04

Color: Mesopic Vision

The color of the light source used for road lighting can influence visual perception of drivers and pedestrians. Although photometric specifications for road lighting are based on the photopic luminous efficiency function, representing the spectral sensitivity of cone visual photoreceptors to light at daytime light levels, visual spectral sensitivity differs at low levels commonly experienced under road lighting at night. At very low light levels, rod photoreceptors, whose sensitivity is

represented by the scotopic luminous efficiency function, dominate vision. At many road lighting levels, both rods and cones contribute to vision and luminous efficiency can be approximated by a combination of photopic and scotopic efficiency [14].

As a consequence, light sources with relatively greater power in the short-wavelength portion of the visible spectrum can be more effective for vision at these so-called mesopic, nighttime light levels. Lamp spectra can be characterized by their scotopic/photopic (S/P) ratio [15], which indicates a light source's ability to stimulate the rods for an equivalent amount of cone stimulation. Table 2 lists S/P ratios for several road lighting sources. The CIE has developed a unified system of photometry [15] bridging the photopic and scotopic systems, as part of CIE Publication 191 [15], to quantify mesopic luminance. Above a luminance of 5 cd/m^2 , the unified luminance and the photopic luminance are equivalent, and below a luminance of 0.005 cd/m^2 , the unified luminance and the scotopic luminance are equivalent. The CIE 191 system uses the photopic luminance and the S/P ratio of the light source under investigation to estimate the unified luminance.

Cross-References

- [Automotive Lighting](#)
- [Glare](#)
- [High- and Low-Pressure Sodium Lamp](#)
- [Induction Lamp](#)
- [Light-Emitting Diode, LED](#)
- [Light Pollution](#)
- [Luminaires](#)
- [Metal Halide Lamp](#)
- [Tubular and Compact Fluorescent Lamp](#)

References

1. Illuminating Engineering Society: American National Standard Practice for Roadway Lighting, RP-8. Illuminating Engineering Society, New York (2000)
2. Commission Internationale de l'Éclairage: Road Lighting as an Accident Countermeasure, CIE 93.

- Commission Internationale de l'Éclairage, Vienna (1992)
3. Bullough, J.D., Rea, M.S.: Intelligent control of roadway lighting to optimize safety benefits per overall costs. Paper presented at the 14th Institute of Electrical and Electronics Engineers conference on intelligent transportation systems, George Washington University, Washington, 5–7 October 2011
4. Boyce, P.R.: *Lighting for Driving*. CRC Press, New York (2009)
5. Bullough, J.D.: *Roadway transportation lighting*. In: Kutz, M. (ed.) *Handbook of Transportation Engineering*, vol. II, 2nd edn. McGraw-Hill, New York (2011)
6. Commission Internationale de l'Éclairage: *Lighting of Roads for Motor and Pedestrian Traffic*, CIE 115. Commission Internationale de l'Éclairage, Vienna (2010)
7. American Association of State Highway and Transportation Officials: *Roadway Lighting Design Guide*, GL-6. American Association of State Highway and Transportation Officials, Washington, DC (2005)
8. Rea, M.S. (ed.): *Illuminating Engineering Society Lighting Handbook: Reference and Application*, 9th edn. Illuminating Engineering Society, New York (2000)
9. McColgan, M., Van Derlofske, J., Bullough, J.D., Vasconez, S.: *Specifier Reports: Parking Lot and Area Luminaires*. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2004)
10. Radetsky, L.: *Specifier Reports: Streetlights for Collector Roads*. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2010)
11. Radetsky, L.: *Specifier Reports: Streetlights for Local Roads*. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2011)
12. Illuminating Engineering Society: *Luminaire Classification System for Outdoor Luminaires*, TM-15. Illuminating Engineering Society, New York (2007)
13. Bullough, J.D.: *Lighting Answers: Dynamic Outdoor Lighting*. National Lighting Product Information Program. Rensselaer Polytechnic Institute, Troy (2010)
14. Rea, M.S., Bullough, J.D., Bierman, A., Freyssinier-Nova, J.P.: A proposed unified system of photometry. *Light. Res. Technol.* **36**(2), 85–111 (2004)
15. Commission Internationale de l'Éclairage: *Recommended System for Mesopic Photometry Based on Visual Performance*, CIE 191. Commission Internationale de l'Éclairage, Vienna (2010)

Roadway Lighting

- [Road Lighting](#)

Runge, Philipp Otto

Rolf G. Kuehni
Charlotte, NC, USA



P. O. Runge, Self-portrait, ca. 1804

Biography

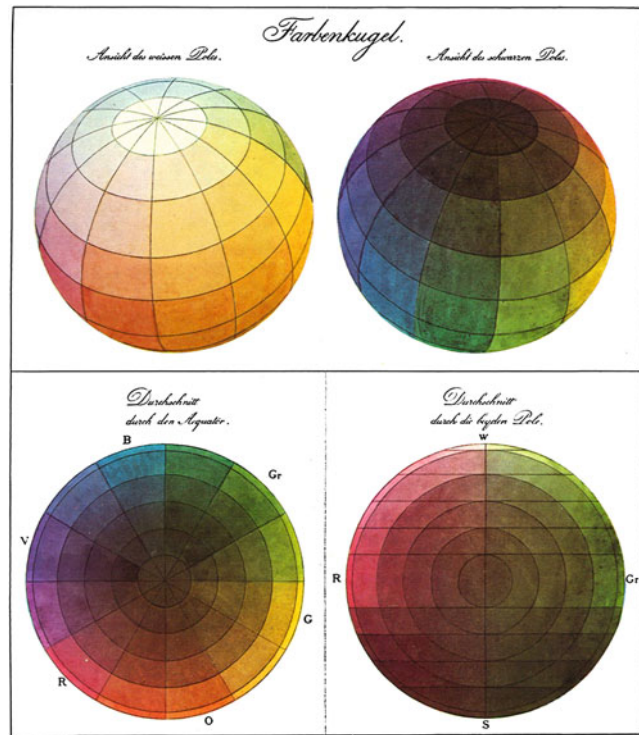
Runge was born on July 23, 1777, the 9th of 11 children of a tradesman and cargo ship owner and his wife in Wolgast, Pomerania, on the Baltic Sea, then under Swedish rule. As a child he was frequently ill with tuberculosis, being often educated at home. In 1795 he began a commercial apprenticeship at his older brother Daniel's firm in Hamburg. In 1799 Daniel supported Otto financially to begin the study of painting at the Copenhagen Academy. In 1801 Otto moved to Dresden to continue his studies, where among others he met his future wife Pauline Bassenge and the painter Caspar David Friedrich. He also began to study extensively the writings of the seventeenth-century mystic Jakob Boehme. In 1803, on a visit to Weimar, Runge unexpectedly met Johann

Wolfgang von Goethe, and the two formed a friendship based on their common interests in color and art. After marrying Pauline in 1804, they moved to Hamburg. But due to imminent war dangers (Napoleonic siege of Hamburg), they relocated in 1805 to his parental home in Wolgast where they remained until 1807, then returned to Hamburg. Together they had four children, the last born on the day after Runge's premature death in 1810. In March of 1810 Runge became ill with tuberculosis again to which he succumbed on December 2 of the year. Runge was of a mystical, deeply Christian turn of mind, and in his artistic work he tried to express notions of the harmony of the universe through symbolism of color, form, and numbers. He considered blue, yellow, and red to be symbolic of the Christian trinity, equating blue with God and the night; red with morning, evening, and Jesus; and yellow with the Holy Spirit (Runge 1841, I, p. 17 [1]). During his short life he became one of the most important German painters of the Romantic period.

Major Accomplishments/Contributions

Runge's interest in color was the natural result of his work as a painter and of having an enquiring mind. Among his accepted tenets was that "as is known, there are only three colors, yellow, red, and blue." (letter to Goethe of July 3, 1806) [2]. His goal was to establish the complete range of colors resulting from mixture of the three primaries, among themselves and together with white and black. In the same lengthy letter, Runge discussed in some detail his views on color order and included a sketch of a mixture circle, with the three primary colors forming an equilateral triangle and, together with their pairwise mixtures, a hexagon. He arrived at the concept of the color sphere sometime in 1807, as indicated in his letter to Goethe of November 21 of that year, by expanding the hue circle into a sphere, with white and black forming the two poles [3]. A color mixture solid of a double-triangular pyramid had been proposed by Tobias Mayer in 1758 and implemented in form of a

Runge, Philipp Otto,
Fig. 1 Hand-illuminated
 illustration from
 P. O. Runge, *Farben-Kugel*,
 Ref. [4]. Top: Image of
 sphere top down (*top left*)
 and bottom up (*top right*).
 Horizontal cross section
 through middle of the
 sphere (*bottom left*) and
 vertical cross section
 (*bottom right*)



triangular pyramid by J. H. Lambert, facts known to Runge. His expansion of that solid into a sphere appears to have had an idealistic basis rather than one of logical necessity. In 1808 he hoped to provide scientific support for the sphere form with disk color mixture experiments. Encouraged by Goethe and other friends, he wrote in 1808 a manuscript describing the color sphere, published in Hamburg early in 1810 as *Farben-Kugel* (Color sphere) by his friend Friederich Perthes [4]. An included hand-colored plate shows two different views of the surface of the sphere as well as horizontal and vertical slices demonstrating the organization of its interior (Fig. 1).

In addition to a description of the color sphere, it contains an illustrated essay on rules of color harmony by Runge and one on color in nature written by Runge's friend Henrik Steffens. Runge's premature death limited the impact of this work. Goethe, who had read the manuscript before publication, mentioned it optimistically in his *Materialien zur Geschichte der Farbenlehre* of 1810 as "successfully concluding this kind of effort" [5]. More detailed similar systems

were published soon thereafter by Gaspard Grégoire in France and by Mathias Klotz in Germany [6, 7].

References

1. Runge, P.O.: *Hinterlassene Schriften*, 2 vols. Hamburg, Perthes 1840/41
2. Letter by Runge to J. W. von Goethe dated July 3, 1806, in Maltzahn, H.: *Philipp Otto Runge's Briefwechsel mit Goethe*, Weimar: Verlag der Goethe-Gesellschaft This letter was included as an appendix in Goethe's *Farbenlehre* of 1810 (1940)
3. Letter by Runge to J. W. von Goethe dated Nov. 21, 1807, in Maltzahn, H., op. cit
4. Runge, P.O.: *Die Farben-Kugel, oder Construction des Verhaeltnisses aller Farben zueinander*, Perthes, Hamburg (1810) (English text translation available on the website www.iscc.org)
5. Goethe, J.W.: *Materialien zur Geschichte der Farbenlehre*. Cotta, Tübingen (1810). Entry on J. H. Lambert
6. Grégoire, G.: *Théorie des couleurs, contenant explication de la table des couleurs*. Brunot-Labbe, Paris (1815)
7. Klotz, M.: *Gründliche Farbenlehre*. Lindauer, München (1816)

Rushton, William A. H.

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

William Albert Hugh Rushton was a British physiologist who made important contributions to our understanding of color vision and perception. He is perhaps best known now for his development of the principle of univariance.

Rushton was born in London on 8 December 1901. He entered Cambridge University as a medical student in 1921 and obtained a degree in physiology in 1925. He received a PhD degree in 1928 working under Prof. E.D. Adrian for his research investigating the flow of current in and around nerves to determine the portion of the current responsible for excitation. As a result of this work he won the Stokes Studentship at Pembroke College in 1929. He spent 2 years at the Johnson Foundation in Philadelphia before returning to a Research Fellowship at Cambridge University. He went to University College Hospital in 1931 to study clinical medicine before obtaining a Lectureship at Cambridge University in 1935 after which he produced a prodigious body of work; he published 37 papers on nerve with 4 colleagues over 25 years and 147 papers on vision with 27 colleagues over a period of 30 years. Upon retirement he spent some time at Florida State University as a Distinguished Research Professor. He was awarded the Royal Medal of the Royal Society in 1970 and continued to publish until he died in June 1980.

Nervous Excitation

Rushton laid the groundwork for the establishment of the modern theory of nervous excitation and propagation by his quantitative analysis of the temporal and spatial factors involved in electrical excitation [1, 2].

Retinal Densitometry

He developed techniques to measure the visual pigments *in vivo*. And made important contributions regarding the distribution of pigments in the retina, the action spectrum of bleaching, and the spectral characteristics of bleaching [3–5].

Cone Pigments

By 1955, Rushton had evidence of visual pigments at the fovea in normal and color-blind subjects. He obtained evidence of two pigments in the medium-long wavelength spectrum in color normals and demonstrated that protonopes and deuteranopes were each lacking one of these pigments [6, 7].

Analytical Anomaloscope

In the early 1960s, Rushton developed an analytic anomaloscope to study color vision. Using this instrument, observers would match a pure spectral light with an additive mixture of red and green in variable proportions [8].

Principle of Univariance

Rushton published two papers in *Scientific American* and gave several notable didactic lectures. In a review lecture at the Physiological Society (London), he described the principle of univariance that: “the output of a receptor depends upon its quantum catch, but not upon which quanta are caught.” Many of the laws of color mixing are a direct consequence of this principle. He also reintroduced the cone pigment triangle, first put forward by Maxwell, as an alternative to the usual representation of color space developed by the CIE [9].

References

1. Rushton, W.A.H.: The time factor in electrical excitation. *Biol. Rev.* **10**, 1–17 (1935)
2. Rushton, W.A.H.: A theory of excitation. *J. Physiol. Lond.* **84**, 42P (1935)

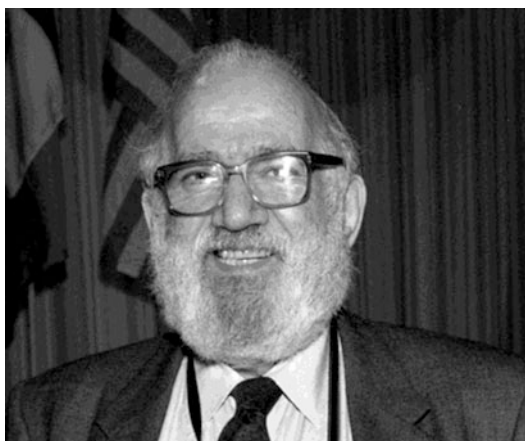
3. Campbell, F.W., Rushton, W.A.H.: The measurement of rhodopsin in the human eye. *J. Physiol. Lond.* **126**, 36P–37P (1954)
4. Campbell, F.W., Rushton, W.A.H.: Measurement of scotopic pigment in the living human eye. *J. Physiol. Lond.* **130**, 131–147 (1955)
5. Rushton, W.A.H.: The difference spectrum and the photosensitivity of rhodopsin in the living human eye. *J. Physiol. Lond.* **134**, 11–29 (1956)
6. Rushton, W.A.H.: A cone pigment in the protonope. *J. Physiol. Lond.* **168**, 345–359 (1963)
7. Rushton, W.A.H.: A foveal pigment in the deuteranope. *J. Physiol. Lond.* **176**, 24–37 (1965)
8. Baker, H.D., Rushton, W.A.H.: An analytical anomaloscope. *J. Physiol. Lond.* **168**, 31P–33P (1963)
9. Rushton, W.A.H.: Pigments and signals in colour vision (Review Lecture). *J. Physiol. Lond.* **220**, 1P–31P (1972)

Further Reading

Barlow, H.B.: William Rushton. 8 December 1901–21 June 1980. *Biograph. Mem. Fellows R. Soc.* **32**, 422–426 (1986)

Saltzman, Max

Ellen C. Carter
Color Research and Application and, Konica
Minolta, Pennsville, NJ, USA



Max Saltzman was a chemist, educator, and scholar. As a steadfast member of the color community in the USA, he often mentored colleagues with thoughtful and generous encouragement, which resulted in him having a profound influence in the field of color science in both industry and the art community.

Saltzman received a B. S. degree in chemistry from the College of the City of New York in 1936. In this early period he worked in the field of medical instrument research and during the war years was in civilian service with the Chemical

Warfare Service. Following the war he settled into his first career in color by joining Harmon Colors, which later became Allied Chemical Corporation, where he worked in various research and management capacities for 26 years. As a native New Yorker, he lived in the city and reverse commuted to New Jersey for this entire time period.

At Harmon he became an expert in pigments, dyes and resins by identifying the colorants, making objective color measurements, defining criteria for selection of colorants for specific applications and use of computers for color matching. We can see his interest and influences through the organizations which he joined: The American Chemical Society, the American Association of Textile Chemists and Colorists, the Inter-Society Color Council, the Society of Plastic Engineers, the Optical Society of America, the Society of Dyers and Colorists (U.K.), the Color Group (Great Britain), the Dry Color Manufacturers' Association, and the Federation of Societies for Coatings Technology.

During this time, Saltzman realized there was a need for academic and industrial training in color technology. He encouraged Dr. Walter Bauer, Dean of the College of Science at Rensselaer Polytechnic Institute to establish a new research laboratory at RPI and to recruit Dr. Fred W. Billmeyer, Jr., to run the undergraduate and graduate program in color science and to expand it to include summer short courses for people from industry. The very popular textbook, *Principles of Color Technology* by Fred W. Billmeyer and Max

Saltzman went through two editions during this period (1966 and 1981) and followed Saltzman's teaching principles of going back to the basics for a thorough understanding of the issue. Saltzman served as an Adjunct Professor for most of the 20 years the Rensselaer Color Laboratory was in existence.

Upon retiring from Allied Chemical in 1973, Saltzman left New York, moving to the west coast where he found time to follow one of his other major interests: the study and identification of ancient dyestuffs in textiles. He began by establishing a color identification laboratory in the Institute of Geophysics and Planetary Physics at the University of California. His research led him to far and wide, for example, he traveled to Peru to collect authentic samples of dyes used in ancient Peruvian textiles. This soon led to his second career in the area of conservation of historical textiles and other objects. He became active in the American Institute for Conservation of Historic and Artistic Works and the International Institute for Conservation of Historic and Artistic Works. In 1984, he was invited to present the George L. Stout Memorial Lecture, a prestigious honor from the American Institute for Conservation of Historic and Artistic Works. When the J. Paul Getty Conservation Institute was established, Saltzman became a consultant there.

Although often working behind the scenes with individuals, Max Saltzman became an internationally recognized authority in the field of color science, receiving the Armin J. Bruning Award (1969) and the Inter-Society Color Council Macbeth Award (1986). Among his many recognitions and awards, the Inter-Society Color Council also honored Max Saltzman's lifetime of contributions by bestowing on him the 2001 Godlove Award. Sadly, he died before the ceremony to receive the award.

Further Reading

1. Billmeyer, F.W., Jr.: Principles of Color Technology, 2 edn. John Wiley & Sons, New York (1966, 1981)
2. English translation of Practical color measurement: a primer for the beginner, a reminder for the expert by Anni Berger Schunn (1994)

3. Saltzman, M.: Color matching via pigment identification. *Dyestuffs* **43**(3), 57–65 (1959)
4. Saltzman, M., Keay, A.M., Christensen, J.: The identification of colorants in ancient textiles. *Dyestuffs* **44**, 241–251 (1963)
5. Saltzman, M., Keay, A.M.: Colorant identification. *J. Paint. Technol.* **39**(509), 360–367 (1967)
6. Saltzman, M.: The Identification of Dyes in Archaeological and Ethnographic Textiles in *Archaeological Chemistry – II*, pp. 172–185. ACS Publications, Washington, DC (1978)
7. Billmeyer Jr., F.W., Kumar, R.: Saltzman m. Identification of organic colorants in art objects by solution spectrophotometry: pigments. *J. Chem. Educ.* **58**(4), 307 (1981)
8. Billmeyer, F.W., Saltzman, M., Kumar, R.: Identification of organic pigments by solution spectrophotometry. *Color. Res. Appl.* **7**, 327–337 (1982)
9. Saltzman, M.: Analysis of dyes in museum textiles or You can't tell a dye by its color. In: McLean, C.C., Connell, P. (eds.) *Textile Conservation Symposium in Honor of Pat Reeves Los Angeles: Los Angeles County Museum of Art*, pp. 27–39 (1986)
10. Saltzman, M.: Identifying dyes in textiles. *Am. Sci.* **80**, 474–481 (1992)

Sampling Problems in Computer Graphics

Kartic Subr

Department of Computer Science, University College London, London, UK

Definition

The term “sampling” is ambiguous since it lends itself to a multitude of interpretations across various fields of study. In computer graphics, sampling problems arise in two distinct contexts. First, continuous functions (signals) are discretized via sampling. For example, a pixelated image represents an underlying continuous image signal sampled at the grid of pixel locations. The general problem is then of reconstructing a high-fidelity approximation of the continuous signal using the discretization. This application of sampling is heavily inspired by the wealth of literature in the digital signal processing community [1]. The second category of sampling problems

in computer graphics applications uses sampled values to estimate general characteristics of the sampled population [2]. For example, in Monte Carlo image synthesis, samples of the radiance incident at each pixel are used to estimate the expected radiance at the pixel, using Monte Carlo integration. This article addresses the relevance and application of each of these classes of sampling problems across various topics in computer graphics.

Two Categories of Sampling Problems

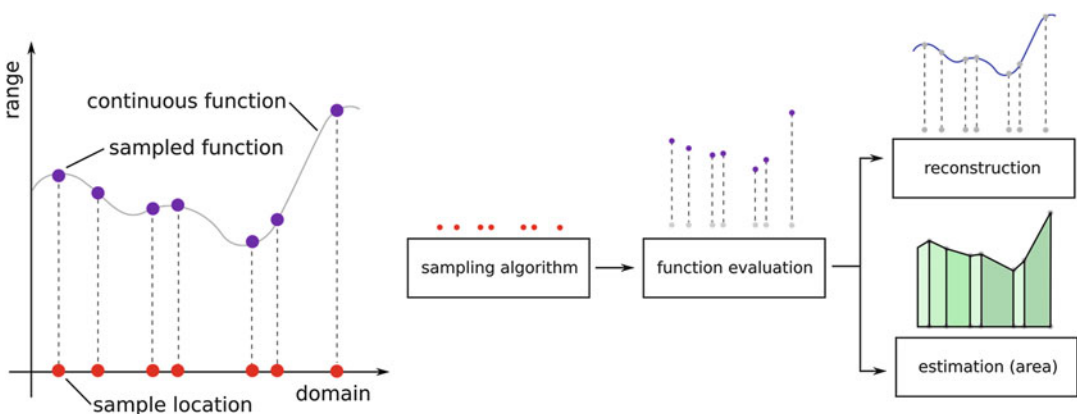
The core problem of sampling a continuous domain involves the selection of a finite set of points in the domain. The sample locations may be chosen by a deterministic algorithm or a stochastic algorithm. The evaluation of a function, defined over this domain, at these sample locations is commonly known as sampling the function. Numerous strategies exist for sampling. The relative benefits of these strategies hinge on the use of the resulting samples.

In computer graphics, the functions are sampled for two purposes (Fig. 1). First, sampled functions are sometimes the only tractable representation for the function under consideration. For example, images are rendered by computing the

amount of light at each pixel on a grid. The second purpose of sampling is to glean key attributes of the function from its sampled values. For example, the expected amount of incident light (radiance) at each pixel is estimated by sampling the infinite space of light paths arriving at that pixel from virtual light sources in the scene. Another example of statistical sampling in graphics is in usability testing, for subjective evaluation, such as rating intuitive interfaces, judging the relative quality of images, etc. This article presents an overview of sampling problems that arise in computer graphics. In addition, it briefly describes the approaches adopted for analysis and assessment of the sampling algorithms. Regardless of the use of the sampled function, the representation of the function by evaluations at a finite set of locations introduces errors. A systematic characterization of these errors is important for assessing and comparing sampling strategies.

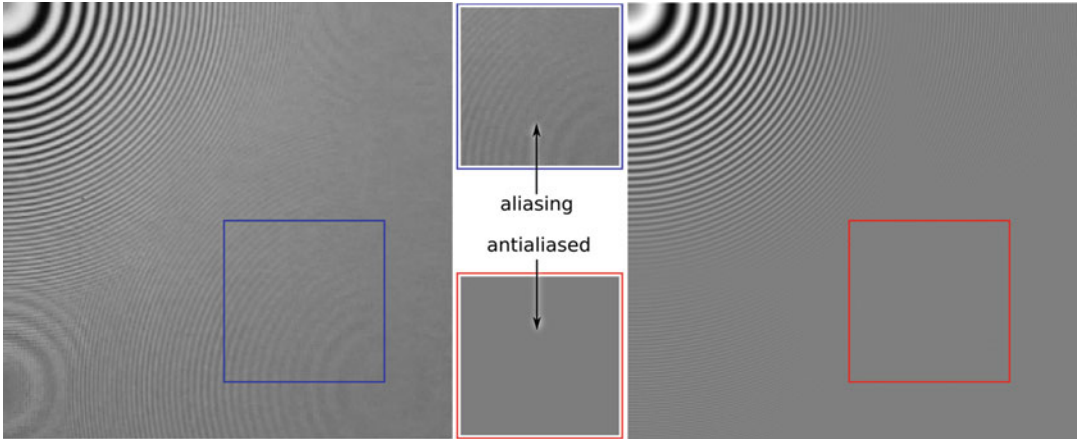
Discretization: Sampling for Reconstruction

Discrete representations are ubiquitous in computer graphics: Images are represented by a spatial grid of locations; the distribution of radiant light energy across wavelengths is often represented



Sampling Problems in Computer Graphics, Fig. 1 Sampling is the process of identifying a set of points (red circles) in the domain of a function. Sampled functions are function values (purple circles) evaluated at the sampled points. In computer graphics, sampled

functions are used for two general purposes: reconstruction of the function (blue curve) and for estimation of characteristics of the function, such as the area under its curve (e.g., the green shaded area)



Sampling Problems in Computer Graphics, Fig. 2 The errors, or artifacts, due to the discretization of a continuous function is called *aliasing*. The zone plate function (over a quadrant) is depicted here, sampled on a

256×256 regular grid. The finite set of samples introduces apparent patterns (*left*) which can sometimes be removed, or reduced, by specially designed reconstruction algorithms (*right*). Zoomed insets are shown in the *middle*

using three pixel colors (red, green, and blue); 3D geometry is commonly expressed as simplicial complexes (e.g., triangle meshes); fluid and other dynamical simulations are performed by propagating physical attributes at discrete locations. A fundamental advantage of discrete representation is that a finite set of data points, or samples, may be used to represent the function under consideration. The process of selection of the locations of these data points in the function domain is called sampling.

The distortion exhibited by the discrete function, due to an insufficient set of data points, is called aliasing. The practical elimination of this error, in general, is still considered an open problem. A thorough discussion of the fundamental problems related to sampling, reconstruction, and aliasing may be found in textbooks on signal processing [1]. The problem of anti-aliasing, or abating the distortion, has enjoyed much attention across several domains of computer graphics.

Images

The choice of a discrete grid, of pixels, to represent an image comes with a penalty – aliasing. If the underlying continuous image exhibits rapid spatial (or temporal) variation, unless the

resolution of the spatial grid (resp. frame rate) is large enough, artifacts appear in the image as shown in Fig. 2. To avoid aliasing, the sampling frequency must be twice the maximum Fourier frequency in the continuous image.

In the frequency domain, if there exists a finite frequency above which the Fourier amplitude spectrum of the image (before sampling) is zero, then aliasing can be avoided by suitably raising the number of samples on the grid. Such signals (images) are known as band limited, and this approach to anti-aliasing is called super-sampling. If the sampling rate is above twice the maximum frequency, then an ideal low-pass reconstruction filter [3] may be designed that is guaranteed to eliminate aliasing. The primary advantage of super-sampling is its simplicity and flexibility. It is widely used in rendering since it can be used in conjunction with complex primitives such as polygons, implicit functions, or parametric surfaces. Common strategies used to generate the sample locations include regular grid sampling, stratified random (jittered) sampling, and sample sets with blue-noise spectral characteristics.

If the underlying continuous signal contains positive energy at arbitrarily high frequencies or

is not band limited, then no finite resolution of the discretization can eliminate aliasing. This is typical, since the existence of any step edge implies arbitrarily high frequencies. In such cases, anti-aliasing may be performed by first blurring, or band limiting, by convolving with a smooth prefilter. Prefiltering is a way of enforcing the image signal before sampling to be band limited. Efficient prefiltered anti-aliasing is tricky for rendering: It involves continuous area integrals with detailed boundary information to find the integration limits; its applicability depends on the primitive. However, it is effectively used in texture mapping, when a high-resolution texture is available, and needs to be mapped on an object that potentially projects onto a small area in the image. Rather than mapping the high-resolution texture and then anti-aliasing at runtime, a popular approach is to query the appropriate level, based on the projected area, from a precomputed, prefiltered hierarchy of the texture. Since anti-aliasing is precomputed with this mip-mapping (multum-in-parvo-mapping) approach, it is widely used in real-time computer graphics applications such as rendering in video games.

Geometry

Polygon meshes are a popular choice for representing 3D geometry in computer graphics. They are simple and intuitive to manipulate locally, and they lend themselves to efficient rendering. Modeling or creating high-quality polygon meshes requires skilled artistry and sophisticated 3D modeling tools. To circumvent the long modeling process, artists often use 3D scans of objects as a starting point to their design. However, scanners today output point clouds as representations of underlying surfaces. The simplification and resembling of these point clouds are considered essential steps before the scanned model is usable for rendering or surface reconstruction.

Similar to the reconstruction of anti-aliased images, reconstruction of the underlying geometry from its point-sampled version is a challenging

problem. The difficulty is exacerbated by the need to generalize elegant sampling theories that are applicable in Euclidean domains, such as for image processing, to arbitrary manifolds. Spectral analysis of manifolds is based on the eigenfunctions of the Laplace–Beltrami operator, just as Fourier analysis is based on trigonometric harmonics. The extension of signal processing on arbitrary manifolds has inspired many useful algorithms for resampling point-sampled geometry, smoothing manifolds, reconstructing simplicial manifolds from point samples, etc. A comprehensive description of sampling and processing of polygonal meshes can be found in [2].

Plenoptic Functions

While rendering images with effects such as defocus (due to a finite camera aperture) and motion blur (due to a finite exposure), the radiance at each pixel is obtained by integrating estimates of incident radiance over the aperture and time. That is, at each pixel (x, y) on the image plane, radiance estimates along a direction (u, v) , so that the ray from the pixel passes through the aperture, and at times (t) are integrated.

The general rendering process involves reconstructing portions of the plenoptic function by estimating radiance at 5D sampled locations (x, y, u, v, t) . The radiance estimate at each sampled location is obtained via numerical integration in a high-dimensional space, accounting for different transport phenomena such as reflection, refraction, scattering in participating media, and occlusion. Hence, the cost of sampling the plenoptic function is high. Although the domain spans a higher-dimensional space, standard signal processing results are still directly applicable. The search for efficient reconstruction algorithms for the portions of the plenoptic function required for rendering has spawned a recent flurry of research. A recent trend in high-quality rendering is to develop ways of approximating frequency domain characteristics of the plenoptic function, using which adaptive sampling and reconstruction filters may be derived.

Physically Based Simulation

The goal of physically based simulation, in computer graphics, is to animate virtual objects by mimicking their physical interactions with the environment. The simulation may be of rigid bodies, deformable objects, or fluids. Although the discretization plays a critical role in deciding the quality – tractability, stability, or fidelity (realism) – of the simulation in all three cases, it is particularly important in fluid simulations.

The introduction of a non-penetration constraint in rigid body simulation introduces discontinuities in attributes such as position and velocities at the moment of collision. Thus, accurately identifying the moment of collision is important. Many methods have been proposed to perform collision detection, and they resort to different discretization [4] of the geometry and time domain.

Deformable objects are simulated using a number of techniques: mass-spring systems, the finite difference method, the finite element method, mesh-free methods, coupled particle systems, and reduced deformable models based on modal analysis. Each of these methods relies on a discretization of the geometry and time, and the choice of discretization has a significant effect on the stability, sensitivity, and fidelity (realism) of the simulation. The list of sampling schemes used in this context includes a cubic octree hierarchy, non-nested hierarchy of meshes (FEM), regular spatial grids (FDM), uniformly distributed samples, and samples with blue-noise spectral profiles. The interested reader is referred to [5, 6] for a survey of the methods.

The central problem in fluid simulation for computer graphics applications is the robust solution of the incompressible Navier–Stokes equations [7]. One of three strategies is popularly adopted to decide the locations where physical attributes such as pressure, viscosity, density, and velocity are iteratively computed at each time step. The first option is a regular grid over the spatial extent of the fluid and is an Eulerian approach to the problem. The second option is to represent the fluid using discrete particles that

measure these attributes as the particles themselves are advected about. The third option is to use a combination of irregularly spaced samples and regular grid. As with reconstructing images and geometry, the sampling rate governs the largest frequency variation that can be faithfully reconstructed. Simply increasing the number of samples in a fluid simulation quickly results in prohibitively expensive computation times, and so adaptive sampling strategies are key to practicable fluid simulation. Octree-based sampling, blue-noise sampling, and regular (grid) sampling are commonly used.

Statistical Sampling: Sampling for Estimation

Another application of sampling, in computer graphics, is to estimate the statistics of a function under consideration from values of the function at sampled locations. An estimator is a combination of the sampling strategy and the process of combining the samples. The fidelity of the mean and the variance of estimates produced by an estimator are both typical measures of quality for estimators. For example, a Monte Carlo estimator for the expected value of a function is the average of the function evaluated (sampled) at uniformly random locations in the domain. This simple estimator converges to the correct answer, in the limit of sampling the domain.

Physically based image synthesis involves estimation of a portion of the plenoptic function, or incident radiance at each (x, y, u, v, t) , where (x, y) is on the image plane. The problem is addressed by sampling at two levels: designing a high-level rendering strategy or estimator and then choosing low-level, specific sampling strategies.

The rendering strategy governs how light paths are sampled. For example, in path tracing, each estimate of the radiance is obtained by sampling the contribution of a single path from the space of light paths originating at (x, y) on the image plane and terminating at a light. In

distributed ray tracing, incident radiance is recursively estimated at each bounce, using multiple incident samples. Metropolis light transport adopts a Markov-chain model to sample in path space, given a pool of initial paths using which the search is seeded. The specific sampling strategy defines how samples are generated. For example, at each bounce (whether path tracing or distributed ray tracing), the outgoing directions may be generated using quasi-Monte Carlo, random, stratified, or importance sampling schemes [8].

Error and Assessment

The graphics literature is rich with numerous sampling algorithms. Ideally, the invention of a core sampling algorithm develops in four stages: motivation for a novel algorithm, the design of the sampling procedure or algorithm, analysis of the proposed algorithm, and finally assessment of the resulting samples. This section describes typical approaches for the latter two stages – and assessment. The motivation and design of specific sampling algorithms are closely tied to the specific application and beyond the scope of this article.

Sampling for Reconstruction

The analysis of sampling algorithms (for reconstruction) is typically performed in the frequency domain. Fourier analysis is typical for Euclidean domains such as images and 4D plenoptic functions. For frequency analysis over spherical domains and spaces of rotations, spherical harmonics are the popular choice in computer graphics. Using the eigenfunctions of the Laplace–Beltrami operator as a frequency basis is gaining popularity for frequency analysis on arbitrary manifolds. Regardless of the choice of frequency bases, the sampling theorem from signal processing is a fundamental test to analyze the sufficiency of sampling. However, most frequency domain analyses are global. For example, it is impossible to use the Fourier transform of an image to predict local regions with high (or low)

variation. A typical trick for improving local analysis is to use windowed transforms. However, any gain in localization due to windowing is at the cost of precision in the frequency domain, in accordance with the uncertainty principle.

The assessment of reconstruction error is domain specific. Generic measures such as root mean squared error or peak signal to noise ratio, although common, are typically insufficient. For example, the “quality” of reconstructed images relies heavily on human perception. Although many studies have been performed in this area, automatic assessment of reconstructed images, including perceptual considerations, is still an open problem. For 3D geometry, the problem of automatic assessment of final appearance is compounded by its dependence on view and shading.

Sampling for Estimation

The typical measures for the assessment of estimators are accuracy (bias) and precision (variance). Bias measures the expected discrepancy between the estimated value and the actual value. Variance measures the spread of predicted values. Bias and variance are difficult to attain simultaneously. Estimators typically strike a trade-off between the two, and their relative importance depends on the application.

Despite the quantifiable measures of bias and variance, automatic assessment of estimators is not prevalent in image synthesis. Bias leads to physically inaccurate images while variance manifests as noise across the pixels. A large bias is generally believed to be less objectionable (perceptually) than a large variance. Variance reduction schemes such as stratified sampling, importance sampling, control variates, and anti-thetic variates are commonly used in Monte Carlo rendering. The resulting estimators are typically compared by manual inspection of their resulting images given a particular virtual scene. Images are generated with increasing numbers of samples and the estimator with the least noise is considered best. In addition, a converged image using an unknown estimator is visually compared against

a reference solution (with a known estimator and a large number of samples). Clearly, such comparisons are not general and only valid on the particular images that the methods are demonstrated on. The development of an automatic method for such assessment is an open problem.

Cross-References

- [Global Illumination](#)
- [Image Quality](#)

References

1. Oppenheim, A.V., Schafer, R.W.: Discrete-Time Signal Processing (Chap 4, ed. 3), (3rd edition) Prentice-Hall Signal Processing Series, Pearson (2009)
2. Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., Levy, B.: Polygon Mesh Processing (2011)
3. Thompson, S.K.: Sampling. Wiley, Hoboken (2012)
4. Witkin, A., Baraff, D.: Physically based modeling: principles and practice. ACM SIGGRAPH course. <http://www.cs.cmu.edu/~baraff/sigcourse/> (1997). Accessed 14 Feb 2013
5. Sifakis, E.D., Barbic, J.: FEM simulation of 3D deformable solids: a practitioner's guide to theory, discretization and model reduction (part 1). ACM SIGGRAPH course. <http://run.usc.edu/femdefo/sifakis-courseNotes-TheoryAndDiscretization.pdf> (2012). Accessed 14 Feb 2013
6. Nealen A, Müller M, Keiser R, Boxerman E, Carlson M, Physically based deformable models in computer graphics. Computer Graphics Forum 2006;25 (4):809–836
7. Bridson, R.: Fluid Simulation for Computer Graphics. A. K. Peters, Wellesly (2008)
8. Shirley, P., Marschner, S.: Fundamentals of Computer Graphics (Chap. 24). CRC Press, Boca Raton (2009)

Sapir-Whorf Hypothesis

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

Schiller

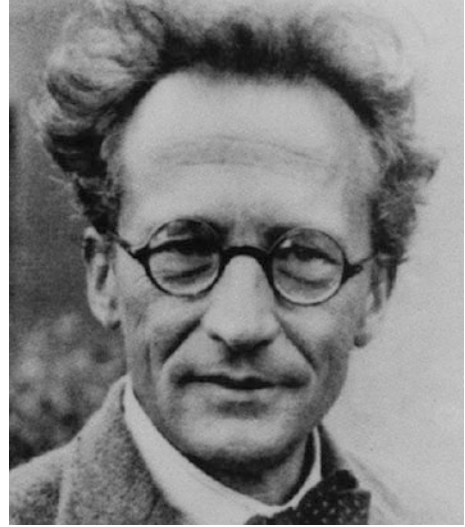
- [Iridescence \(Goniochromism\)](#)

Schrödinger, Erwin

Michael H. Brill¹ and Rolf G. Kuehni²

¹Datacolor, Lawrenceville, NJ, USA

²Charlotte, NC, USA



Erwin Schrödinger (1887–1961) was born in Erdberg, Austria, to a father of Austrian and a mother of mixed Austrian-English descent. He studied physics in Vienna under Franz Exner and became his assistant in 1911. He was influenced early by the writings of the German philosopher Arthur Schopenhauer resulting, among other things, in his interest in color theory. Three of his papers on color were written in Vienna after he concluded his military service in 1918 and before assuming a position at the University of Zürich in 1921. There he published six more papers on color and vision. His interest in color, which did not extend beyond his years in Zürich, apparently was due to his deep involvement with the works of Schopenhauer who in turn was influenced by Goethe.

In Zurich, Schrödinger did his most important work, on quantum wave mechanics, for which he shared with Paul Dirac the 1933 Nobel Prize in physics. In 1940, Schrödinger moved to Ireland, where he remained until his retirement in 1955. Then he returned to Vienna. Among his

Schrödinger, Erwin,
Fig. 1 Schrödinger's
 portrait on the Austrian
 1000-Schilling note



achievements is the thought experiment known as Schrödinger's Cat.

Schrödinger's contributions to color include the following:

- (a) He offered the first mathematical proof [1] of the theorem sketched by Wilhelm Ostwald that, under any illuminant, the object-color tristimulus locus is contained by a two-dimensional manifold generated by reflectances that are 1 or 0 at each wavelength and with at most two transitions between 0 and 1. Robert Luther later called these reflectances "optimal colors." The maximality of two transitions depends on the convexity of the spectrum locus, as is clear in his very brief proof: *"If a pigment has three transition points that in the chromaticity diagram are not located on a straight line, by moving the transition points its reflectance can be changed in a manner that results in a lighter pigment of the same chromaticity coordinates. Therefore, it cannot be optimal."*
- (b) He was the first to use differential geometry for color space, to import brightness as one of the space's coordinates, and to infer total color difference (in units of just-noticeable differences) through arclength along geodesics [2, 3].
- (c) He was the first to offer a mathematically detailed connection of the Young-Helmholtz three-color theory and Hering's opponent-color theory [4], after Helmholtz offered such a concept in 1896 and von Kries that of a zone theory in 1905.

In addition to this work, Schrödinger also published several other works relating to color science [5–9].

Besides Isaac Newton, we are unaware of any color scientist other than Schrödinger whose portrait appeared on paper money (Fig. 1). (Goethe satirized paper money [*Faust*, Part 2, Act I, Scene 4], but did not appear on it.).

References

1. Schrödinger, E.: Theorie der Pigmente von größter Leuchtkraft. *Ann. Phys.* **62**(4), 603–622 (1920) (translation, Theory of pigments of greatest lightness, is at <http://www.iscc.org/resources/translations.php>)
2. Schrödinger, E.: Grundlinien einer Theorie der Farbenmetrik im Tagessehen. *Ann. Phys.* **63**(4), 397–426; 427–456; 481–520 (1920) (translation, Outline of a theory of color measurement for daylight vision, is in D MacAdam, Sources of Color Science, MIT Press, 1970)
3. Schrödinger, E.: Die Gesichtsempfindungen, Müller-Pouillet's Lehrbuch der Physik 2/1, 11. Auflage, 456–560, Braunschweig: Vieweg (1926). (11 pages translated, Thresholds of color difference, in D. MacAdam, Sources of Color Science, but chapter title and length of chapter not indicated)
4. Schrödinger, E.: Über das Verhältnis der Vierfarben- zur Dreifarben-theorie, Sitzungsberichte der Akademie der Wissenschaften in Wien. Mathematisch-naturwissenschaftliche Klasse, Abteilung 2a. **134**, 471–490 (1925) (translation, On the relationship of four-color theory to three-color theory, is in Color Res. Appl. **19**, 37–47 (1994))
5. Farbenmetrik. *Z. Phys.* **1**, 459–466 (1920) (no known translation)
6. Schrödinger, E.: Ton und Farbe, Neue Zürcher Zeitung, 3. Februar, (1923) (In a Swiss newspaper)

7. Schrödinger, E.: Über den Ursprung der Empfindlichkeitskurven des Auges. *Naturwissenschaften* **12**, 925–929 (1924) (no known translation)
8. Schrödinger, E.: Über Farbenmessung. *Phys. Z.* **26**, 349–352 (1925) (no known translation)
9. Schrödinger, E.: Über die subjektiven Sternfarben und die Qualität der Dämmerungsempfindung. *Naturwissenschaften* **13**, 373–376 (1925) (not aware of translation)

Scintillating Scotoma

Muneto Tatsumoto

Department of Neurology, Dokkyo Medical University, Tochigi, Japan

Synonyms

[Aura](#)

Definition

Migraine is one of the most common neurological disorders and is observed in about 11 % of the world population [1]. Migraine attacks are characterized by unilateral, throbbing and moderate-to-severe headache lasting 4–72 h, typically accompanied by nausea and photo- and phonophobia [2]. In approximately 20 % of patients, the migraine can also be preceded by transient neurological symptoms (aura) that usually develop gradually over >5 min and last for <60 min. Auras are most frequently visual (>90 %), but may also involve other sensory symptoms (pins and needles, numbness), hemiparesis or speech deficits. Visual symptom (aura) often has a bimodal progression with positive symptom (**scintillating**) followed by negative symptom (**scotoma**). To date, the trigger mechanisms of migraine with aura are cortical spreading depression (CSD). Migraine with aura (MWA) is diagnosed according to the operational diagnostic criteria of The International Classification of Headache Disorders 3rd edition (beta version) [2]. Although, in migraine with aura,

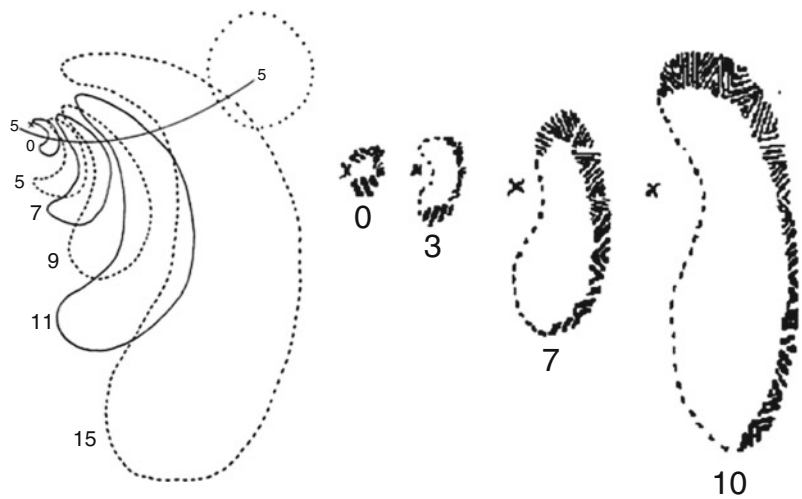
headache which fulfills the same features as migraine without aura (MWOA) frequently occurs following the aura, there are also cases in which, along with aging, there is no headache that fulfills the features of MWOA or in which there is only the aura but no headache [3].

Overview

The patterns of the visual aura are diverse, but scintillating scotoma is representative. Scintillation is a positive symptom in which light is seen despite the lack of any input visual stimulus, and scotoma is the negative symptom in which an input visual stimulus, in the same locus as the scintillation after it has disappeared, is not visible for some time. The phenomenon of the extension of these positive (scintillation) and negative (scotoma) symptoms within the visual field is presumed to be in correspondence with changes in neural activity in the occipital lobe. It is understood that these changes in neural activity are a phenomenon known as cortical spreading depression (CSD). CSD is a wavelike, self-propagating phenomenon of neuronal depolarization, which originates in the cerebral cortex (2–6 mm/min), as well as of inhibition of neuronal activity.

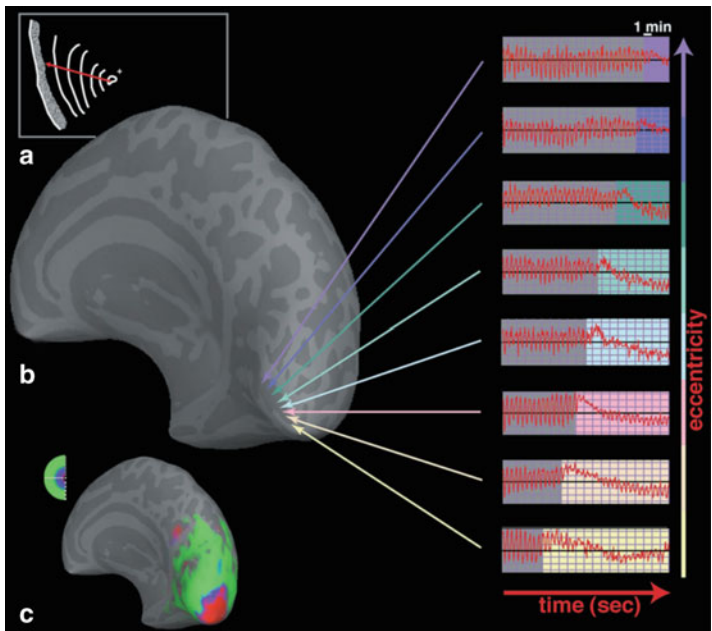
Lashley made a detailed study of his own scintillating scotoma, calculated the speed of migration of the scintillating scotoma in the visual field, and showed that it migrates at approximately 3 mm/min in the visual cortex of the occipital lobe [4] (Fig. 1).

With CSD in an experimental rabbit model, a transient inhibition of neural activity is observed upon electrical stimulation of the cerebrum, which recovers to the state of normal spontaneous activity after a certain period of time [5]. Also, when records were made by an array of recording electrodes on the cerebral surface of a rabbit, a phenomenon was observed in which transient inhibition of cerebral activity was sequentially propagated after electrical stimulation. The speed of propagation of CSD in the experimental rabbit model and the speed of propagation of the scintillating scotoma reported by Lashley were similar, both having been found to be approximately



Scintillating Scotoma, Fig. 1 The analytical drawings of the spatiotemporal evolution of spreading migraine hal-lucinations. *Left:* a scintillating scotoma, spreading at the intervals shown. The blindspot is shown as the *dotted circle*, and the fixation point as *x*; alternate outlines are *solid* and *dotted* to avoid confusion; scintillations were

confined to the region above the line s-s. *Right:* successive forms of a scintillating scotoma, mapped at different time intervals as indicated; the fixation point is shown as *x* (From Lashley [4] . Copyright © 1941 American Medical Association. All rights reserved)



Scintillating Scotoma, Fig. 2 Spreading suppression of cortical activity during migraine aura, as shown by functional MRI. **(a)** A drawing of the progression over 20 min of the scintillating field defect, as described by the patient. **(b)** The spread of the signal from the posterior pole towards more anterior regions representing the peripheral visual field; MR perturbations develop earlier in the foveal regions than in the periphery, consistent with the

progression of the aura from the center to the periphery. **(c)** The MR maps of retinotopic eccentricity acquired during interictal scans. As shown in the logo in the *upper left*, voxels that show retinotopically specific activation in the fovea are coded in *red*. Parafoveal eccentricities are shown in *blue*, and more peripheral eccentricities are shown in *green* (From Hadjikhani et al. [9] Copyright 2001 National Academy of Sciences, U.S.A.)

3 mm/min. The electrophysiological characteristics of CSD which have been indicated include, (1) CSD occurs under specific conditions, (2) its extension is autonomous, (3) the speed of extension is 3 mm/min, (4) inhibition of each locus is sustained for 1–3 min, with a total refractory period longer than 1 min and (5) all intrinsic/induced neural activities are inhibited [6].

The first evidence for the CSD hypothesis of migraine in the human cerebrum was a study conducted of regional cerebral blood flow (rCBF) in the aura phase using Xe-CT, in which spreading oligemia originating in the occipital lobe, similar to CSD, was observed [7]. The oligemia preceding hyperemia is similar to the hemodynamics, which is observed in experimental CSD. Woods, et al., observed, by positron emission tomography of cases in which migraine arose accompanied by abnormal vision, spreading hypoperfusion in which, together with the onset of an episode, an area of reduced rCBF gradually extended forward from the bilateral occipital lobe [8]. The best-known report related to this is that of Hadjikhani, et al. [9]. In one of the three patients it was possible to start the test before the onset of aura since an episode was induced with near certainty upon engaging in basketball practice. Upon functional MRI with presentation of visual stimulus, the signal in the visual cortex became unresponsive to stimuli as the aura began and the baseline signal slowly increased at the same time. This originated in the V3a area, extended with a speed of 3.5 mm per minute to the adjacent occipital cortex including the primary and secondary visual cortices, and also coincided with the areas of the visual aura in the visual field. Subsequently, the cerebral response to visual stimuli remained weak over more than 15 min and the baseline diminished gradually to become lower than before the episode (Fig. 2).

The transient rise in this signal was similar to the change in blood flow, which was observed in accompaniment to CSD in the experimental model. This report was evidence in support of the view that the phenomenon of CSD is involved in the pathology of scintillating scotoma in migraine.

Cross-References

- [Glare](#)
- [Lighting controls](#)

References

1. Robbins, M.S., Lipton, R.B.: The epidemiology of primary headache disorders. *Semin. Neurol.* **30**, 107–119 (2010)
2. Society, H.C.C.o.t.I.H.: The international classification of headache disorders, 3rd edition (beta version). *Cephalalgia* **33**, 629–808 (2013)
3. Aiba, S., Tatsumoto, M., Saisu, A., Iwanami, H., Chiba, K., Senoo, T., Hirata, K.: Prevalence of typical migraine aura without headache in Japanese ophthalmology Clinics. *Cephalalgia* **30**, 962–967 (2010)
4. Lashley, K.S.: Patterns of cerebral integration indicated by the scotomas of migraine. *Arch. Neurol. Psychiatry* **46**, 331–339 (1941)
5. Leão, A.A.P.: Spreading depression of activity in the cerebral cortex. *J. Neurophysiol.* **7**, 359–390 (1944)
6. Somjen, G.G.: Mechanisms of spreading depression and hypoxic spreading depression-like depolarization. *Physiol. Rev.* **81**, 1065–1096 (2001)
7. Olesen, J., Larsen, B., Lauritzen, M.: Focal hyperemia followed by spreading oligemia and impaired activation of rCBF in classic migraine. *Ann. Neurol.* **9**, 344–352 (1981)
8. Woods, R.P., Iacoboni, M., Mazziotto, J.C.: Bilateral spreading cerebral hypoperfusion during spontaneous migraine headache. *N. Engl. J. Med.* **22**, 1689–1692 (1994)
9. Hadjikhani, N., Sanchez del Rio, M., Wu, O., Schwartz, D., Bakker, D., Fischl, B., et al.: Mechanisms of migraine aura revealed by functional MRI in human visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4687–4692 (2001)

Screening Plates

- [Pseudoisochromatic Plates](#)

Simple Colors

- [Primary Colors](#)

Simulation

- [Color Category Learning in Naming-Game Simulations](#)

Simulation and Image Synthesis of Natural Phenomena

► Rendering Natural Phenomena

Simultaneous color contrast

Alessandro Soranzo
Sheffield Hallam University, Sheffield, UK

Definition

Simultaneous color contrast is the condition whereby two surfaces with the same spectral composition are perceived to have a different color when they are placed against different chromatic backgrounds (see Fig. 1).

History and Theories

The term “simultaneous” was introduced by Chevreul to “distinguish this phenomenon to the ‘successive’ contrast, where two colors appear in succession upon the same retinal area” [1, p. 264].

The term “contrast” refers to the fact that the perceived color of the surfaces is “contrasted” by the color of the surround. As the term “contrast” is also used in the literature to indicate the relative

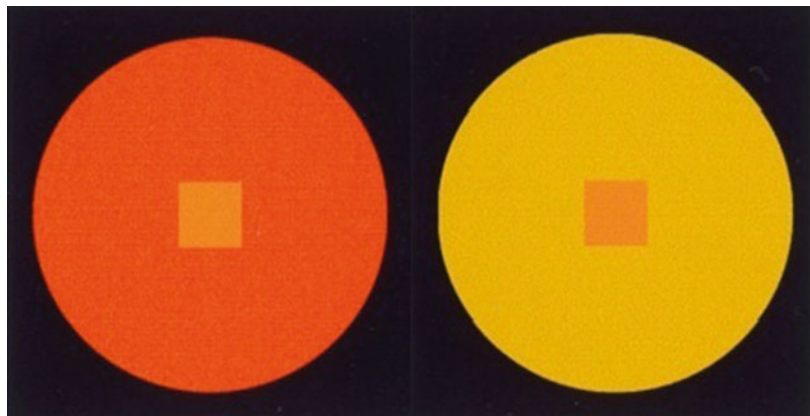
intensity of the stimulation, some authors prefer the term “induction” over contrast.

When the squares and the backgrounds are achromatic, this phenomenon is named Simultaneous Brightness (or Lightness) Contrast effect (see Fig. 2).

This is one of the most studied phenomena in visual perception and has been the focus of centuries of debate that has interested scientists and philosophers since Aristotle’s time [2]. In the nineteenth century controversy raged between Hering [3], who supported an explanation based on retinal neuron interaction processes, and Helmholtz [1], who was in favor of an explanation based on higher-level processes, involving assumptions about the configuration as a whole.

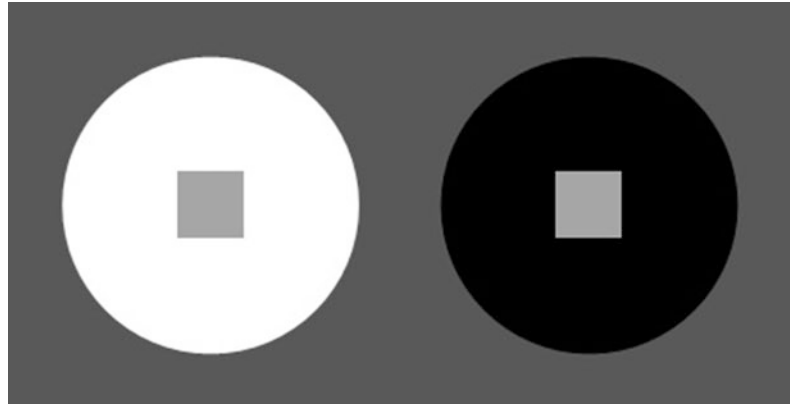
The retinal-based interpretation (also referred to as the “low-level” interpretation) was particularly in vogue during the 1960s mainly because of the physiological discovery of the lateral inhibition process in the limulus retina [4]. To explain the contrast phenomenon the retinal-based interpretation focuses on the notion that the receptors that are stimulated by the background then send inhibition to the receptors that are stimulated by the surrounded area. According to this view, the simultaneous color contrast shown in Fig. 1 occurs because the receptors stimulated by the red background inhibit the red sensitive receptors stimulated by the orange square, leading to a yellowish appearance. Conversely, the receptors stimulated by the yellow background inhibit the yellow sensitive receptors stimulated by the orange square, leading to a reddish appearance.

Simultaneous color contrast, Fig. 1 The spectral components of the *squares* in the centers of the *red* and *yellow* surrounds are identical. The appearance of the color is different, however. The *two squares* appear to acquire the opposite spectral component of their surrounds



Simultaneous color

contrast, Fig. 2 The *grey squares* in the centers of the *white* and *black* surrounds are the same. The appearance is different, however: the *square* to the *left* appears darker than the *square* to the *right*



A similar explanation is provided for the contrast phenomenon seen in Fig. 2. The receptors stimulated by the light background send inhibition to the receptors stimulated by the patch that the background surrounds, causing perceptual darkening. On the other hand, the receptors stimulated by the dark background send little inhibition to nearby receptors, and therefore, there is no darkening effect [5].

During the last few decades, however, the balance has been shifted toward more high-level theories. This is because a series of phenomena have been presented that are difficult to explain on the basis of retinal interactions. One of the best-known demonstrations is the Benary cross (see Fig. 3), discovered by Max Wertheimer and studied further by Benary [6].

Although both gray triangles are surrounded by the same amount of black and white, the left triangle appears lighter than the right triangle. Even more interesting is the contrast phenomenon found by Agostini and Galmonte [7]. In this case, a gray region surrounded by a dark area appears darker than an identical gray region surrounded by a light area (see Fig. 4).

The interested reader can find more of these effects elsewhere [8–12]. As these examples are difficult to explain on the basis of retinal interactions, they have been interpreted according to the principles of perceptual organization [13, 14]; that is, the perceived color of a surface is determined primarily by the global contrast between the surface and the color of the surfaces to which it perceptually belongs



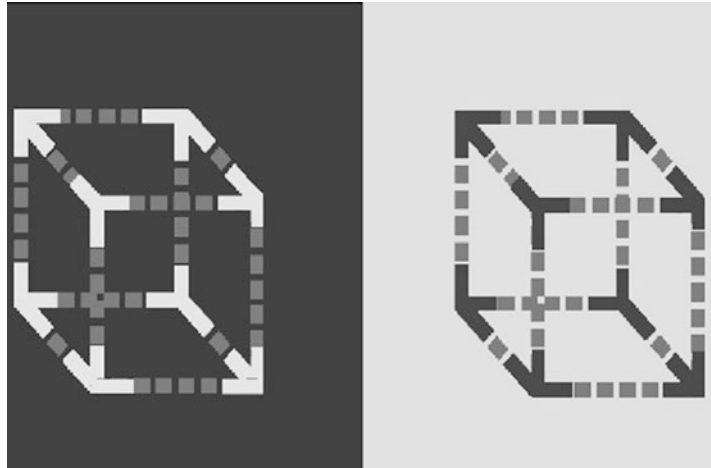
Simultaneous color contrast, Fig. 3 The Benary cross. The two *gray triangles* are the same. Although they are surrounded by the same amount of *white* and *black* area the *left triangle* appears lighter than the *right triangle*

(where perceptual belongingness refers to the grouping of a set of apparent elements into a perceived whole).

Although there are some exceptions (e.g., [15–18]) many theorists have now refused low-level explanations and have suggested that the retinal image is decomposed into separate components by mechanisms operating at a higher level of the visual process. Agreement among scientists, however, is still far from being achieved, and the debate is now between the *framework* and the *layer* type of decomposition models.

Simultaneous color

contrast, Fig. 4 The dashed elements of the cubes are the same *gray* but those to the *left* appear darker than those to the right despite the fact that they are surrounded by a darker area



The framework model maintains that the visual system, in agreement with the belongingness principles, processes the light reflected by surfaces that reaches the eyes by grouping it into a set of contiguous frameworks. According to this model, contrast effects are generated by the geometric and photometric relationships among the surfaces within the same perceptual group, i.e., within the same framework [14, 19].

Conversely, the *layer* model claims that the visual system operates by splitting the light reaching the eyes into separate overlapping layers which correspond to separate physical contributions (illumination, reflectance, transparency, etc.). According to this model, contrast effects are generated when the light reaching the eyes is misattributed to the surface color and illumination. In other words, part of the light that the visual system should attribute to the surface color is instead attributed to the illumination or vice versa [20–27].

In an attempt to disentangle these two models Soranzo, Lugin, and Wilson [28] studied the contrast phenomenon in the Virtual Reality Cave. This apparatus permitted the manipulation of the geometric relationships between the surfaces while at the same time maintaining the same photometric relationships (i.e., the amount of light reaching the observers' eyes remained constant). However, findings indicated that both of the decomposition processes which the framework and the layer models advocate are *jointly*

accountable for the color contrast phenomenon. The color contrast phenomenon may therefore be attributed to the summative effect of the framework and the layer processes.

Cross References

- ▶ [Anchoring Theory of Lightness](#)
- ▶ [Chevreul, Michel-Eugène](#)
- ▶ [Color Contrast](#)
- ▶ [Perceptual Grouping and Color](#)

References

1. Helmholtz, H.V.: *Helmholtz's Treatise on Physiological Optics*. Optical Society of America, New York (1866/1964)
2. Wade, N.J.: Descriptions of visual phenomena from Aristotle to Wheatstone. *Perception* **25**, 1137–1175 (1996)
3. Hering, E.: *Outlines of a Theory of The Light Sense*. (translated by Hurvich, L. M., Jameson, D.). Harvard University Press, Cambridge, MA (1874/1964)
4. Hartline, H.K., Wagner, H.G., Ratliff, F.: Inhibition in the eye of Limulus. *J. Gen. Physiol.* **39**, 651–673 (1956)
5. Cornsweet, T.N.: *Visual Perception*. Academic, New York (1970)
6. Benary, W.: Beobachtungen zu einem Experiment über Helligkeitskontrast. *Psychol. Forsch.* **5**, 131–142 (1924)
7. Agostini, T., Galmonte, A.: Perceptual organization overcomes the effects of local surround in determining simultaneous lightness contrast. *Psychol. Sci.* **13**, 88–92 (2002)

8. White, M.: A new effect of pattern on perceived lightness. *Perception* **8**(4), 413–416 (1979)
9. Knill, D.C., Kersten, D.: Apparent surface curvature affects lightness perception. *Nature* **351**, 228–230 (1991)
10. Adelson, E.H.: Perceptual organization and the judgment of brightness. *Science* **262**, 2042–2044 (1993)
11. Agostini, T., Proffitt, D.R.: Perceptual organization evokes simultaneous lightness contrast. *Perception* **22**(3), 263–272 (1993)
12. Lotto, R.B., Purves, D.: The effects of color on brightness. *Nat. Neurosci.* **2**(1), 1010–1014 (1999)
13. Wertheimer, M.: Untersuchungen zur lehre von der Gestalt. *Psychologische Forschung* **4**, 301–350 (1923)
14. Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., Economou, E.: An anchoring theory of lightness perception. *Psychol. Rev.* **106**, 795–834 (1999)
15. Kingdom, F., Moulden, B.: A multi-channel approach to brightness coding. *Vision Res.* **32**, 1565–1582 (1992)
16. Blakeslee, B., McCourt, M.E.: A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Res.* **44**, 2483–2503 (2004)
17. Todorovic, D.: Lightness, illumination, and gradients. *Spat. Vis.* **19**(2–4), 219–261 (2006)
18. Robinson, A.E., Hammon, P.S., de Sa, V.R.: Explaining brightness illusions using spatial filtering and local response normalization. *Vision Res.* **47**(12), 1631–1644 (2007)
19. Gilchrist, A.L.: Seeing black and white. New York: Oxford University Press (2006)
20. Eagleman, D.M., Jacobson, J.E., Sejnowski, T.J.: Perceived luminance depends on temporal context. *Nature* **428**, 854–856 (2004)
21. Gilchrist, A.: The perception of surface blacks and whites. *Sci. Am.* **24**(3), 88–97 (1979)
22. Gilchrist, A.L.: Lightness contrast and failures of lightness constancy: a common explanation. *Percept. Psychophys.* **43**(5), 415–424 (1988)
23. Schirillo, J.: Surround articulation I. Brightness judgments. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* **16**, 793–803 (1999)
24. Soranzo, A., Agostini, T.: Does perceptual belongingness affect lightness constancy? *Perception* **35**, 185–192 (2006)
25. Soranzo, A., Agostini, T.: Photometric, geometric and perceptual factors in illumination-independent lightness constancy. *Percept. Psychophys.* **68**(1), 102–113 (2006)
26. Soranzo, A., Galmonte, A., Agostini, T.: Lightness constancy: ratio invariance and luminance profile. *Atten. Percept. Psychophys.* **71**(3), 463–470 (2009)
27. Soranzo, A., Galmonte, A., Agostini, T.: The luminance misattribution in lightness perception. *Psihologija* **43**(1), 33–45 (2010)
28. Soranzo, A., Lugin, J.-L., Wilson, C.: The effects of belongingness on the SLC phenomenon: a VR study. *Vision Res.* **86**, 97–106 (2013)

Sky Glow

► [Light Pollution](#)

Sodium Gas Discharge Lamps

► [High- and Low-Pressure Sodium Lamp](#)

Soft Colors

► [Pastel Colors](#)

Solid-State Light Source

► [Light-Emitting Diode, LED](#)
 ► [Light-Emitting Diode, OLED](#)

Sparkle

► [Texture Measurement, Modeling, and Computer Graphics](#)

Spectral Luminous Efficiency

János Schanda
 Veszprém, Hungary

Synonyms

[Eye sensitivity functions](#); [Luminous efficiency function](#); [Relative visibility function](#); $V(\lambda)$, $V'(\lambda)$

János Schanda: deceased

Definition

Spectral luminous efficiency (of a monochromatic radiation of wavelength λ) [$V(\lambda)$ for photopic vision; $V'(\lambda)$ for scotopic vision] (see CIE e-ILV [1])

Ratio of the radiant flux at wavelength λ_m to that at wavelength λ such that both produce equally intense luminous sensations under specified photometric conditions and λ_m is chosen so that the maximum value of this ratio is equal to 1.

Unit: 1

Overview

By the end of the nineteenth century, photometry achieved a level where it became necessary to describe the light perception efficiency of radiation of different wavelengths. Several methods were used to determine the spectral luminous efficiency function (called also spectral visibility). The two most important ones were direct comparison and the flicker method (based on the fact that if two lights are seen alternatively with increasing frequency, first the sensation of hue difference is lost and only later the sensation of intensity difference; choosing a frequency when color sensation is already lost but the intensity flicker is still perceived (in the order of 15 Hz), one can equate the “intensity” of the two lights by minimizing the flicker). Based on many investigations, CIE agreed to a table showing the relative visibility of lights of different wavelengths compared to the light with highest visibility in 1924 [2].

Spectral luminous efficiency $V(\lambda)$ for foveal vision at photopic levels: The present standard [3] is based on the table recommended in 1924, practically only interpolating and extrapolating the values with a very minor smoothing made. This $V(\lambda)$ function is the basis of present-day photometry, and due to the fact that the CIE 1931 colorimetry was based on photometry, the $2^\circ \bar{y}(\lambda)$ function was chosen to be equal to the $V(\lambda)$ function. The $V(\lambda)$ function is – more or less – valid (see later) at high luminance levels (accordingly, the $V(\lambda)$ function should be used

above 5 cd/m^2 luminance, and it should mimic the spectral sensitivity of one of the cone signal channels (composed from the L and M cone signals) of human vision for foveal vision (part of the human retina where the spatial sensitivity (visual acuity) is highest. This part of the retina is covered by a yellowish layer (macula lutea) that functions as a color filter. At larger eccentricity angles, where the macula lutea does not cover the fovea anymore, the eye sensitivity is different, see $V_{10}(\lambda)$ later).

Spectral luminous efficiency $V'(\lambda)$ for foveal vision at scotopic levels: At low light levels, the rods take over the role of vision; this is the range of scotopic vision. The spectral sensitivity of the rods differs from that of the cones. The spectral luminous efficiency for scotopic vision was internationally agreed in 1951 [5] and was standardized in 2004 [4]; it should be used if the luminance is smaller than 0.005 cd/m^2 .

Mesopic spectral luminous efficiency $V_{\text{mes}}(\lambda)$: Between scotopic and photopic luminance levels, both rods and cones contribute to the spectral luminous efficiency. CIE published in 2010 a method to calculate the luminance level-dependent spectral luminous efficiency function in this mesopic range [4].

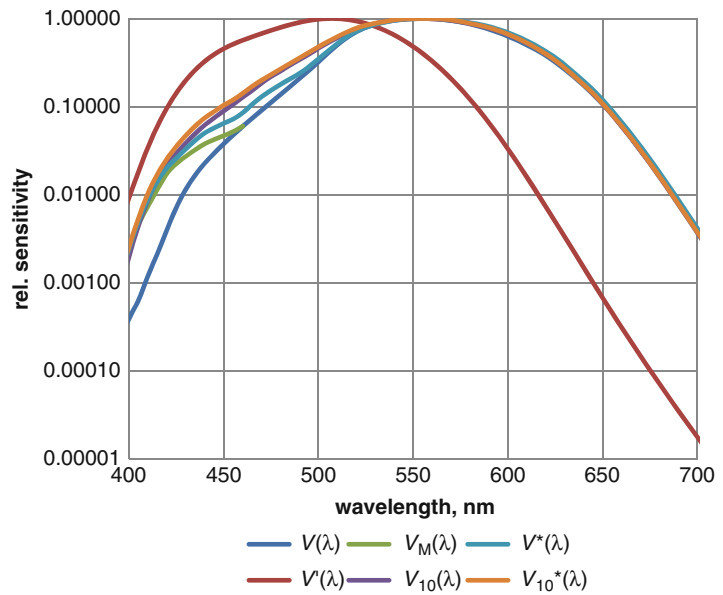
Some Further Visibility Functions

Modified 2° spectral luminous efficiency: It turned out quite early that the values of the 2° spectral luminous efficiency function are lower as would correspond to average human vision, but it took till 1990 that CIE adopted the $V_M(\lambda)$ function [6] and recommended it for applications in visual sciences. For further details on how this function has been derived and what forerunners have been published, see above publication.

Spectral luminous efficiency for large visual fields: There was no internationally accepted spectral luminous efficiency function till 2005, when CIE adopted the $V_{10}(\lambda)$ function based on the 10° standard colorimetric observer's $\bar{y}_{10}(\lambda)$ function [7].

Cone fundamental-based spectral luminous efficiency functions: Recent investigations have

Spectral Luminous Efficiency, Fig. 1 Some more important spectral luminous efficiency functions



shown that in human vision, one of the visual channels provides a luminance-like signal to the brain. Based on detailed investigations, CIE determined the average cone fundamental spectral response functions (spectral responsivities at human eye cornea level) and recommended color matching functions that correspond better to average human color perception as the CIE 2° and 10° standard observer functions [8].

Based on the above, Stockman and coworkers determined an average spectral luminous efficiency function for daylight adaptation [9] taking world population genotype distribution into consideration (the genotype influences the cone fundamental spectral sensitivity $V^*(\lambda)$).

In the following figure, we summarize the different spectral luminous efficiency functions in logarithmic representation. As can be seen, the traditional – and still official – photometric curve is way too low beyond 450 nm, having as a consequence erroneous photometric evaluation of blue LED lights. The $V_M(\lambda)$ curve makes some correction to this discrepancy. The $V^*(\lambda)$ curve is the most recent modification [9] based on cone fundamental signals, average population, and daylight adaptation (earlier $V(\lambda)$ functions have not taken any adaptation into consideration). The $V_{10}(\lambda)$ and $V_{10}^*(\lambda)$ curves differ from each

other only very slightly, while the $V'(\lambda)$ function – as it represents the input from a quite different type of light-sensitive cells – shows markedly different character (Fig. 1).

One has to emphasize, however, that up to now (2013), all practical photometric calculations are based on the CIE 1924 $V(\lambda)$ function and the Meter Convention has endorsed only the $V(\lambda)$ and the $V'(\lambda)$ functions. All other visibility functions have been recommended only for trial and some eye research applications. With the increasing use of LED-based light sources that contain a pronounced spectral maximum in the blue part of the spectrum, prediction errors using the $V(\lambda)$ function become apparent, and the use of more modern spectral luminous efficiency functions might become necessary.

Cross-References

- [CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments](#)
- [CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength](#)
- [CIE Special Metamerism Index: Change in Observer](#)

References

1. CIE e-ILV item 17–1222 in <http://eilm.cie.co.at/term/1222>
2. CIE Compte Rendu 6e session, p. 67
3. CIE Photometry – The CIE System of Physical Photometry CIE S 010/E:2004/ ISO 23539:2005(E)
4. CIE Recommended System for Mesopic Photometry based on Visual Performance, CIE 191:2010
5. CIE Compte Rendu 12e session, vol. 3, p. 37
6. CIE 1988 Modified 2° Spectral Luminous Efficiency Function for Photopic Vision”, $VM(\lambda)$, CIE 86–1990
7. CIE 10° photopic photometric observer”, $V_{10}(\lambda)$, CIE 165:2005
8. CIE Fundamental Chromaticity Diagram with Physiological Axes – Part 1, CIE 170-1:2006, and draft of Part 2
9. Sharpe, L.T., Stockman, A., Jagla, W., Jägle, H.: A luminous efficiency function, $V^*(\lambda)$, for daylight adaptation. *J. Vis.* **5**, 948–968 (2005)

Spectral Power Distribution

Yi-Chen Chuang

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Spectral power distribution curves](#)

Definition

Spectral power distribution (SPD) is a function of wavelength and describes the amount of optical radiation across part of a spectrum. SPD usually describes optical radiation emitted by a light source, but generally, SPD can be applied to describe optical radiation reflected by an object, or transmitted from a transparent/semi-transparent object.

According to the definition of the Illuminating Engineering Society of North America (IESNA), SPD is “a pictorial representation of the radiant power emitted by a light source at each wavelength or band of wavelengths in the visible region of the electromagnetic spectrum.” Ideally,

SPD should be expressed in units of W/nm. However, in some of the literature, SPD is a general term used to describe any radiometric or photometric quantity as a function of wavelength (with corresponding units). In many conventional cases, SPD is referred to as radiant power per unit area as a function of wavelength, and this should be correctly termed spectral irradiance distribution (or radiance exitance distribution, depending on the context) expressed in units of W/m²/nm. Although based on the definition of the IESNA, SPD is limited to visible light. SPD is generally applicable to ultraviolet radiation, visible light, and infrared radiation. In color applications, SPD in the visible light region (380–780) nm is the main focus of interest, but SPD in the ultraviolet region is also important because it influences the color perception of fluorescent materials.

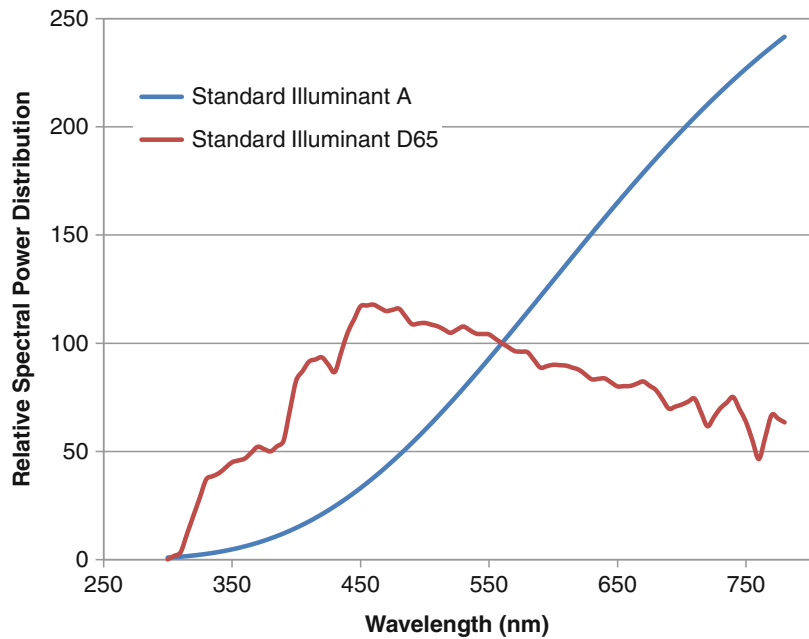
As the overall power levels of light sources can vary over many orders of magnitude, SPDs are often normalized for easy comparisons of color properties. Such SPDs are referred to as relative SPDs (dimensionless). A relative SPD is usually normalized so that it has a maximum value of 100 (or 1).

Overview

Spectral power distribution is a function of wavelength and describes the amount of optical radiation within a particular range of a spectrum [1–4]. SPD is used to describe not only light sources but also illuminants. The difference between light source and illuminant is that a light source is an actual physical entity, whereas an illuminant may only be a set of values recording its SPD. For example, International Commission on Illumination (CIE) Standard Illuminant A is the SPD of CIE Source A (approximated by an incandescent, tungsten filament light source). Many CIE Standard Illuminants do not exist, that is, there is no actual light source that can realize the SPD defined for such an illuminant. For example, the CIE D series of illuminants represents various phases of daylight, but no source is available for the D illuminants [1, 5]. Figure 1 shows

Spectral Power Distribution,

Fig. 1 Spectral power distribution of International Commission of Illumination (CIE) Standard Illuminants A and D65



the SPDs of CIE Standard Illuminant A and Standard Illuminant D65 from 300 to 780 nm, as examples of SPD [6].

In color science, SPD describes the color characteristics of a light source as a function of wavelength over the visible region. SPD is also often used to describe the color characteristics of an object because optical properties such as transmittance, reflectance, and absorbance in addition to the detector response are typically dependent on the incident wavelength [7]. Knowledge of SPD contains all the basic physical data about the light and serves as the starting point for quantitative analyses of color. However, SPD cannot provide the human perception of objects [8]. Note that human perception is based on three components: SPD of light, reflectance of objects, and the color-matching function of human observers.

The SPD of light sources can be measured using a monochromator with optical detectors, or a spectroradiometer. An example of experimentally measured spectral irradiance distribution of a 1000 W tungsten lamp representing SPD is shown in Fig. 2. SPDs of other light sources can be found

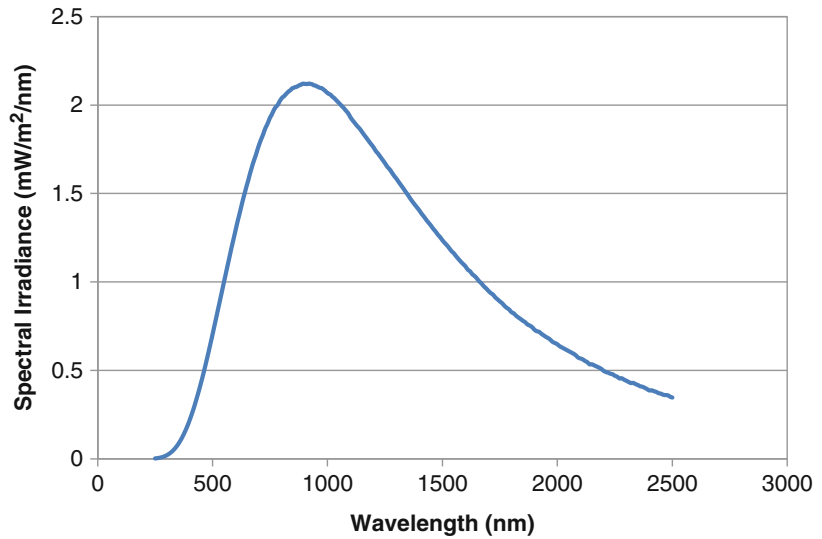
in a lamp manufacturer's manual, for example, or SPD simulators on-line for educational purposes [8].

By engineering SPDs of different light sources, it is possible to obtain the desired color property of light. These technologies have been applied to solid state lighting, such as white light emitting diodes (LEDs). Samples of SPDs of white LEDs are shown in Fig. 3. The "apparent white" of white LEDs is a "mixture" of different colors of light, where blue light is essential, as shown in Fig. 3. There are many ways to obtain precise color quality parameters, such as correlated color temperatures, illuminance, color rendering, luminous efficacy, etc. Note that different SPDs of light can result in similar apparent colors of light, but quite different color quality.

On the other hand, the apparent color of an object can be described by the SPD of the reflected light from an illuminated surface and is the product of the percentage of reflectance of the surface and the SPD of the light that falls on the surface. Thus, an object may appear differently affected by different light illuminations.

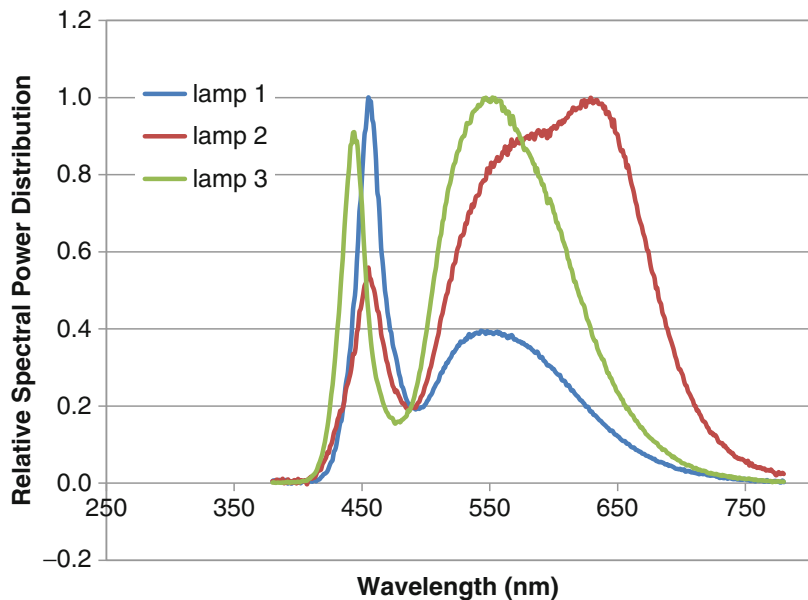
Spectral Power Distribution,

Fig. 2 Spectral power distribution of a 1000 W tungsten lamp



Spectral Power Distribution,

Fig. 3 Sample spectral power distribution of white light emitting diodes



Cross-References

- [CIE Color-Rendering Index](#)
- [CIE Standard Illuminants and Sources](#)

References

1. Webster, J.G. (ed.): The Measurement, Instrumentation, and Sensors Handbook. Springer, Heidelberg (1999)
2. Peres, M.R.: The Focal Encyclopedia of Photography. Focal Press, Amsterdam (2007)
3. Fairchild, M.D.: Color Appearance Models. Wiley, Hoboken (2005)
4. Grum, F.C.: Optical Radiation Measurements, vol. 1. Academic, New York (1979)
5. CIE, Technical Report: Colorimetry, 3rd ed., Publication 15:2004, CIE Central Bureau, Vienna (2004)
6. Joint ISO/CIE Standard: ISO 10526:1999/CIE S005/E-1998 and CIE Standard Illuminants for Colorimetry
7. McCluney, W.R.: Introduction to Radiometry and Photometry. Artech House, Boston (1994)
8. <http://hyperphysics.phy-astr.gsu.edu/hbase/vision/spd.html> HyperPhysics (©C.R. Nave, 2012)

Spectral Power Distribution Curves

► Spectral Power Distribution

Spectrometer

Yi-Chen Chuang
Center for Measurement Standards, Industrial
Technology Research Institute, Hsinchu, Taiwan

Synonyms

Spectroradiometer

Definition

A spectrometer is an optical instrument designed to measure the spectrum and stimulus of a light source or self-luminous object. The prefix “spectral” is an abbreviation of the phrase “spectral concentration of,” which is defined by the International Commission on Illumination (CIE) as the “quotient of the radiometric quantity taken over an infinitesimal range on either side of a given wavelength, by the range”.

Overview

Basic Principles

A spectrometer receives light, disperses the light, and converts the optical signals (or more generally, electromagnetic radiation signals) to electrical signals. It provides the radiometric quantities in narrow wavelength intervals. Because many sources have line structures, it is useful to sample the spectrum with narrow bandwidth and wavelength increments [1, 2]. In general, spectral irradiance varies from point to point on the surface. In practice, it is important to note how radiant flux varies with direction, the size of the solid angle subtended by the source at each point on the surface, and the orientation of the surface. Given

these considerations, it is often more prudent to use a stricter form of the formula, to account for these dependent relationships [3]. Spectral irradiance is the desired measurement for general spectroradiometry. In practice, the average of the spectral irradiance is measured, shown mathematically as the approximation:

$$E(\lambda) = \frac{\Delta\Phi}{\Delta A \Delta\lambda}$$

where Φ is the radiant flux of the source (SI unit: watt, W) within a wavelength interval $\Delta\lambda$ (SI unit: meter, m), incident on the surface area, A (SI unit: square meter, m²), and the calculated result E is the spectral irradiance. It is more useful to measure wavelength in nanometers; therefore, the submultiples of the SI units of spectral irradiance are often used, for example, $\mu\text{W}/(\text{cm}^2 \cdot \text{nm})$ [3].

A spectrometer is very similar to a spectrophotometer, another optical instrument used to measure an electromagnetic radiation spectrum. However, there are some differences. The name spectrometer is applied to instruments that operate over a very wide range of wavelengths to measure self-luminous objects. An optical spectrometer could include the UV–VIS–IR spectrum; spectrophotometers are usually used in spectral regions near the visible light spectrum. In general, the reflectance or transmittance is measured. For color measurements, spectrometers are used to measure the color of a light source, while spectrophotometers are used to measure the color of an object [2].

Applications

Spectrometers take the spectral radiance of measurements and spectral irradiance is typical. Based on the spectral data measured, the radiometric, photometric, and colorimetric quantities of light can be calculated, providing a complete description of the color of the source [3]. Thus, the spectrometer is an important tool in the color industry as the basis for calibration. Quantities commonly used in the color industry include CIE tristimulus values, CIE chromaticity

coordinates (x, y), correlated color temperature (CCT), color rendering index (CRI), etc. [2].

An automatic scanning spectrometer includes a mechanism to change the wavelength selected by the monochromator and to record the resulting changes in the measured quantity as a function of the wavelength. Many commercial spectrometers have included built-in post-calculation functions for easy usage of the spectrometer for various applications. The latter is the basic configuration of a spectrometer, which usually includes a dispersing device and a detecting device (i.e. an optical detector or alternative).

This configuration allows the simultaneous analysis of colors. Such an instrument can record a spectral function without mechanical scanning, which may be corrected in terms of resolution or sensitivity, for instance. The absorption spectrophotometer measures the absorption of light by a sample as a function of wavelength.

References

1. Spectroradiometer: http://en.wikipedia.org/wiki/Spectroradiometer#cite_note-3.
2. Janos, S.: Colorimetry: Understanding the CIE System. Wiley, Hoboken (2007)
3. IES, Illuminating Engineering Society. Lighting measurements: a basic approach to understanding the principals of lighting science. Handout. Comp. Nlena and Associates, Inc. Retrieved from <http://www.ies.org/pdf/education/IES-Color-3-Webcast-Handout.pdf>.
4. Stroebel, L.D., Zakia, R.D.: Focal Encyclopedia of Photography, 3rd edn, p. 115. Focal Press, London (1993). ISBN 0-240-51417-3

Spectroradiometer

► [Spectrometer](#)

Spectroreflectometer (for Reflectance Measurement)

► [Instrument: Spectrophotometer](#)

Sports Lighting

Gilles Page

Philips Lighting, Miribel, France

Definition

The purpose of sports lighting is to create the ideal conditions of visibility for those who experience sports facilities.

User Requirements

The great interest of the media in sports events has an important effect on the development of the sports themselves. This, together with the increased leisure time, has resulted in a booming number of active sportsmen and women all over the world, either taking part in the sport as a leisure activity or in amateur and professional competitions.

Consequently numerous sports complexes have been developed in which all kinds of sports are practiced, from the general physical training sessions to professional activities. In the organization of sports activities and events, the quality of the facilities available is for the utmost importance for pleasure and in the outcome of the competition. Space, a good playing surface, heating/cooling, safety provisions, etc., are all of major importance.

One essential contribution to a successful sports venue, which is still not always recognized as such, is the standard and quality of the lighting system. This, of course, is not only for TV transmissions, where lighting levels of sufficient quality are an absolute necessity to assure an optimal television transmission, but also for the general practicing of the sports, guaranteeing, for example, a clear view of the components, the objects like the ball, and of course, the indispensable marking of the competitive area itself.

The users of sports grounds can be distinguished according to their activities:

- The players
- The technical staff, referees, and team officials
- The spectators arriving, watching the game, and leaving the sports facility
- Both the television and film crews and the photographers recording the event

The players, referees, and officials must be able to see clearly all that is going on in the competition area so that they can produce their best possible performance and make accurate decisions.

Spectators should be able to follow the performance of the players and the action in an agreeable environment. The latter requirement means that they must be able to see not only the competition area but also its immediate surroundings. The lighting should also enable spectators to safely enter and leave the sports facility. With large groups this safety aspect is very important.

For television and film coverage, the lighting must be able to ensure high quality color images, not only of the overall action but also close-ups of the players and spectators alike. Closeup images are important to convey the emotions and atmosphere in a stadium to viewers watching at home.

Sports Lighting Classes

As the competence level of athletes increases so does the speed of the action; the visual task becomes more difficult, requiring more light of a higher quality.

In general, five levels of sporting activity are recognized: international and national competition, regional competition, local competition, training, and recreational. These levels are related to the standards of play and the viewing distances of the spectators. These five levels do not all require the same quality of lighting. Lower standards are clearly acceptable for recreation than for national competitions. To cover all five activities levels, three lighting classes are given in Table 1 and are defined:

Class I: Top-level competition such as national and international matches, which generally

Sports Lighting, Table 1 Lighting classes according to CEN [1]

Level of activity	Class I	Class II	Class III
International and national competition	•		
Regional competition	•	•	
Local competition	•	•	•
Training		•	•
Recreational			•

involve large spectator capacities with potentially long viewing distances. Top-level training may also be included in this class.

Class II: Mid-level competition such as regional or local club matches which generally involve medium-size spectator capacities with medium viewing distances. High-level training may also be included in this class.

Class III: Low-level competition such as local or small club matches which do not usually involve spectators. General training and recreation also come into this class.

Lighting Criteria

Knowing the general user requirements it is possible to determine the lighting criteria for each of the different level of activity. The purpose of this section therefore is to identify these lighting criteria and, wherever possible, to derive the lighting parameters of interest in each case.

Illuminance on a Horizontal Plane

The illuminated playing surface forms a major part of the field of view for the players, officials, and spectators. It is the illuminance on this horizontal plane, at ground level, commonly referred to as the horizontal illuminance (E_h), which chiefly serves to establish the adapted state of the eye, by creating a stable visual background against which players and objects will be seen. Because of this an adequate horizontal illuminance on the field is important.

To ensure safety of movement for the spectators when entering and leaving the stands or surrounds, adequate horizontal illuminance in these areas is also required.

Vertical Illuminance

Because it is important to follow the progress of the players and the ball which can be approximated as vertical surfaces, it is the illuminance on the vertical surfaces (E_v) which describes it best. However experience shows that where the human observer is concerned, there is a good relationship between vertical and horizontal illuminance. Therefore, for all activities with the exception of television broadcasting, if the specified horizontal illuminance is provided and the design rules are followed, the vertical illuminance will be sufficient.

The scene illuminance, and more particularly the vertical illuminance, has a major influence on the quality of the final television or film picture. To guarantee an optimal view and identification of players from all directions, specified illuminances on vertical planes at a height of 1.5 m are required.

For television or filming with fixed camera positions, it is sufficient to ensure that the illuminances on planes at right angles to the camera positions are adequate. In the case of an unrestricted choice of camera positions, the vertical illuminance on planes facing all four sides of the competition area should be taken into account.

Uniformity

Good illuminance uniformity is important in order to avoid both adaptation problems for players and spectators. If the uniformity is not adequate, there is a risk that an object or player details, like sponsor logos and emotions, will not be clearly seen at certain positions on the competition area.

Uniformity is expressed as the ratio of the lowest to the highest ($U1 = E_{min}/E_{max}$) illuminance and as the ratio of the lowest to the average ($U2 = E_{min}/E_{average}$) illuminance.

Even when the uniformity ratios as defined above are acceptable, changes in illuminance can be disturbing if they occur over a too short distance. This problem is most likely to arise when panning a television camera. Therefore, the illuminance uniformity for TV/film coverage at a certain grid point has to be expressed as a percentage change from the average adjacent grid points. This is called the uniformity gradient.

Glare Restriction

Glare is a subjective factor, for which a practical evaluation system has been described for outdoor sports applications, by CIE (*CIE 112 glare evaluation system for use within outdoor sports and area lighting*), based on extensive field tests [2, 3]. CIE 112 defines a so-called glare rating factor (GR) for which an assessment scale is given from 10 to 90. The lower the value of GR, the better glare situation is. A maximum GR value of 50 is generally specified in high-end sports projects. The validity of the system is restricted to viewing directions below eye level. The development of the glare rating GR assessment method has led to a reliable system that can be used to make a qualitative statement about the suitability of a lighting system before it is installed.

This assessment method is supposed to be used both for predicting degree of glare at the design stage and checking glare situation of installation, but in reality it is almost never used for the latter situation as it is quite impossible (very difficult) to measure reliable glare on site.

It is also a reason why it is important to refer to guidelines for the positioning of floodlights.

Modeling and Shadows

Modeling is the ability of the lighting to reveal form and texture. This “modeling” ability is particularly important to provide a pleasant overall impression of the athletes, objects, and spectators on and around the pitch.

An installation where light only comes from one direction will cause harsh shadows and poor modeling. The players will appear to be pushed into the background and will only be seen from the same direction as which the light comes. From all other directions they will be seen as dark silhouettes only.

Color and Color Properties of Lamps

Color perception is important in most sports, and while some color distortion, attributable to the artificial lighting, is acceptable, there should not be problem being able to distinguish between different colors.

There are two important aspects of the color properties of lamps:

The color appearances of the light. This is the color impression of the total environment created by the lamp.

The color-rendering properties of the lamps used or CIE color-rendering index. This describes how faithfully a range of colors can be reproduced by a light source.

Both the color appearance and the color-rendering properties of lamps are completely dependent upon the spectral energy distribution of the light emitted. An indication of the color appearance of a lamp can be obtained from its correlated color temperature in degrees Kelvin indicated as T_k , which vary mainly between 2,000 and 6,500 K. The lower the color temperature, the “warmer” the color impression of the light is; the higher the color temperature, the “cooler” or more bluish the impression of the light.

The color-rendering properties of a light source can be indicated by its color-rendering index R_a expressed as numerical values from 0 to 100. A light source having a color-rendering index of 100 will represent all scene colors faithfully and can be compared with daylight.

Lighting Recommendations

Taking the various lighting criteria discussed in the previous section as a basis, it is possible to draw up two sets of lighting recommendations. The first set, presented in Tables 2, and 3, concerns the lighting needs of players, officials, and spectators and covers a wide range of sporting activities and various levels of play, respectively, for outdoor and indoor activities.

The second set, presented in Table 4 and 5, concerns the different lighting needs for TV coverage.

These tables are the ones published in the GAISF guide for the artificial lighting of indoor and outdoor sports venues (GAISF General Association of International Federations) [4, 5]. Other

publications from national or international sports federations exist and may provide slight differences in the recommendation values; the interest of the GAISF guide is to summarize in one document all the recommendations for the most practiced sports activities and to be aligned with the recommendations from the European standard (EN 12193 Sports Lighting) [6].

It should be noted that different practices are used in the USA; the Illuminating Engineering Society (IES) [7] publishes illuminance recommendations for sports and recreation in its lighting handbook [8] that differ from the ones defined by GAISF guide and EN12193; the main differences are the following:

- Four lighting classes defined.
- Nearly all sports activities have both a horizontal illuminance and a vertical illuminance component.
- Illuminance criteria are based on visual ages of more than half the intended observers.

As a consequence the recommended lighting levels are generally higher than the ones defined in the GAISF guide/EN 12193 Norm.

Before consulting these tables, the following points should be noted.

Horizontal Illuminance

The horizontal illuminance needed on the playing area for players, officials, and spectators depends on several factors like:

- The size of the playing object
- The reflectance and color of the playing object relative to the reflectance and color of the background surfaces
- The viewing distance
- The speed of the playing object
- Whether the sports facility is located indoors or outdoors

Levels of illuminance for outdoor facilities are usually lower than for the same sports when

Sports Lighting, Table 2 Lighting recommendations for non-televised outdoor activities

Class	Horizontal illuminance (lx)	Uniformity min/ave	Color rendering	Glare rating
American football, athletics, basketball, cycle racing, equestrian, fistball, football, netball, rugby, and volley				
I	500	0.7	>60	<50
II	200	0.6	>60	<50
III	75	0.5	>20	<55
Note: for Class III athletics and equestrian sports, the minimum illuminance is 100 lx as per EN12193				
For athletics this can be reduced to 50 lx for running sports. For cycling Class II: 300 lx (0.7, Class III: 100 lx				
Swimming (aquatic sports)				
I	500	0.7	>60	<50
II	300	0.7	>60	<50
III	200	0.5	>20	<55
Note: for diving, vertical uniformity should also be considered. Class I: 0.8 Eh/Ev. Class II: 0.5 Eh/Ev. Class III: 0.5 Eh/Ev				
Tennis				
I	500	0.7	>60	<50
II	300	0.7	>60	<50
III	200	0.6	>20	<55
Note: values refer to “total playing area” as defined by ITF				
Baseball, bandy, cricket, hockey, ice hockey				
I	750	0.7	>60	<50
II	500	0.7	>60	<50
III	300	0.7	>20	<55
Outfield for baseball, cricket, and softball				
I	500	0.5	>60	<50
II	300	0.5	>60	<50
III	200	0.3	>20	<55
Bobsleigh and luge				
I	300	0.7	>60	<50
II	200	0.5	>60	<50
III	50	0.4	>20	<50
Boules sport (lawn, raff, and petanque)				
I	200	0.7	>60	<50
II	100	0.7	>20	<50
III	50	0.5	>20	<55
Archery				
I, II, III	200	0.5	>60	
Vertical illuminance	Target	Uniformity min/ave		
I, II, III	750	0.8		
Alpine and freestyle skiing				
I	150	0.5	>60	<50
II	100	0.4	>20	<50
III	50	0.3	>20	<55
Ski jump landing area				
I	300	0.7	>60	<50
II	200	0.6	>20	<50
III	200	0.6	>20	<55

Sports Lighting, Table 3 Lighting recommendations for non-televised indoor activities

Class	Horizontal illuminance (lx)	Uniformity min/ave	Color rendering	Glare rating
Aikido, basketball, body building, cycle racing, fistball, floorball, football, handball, jujitsu, judo, karate, korfbal, netball, powerlifting, sambo, sepak takraw, school sports (physical education), sumo, taekwondo, volleyball, weight lifting, wrestling, wushu				
I	750	0.7	>60	n/a
II	500	0.7	>60	n/a
III	200	0.5	>20	n/a
Boxing				
I	2,000	0.8	>80	n/a
II	1,000	0.8	>80	n/a
III	500	0.5	>80	n/a
Note: vertical illuminance at 1.5°m should be >50% of Eh				
Athletics, dancing, equestrian sports, gymnastics, roller sports, and wall climbing				
I	500	0.7	>60	n/a
II	300	0.6	>60	n/a
III	200	0.5	>20	n/a
Note: for wall climbing, Class I: 500 lx vertical. Class II: 300 lx vertical. Class III: 200 lx vertical				
Swimming (aquatic sports)				
I	500	0.7	>60	n/a
II	300	0.7	>60	n/a
III	200	0.5	>20	n/a
Note: for diving, vertical uniformity should also be considered. Class I: 0.8 Eh/Ev. Class II: 0.5 Eh/Ev. Class III: 0.5 Eh/Ev				
Tennis				
I	750	0.7	>60	n/a
II	500	0.7	>60	n/a
III	300	0.5	>20	n/a
Note: values refer to “total playing area” as defined by ITF				
Badminton, basque pelota, cricket, cricket nets, curling, fencing, hockey, ice hockey, ice skating, racquetball, squash, and table tennis				
I	750	0.7	>60	n/a
II	500	0.7	>60	n/a
III	300	0.7	>20	n/a
Note: for fencing, Class I: 500 lx vertical. Class II: 300 lx vertical. Class III: 200 lx vertical				
Cricket net, Class I: 1,500 lx (0.8).Class II: 1,000 lx (0.8).Class III: 750 lx (0.8)				
Billiards				
I	750	0.8	>80	
II	500	0.8	>80	
III	500	0.8	>80	
Boules sport (lawn, raff, and petanque)				
I	500	0.8	>60	n/a
II	500	0.8	>60	n/a
III	300	0.5	>20	n/a
Bowling, archery, shooting				
I	200	0.5	>60	n/a
II	200	0.5	>60	n/a
III	200	0.5	>60	n/a
Vertical illuminance	Pins	Target 25 min	Target 50 min	min/ave
I, II, III	500	1,000	2,000	0.8

Sports Lighting, Table 4 Lighting recommendations for televised events for major events

Major events	Horizontal illuminance			Vertical illuminance			Color rendering (Ra)	Glare rating
	Average (lx)	Uniformity min/ave	Uniformity min/max	Average (lx)	Uniformity min/ave	Uniformity min/max		
HDTV	1,500–3,000	0.8	0.7	2,200	0.7	0.6	>90	<50
Slow-motion camera	1,500–3,000	0.8	0.6	1,800	0.7	0.5	>80	<50
Fixed camera	1,500–3,000	0.8	0.6	1,400	0.7	0.5	>80	<50
Mobile camera	1,500–3,000	0.8	0.6	1,200	0.5	0.3	>80	<50

Average horizontal and vertical illuminance ratios: It is recommended that the ratio for horizontal illuminance (field of play) is between 0.75 and 1.5 of the vertical illuminance for cameras. Where there is HDTV all horizontal values for other cameras are as for HDTV

Sports Lighting, Table 5 Lighting recommendations for televised events for national events

National events	Horizontal illuminance			Vertical illuminance			Color rendering (Ra)	Glare rating
	Average (lx)	Uniformity min/ave	Uniformity min/max	Average (lx)	Uniformity min/ave	Uniformity min/max		
Camera	1,000–2,000	0.7	0.5	1,000	0.6	0.4	>80	<50

played indoors; this is due to the corresponding improvement in contrast and adaptation that occurs when viewing a lit object against a dark background such as the night sky.

The illuminances on a horizontal plane needed for competitions with TV coverage are not listed. Such levels are secondary to the requirements expressed in terms of vertical illuminance.

To ensure the television picture has a well-balanced brightness, the ratio between the average vertical and horizontal illuminance should be as closely matched as possible but should not exceed the ratio of 0.5 to 2 times. That is, the horizontal illuminance should not be less than half the vertical illuminance or greater than twice vertical illuminance.

Vertical Illuminance

In practice, the vertical illuminance required for players and spectators will usually automatically be obtained if the requirements regarding the horizontal illuminance are fulfilled.

Where television broadcasting is a requirement, it is necessary to provide an adequate vertical illuminance across the scene viewed by the camera. If

the vertical illuminance is not sufficient, good quality broadcast pictures will not be possible.

Television cameras are not able to adapt to changes and fluctuations in lighting level as quickly as the human eye. Therefore this limitation must be taken into account when designing lighting systems for televised events.

In most of the sports activities, television cameras will be viewing predominantly vertical objects between the ground and a height of 1.5 m above the ground. In that case, the illuminance should be specified as the illuminance on a vertical plane 1.5 m above the ground facing the direction of the television camera. In other cases, when the main action takes place at a different height, the illuminance should be specified where the main action takes place; for example, for diving the vertical illuminance should be specified on a vertical plane along the diving fall and into the direction of the camera.

It is now common that many cameras are used, distributed around the arena to obtain closeup action shots from alongside each event area. So it is important to ensure that the vertical illuminance towards each camera is sufficient.

To guarantee recommended average illuminances, during the whole period of operation of an installation, the illuminances should never fall below the indicated values. The recommended average illuminances (horizontal and vertical) listed in Tables 2 to 5 are therefore “maintained” values. To arrive at “initial values,” these maintained values have to be multiplied by a factor of 1.25.

Illuminance Uniformity

The uniformity requirements for both horizontal and vertical illuminance are more stringent for TV and filming than for the players and spectators, the human eye being less sensitive to nonuniformities than are cameras.

It is principally the horizontal uniformity that determines the brightness range of the overall scene, so this has to satisfy the more stringent requirements.

Color Properties of Lamps

In sports application where color recognition is an important factor especially in case of media coverage, it is essential to use high-color-rendering light sources. In most of the specifications for high-end sports events with TV coverage, it is recommended to use Ra 80 or even 90 lamps.

With the advent of HDTV, it is important that true color scenes are properly seen. For non-televised events, then a lower color-rendering index around 65 is generally specified.

Except in case of TV and film coverage, the color temperature quoted in the recommendations is merely to ensure an acceptable ambiance.

In order to acquire natural color scenes, the camera needs to be white balanced. Because a camera system can only adapt to one color temperature at a time, lighting systems used in combination with daylight should have a color temperature close to that of daylight. In addition the preferred photographic films for sports usage are daylight balanced to around 5,500 K, so light sources with a color temperature between 5,500 and 6,000 K are generally recommended in case of TV coverage.

Flicker Effect

A particular problem for super slow-motion cameras is 50 Hz flicker due to the phasing of the light.

Cameras perceive changes in light levels due to the uneven ratio between the camera scanning frequency and the alternating amplitude of artificial lights powered by mains frequency. This effect is only visible during slow-motion replay.

An AC lamp “flashes” at twice the mains frequency. Due to the speed of the flashes they are imperceptible to the human eye and the illusion is created that the lighting is continuous.

The exposure time of a standard camera corresponds to twice the lighting period. Therefore, a standard camera is integrating each field with an equal amount of light. However, when shooting with a high-speed camera, the behavior of artificial lighting will be shown on the picture, exactly as it is occurring.

In triple speed mode (150 Hz) the exposure time is one third of the exposure time in single speed mode. The alternating light amplitude causes different video levels for the three fields. This effect is only visible during slow-motion replay.

Glare Restriction

The visual performance is reduced by glare which shall therefore be limited.

In sports lighting applications, mounting height of the luminaire is a key factor influencing the quality of an installation. This parameter has a big impact on glare and spill light; so it has to be carefully defined, in relation with the aiming pattern, in order to limit glare while complying with the required lighting level and uniformities.

Additionally it is wise to avoid installing floodlights in the main viewing directions of the athletes.

Environmental Aspect

The control of stray light from outdoor lighting installations can also be disturbing to people outside the lighted area, for example, to traffic on adjacent roads and to inhabitants of houses in the neighborhood of the sports facility. In some cases local authorities or municipalities have their own

guidelines on such matters, and their advice should always be sought.

It should be noted that stray light falling outside the playing area cannot be effectively controlled by lowering the mounting heights of the floodlights. This only results in luminaires being aimed in a more vertical direction causing even more glare.

In large stadiums it is now common practice for large video screens or scoreboard systems to be used. In these cases care must be taken to ensure that luminaires are not positioned too close to these screens as this could cause glare for persons looking up at the screen.

Safety Lighting for Participants

Participant safety is ensured by the safe stopping of an event which might otherwise be dangerous to continue in the absence of lighting.

European Norm EN 12193 about sports lighting specifies safety lighting for some sports activity where the absence of lighting could create a big danger.

The safety lighting shall come on the instant the general lighting fails and last for at least the period specified.

Emergency Lighting

For the purpose of safety and orientation for the spectators, as well as anti-panic lighting in case of mains power failure or emergencies, it is recommended to refer to European Norm EN 1838 (Lighting applications – Emergency lighting) [9].

Continuation of a Sport

For continuation of a sport, the lighting level shall be at least the Class III level specified for that sport (see Tables 2 and 3).

Broadcasting Continuity

The delaying or cancelation of a high-end sports event due to the loss of lighting is unacceptable. Where this is likely to cause a problem, e.g., at professional games and/or where television broadcasting is involved, then provision must be made to give an alternative solution.

Especially for major events broadcasted on TV, it is mandatory that, in the event of power failure, the continuity of the television broadcast shall be guaranteed. The chosen solution has to fit the specification and be relevant for the competition or event to be held in the stadium.

The actual requirement for an alternative power supply and switching method should be specified by sport's governing bodies and their contractual obligations towards the host broadcaster.

The main points to consider are the following:

- Time delay between switching from one power source to another, including the restrike time of lamps
- Lighting level to ensure broadcasting continuity

The highest standard would run the lighting via uninterruptible power supply (UPS) system with an alternative power supply available to supply 100 % of the system. This system will provide a complete system with no reduction in lighting quality.

For lower standards (lower continuity lighting level and/or time delay accepted), other alternatives are available depending on several factors like the quality of the local/national power network, investment costs, and maintenance costs.

In that case, many solutions can be defined, coupling several supply sources (electricity network, running generators, backup generators) and using hot restrike luminaires (or not). (High-intensity discharge lamps used for sports lighting applications will extinguish from even a momentary loss of power supply and will then need to cool down (10–15 min) before being able to start again. Hot restrike luminaires have the facility to immediately restart the lamp after a momentary power loss.)

Spectator Area Lighting

Much of the existing documents about sports lighting mainly specify lighting recommendation on the field of play without taking care of the spectator areas.

For the visual comfort of the spectators, the European norm EN 12193 recommends a minimum value of 10 lx in the spectator area.

In case of TV coverage it could happen that the broadcaster would specify a lighting level in the spectator area as a ratio of the field of play lighting level in order to suit his production style in Tables 2, 3, 4 and 5.

Cross-References

► [CIE Color-Rendering Index](#)

References

1. CEN Comité Européen de Normalisation (European Committee for Standardization) – Brussels
2. CIE Commission Internationale de l'éclairage (International Commission of Illumination) – Vienna
3. CIE 112: Glare evaluation system for use within outdoor sports and area lighting – CIE Central bureau, Vienna, Austria, 1994
4. GAISF General Association of International Federations – GAISF has been renamed Sport Accord in 2009 – Lausanne
5. GAISF guide to the artificial lighting of indoor and outdoor sports venues – 2006
6. NF EN12193 European standard – sports lighting – l'association Française de Normalisation, 11 rue Francis de Pressensé, 93571 La plaine Saint Denis, France, March 2008
7. IES Illuminating engineering Society – New York
8. IES Lighting Handbook, 10th edn. Reference and application – Illuminating Engineering Society of North America, 120 Wall Street, New York, USA, 2011
9. NF EN 1838 European standard – emergency lighting – l'association Française de Normalisation, 11 rue Francis de Pressensé, 93571 La plaine Saint Denis, France, 2013

Stain

► [Pigment, Ceramic](#)

Standard Daylight Illuminants

► [Daylighting](#)

Standard Illuminating and Viewing Conditions

► [Standard Measurement Geometries](#)

Standard Lamp

Cheng-Hsien Chen

Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan

Synonyms

[Reference standard lamp](#)

Definition

A standard lamp is used as a reference in photometric, radiometric, and colorimetric measurements for which the calibration is traceable to a primary photometric, radiometric, or colorimetric standard. According to the basic and general concepts and the associated terms of the International Vocabulary of Metrology, the measurement standard is routinely used to calibrate or verify measuring instruments or measuring systems. In color applications, the standard lamp is the measurement standard that transfers the quantities from the reference measurement standard in a metrological organization [1, 2].

Overview

Colorimeters and spectrometers need to be calibrated against a standard lamp with known standardized data. The International Commission on Illumination (CIE) standard illuminant A is intended to calibrate typical, domestic, tungsten-filament lamps. Its relative spectral power distribution is that of a Planckian radiator at a temperature of approximately 2,856 K according to the International Temperature Scale of 1990 (ITS-90) with $c_2 = 14,388 \mu\text{m}$. Also, the definition of CIE



Standard Lamp, Fig. 1 Various light-emitting diode (LED) transfer standards developed by national metrology institutions

source A is used to realize the CIE illuminant A. CIE standard illuminant A can be realized by CIE source A, defined as a gas-filled, tungsten-filament lamp operating at a correlated color temperature of 2,856 K. The standard lamp can be a CIE source A to transfer the quantities from the reference measurement standard [3].

To establish the traceability of a measurement system, a working standard lamp with the measurement data is assigned by the substitution procedure to compare the primary standard. A working standard lamp is usually calibrated with reference to a secondary standard lamp that is operated under similar conditions and has a certain certification procedure to verify the performance of quantities with a degree of uncertainty. In general, the working standard lamp is applied for the transfer of photometric and radiometric quantities, such as luminance, irradiance, luminous intensity, luminous flux, CIE chromaticity coordinates, etc. The stable and repeatable transfer standards with a low degree of uncertainty, which have similar characteristics to those of the test lamp, are desired for the measurement system

to transfer photometric and colorimetric units from a national metrology institution (NMI).

Some NMIs are now in charge of developing the light-emitting diode (LED) transfer standard to provide the LED industry with a measurement system that is traceable to SI units, as shown in Fig. 1. Also, the transfer standard carries the photometric and radiometric quantities, which are the alternative standards to the traditional gas-filled lamp.

There are three typical LED standard lamps that have been developed in the NMIs. The traditional LED package is the 5-mm diameter package, which has been used since the early 1960s. The LED die is mounted in the reflector cavity, which bonds two wires from the LED die and connects to the frame. There are many colors of LED packaged in transparent epoxy. Some white ones are mixed the phosphors in transparent epoxy lens. The typical LED package – a 5-mm lamp LED transfer standard – has been developed by the seasoning system and the correction model of temperature and photometric quantities has been applied [4]. Although LEDs have many advantages over conventional sources in becoming

a transfer standard, they generate a large amount of heat. The heat not only influences the optical characteristics of LEDs, but also alters the stability and repeatability. Therefore, it is important to design a thermal control system to overcome the temperature problem. The high-powered LED should have a suitable temperature control module to reduce the high temperature on the LED die. Therefore, some NMIs have developed a high-powered LED with a thermal control module for the LED standard lamp that has two light distributions for different applications [5]. We can use the goniophotometer measurement method to provide stable photometry and colorimetry quantities for producing the high-powered LED standard lamp [6].

In addition, some NMIs have developed controllable light distribution and have a lower angular correlated color temperature distribution (ACCTD) LED transfer standard for the various applications of LED measurement systems [7]. Also, NMIs provide the temporal and luminous standard lamp for the display measurement instrument [8].

Application (For Example, How to Use the Standard Lamp, Spectral Radiometer)

We can use the spectroradiometer to measure the spectral power distribution of the luminaire. The typical sensor of the spectroradiometer for the visible wavelength is silicon-based. The spectral response is not the same for the measurement range. Moreover, the spectral efficiency of the grating is not the same either. Therefore, we should use the standard lamp to calibrate the spectroradiometer using the replacement method. The standard lamp carries the spectral irradiance standard for the measurement of spectral range calibrated by the metrological institute. We use the spectroradiometer to measure the standard lamp, then correct the spectral response of the spectroradiometer. Furthermore, using the gas lamp (Xe, Ar, Ne, and Kr), which has distinct spectral lines, the spectral reference calibrates the spectroradiometer for wavelength correction. Therefore, we can use the calibrated spectroradiometer to measure the device under test (DUT).

Cross-References

- [Spectral Power Distribution](#)
- [Spectroradiometer](#)

References

1. International Commission on Illumination.: ILV: International Lighting Vocabulary. CIE S017/E: 2011, Vienna (2011)
2. The Joint Committee for Guides in Metrology.: International Commission on Illumination: International Vocabulary of Metrology – Basic and General Concepts and Associated Terms. JCGM 200:2012, 3rd edn. BIPM, Sevres Cedex, France (2012)
3. János, S. (ed.): Colorimetry: Understanding the CIE System. Wiley, Hoboken (2007)
4. Park, S.C., Kim, Y.W., Lee, D.H., Park, S.N.: Preparation of a standard light-emitting diode (LED) for photometric measurements by functional seasoning. *Metrologia* **43**, 299–305 (2006)
5. Gerloff, T., Lindemann, M., Shirokov, S., Taddeo, M., Pendsa, S., Sperling, A.: Development of a new high-power LED transfer standard. In: Proceedings of CIE 2012 Lighting Quality & Energy Efficiency, Hangzhou, China pp. 125–127. (2012).
6. Zama, T.: Developing a new photometric standard for light emitting diode (LED) – ensuring reliable photometric evaluation of LED. *AIST Today* **40**, 17 (2011)
7. Chen, C.H., Chang, Y.Y., Ting, Z.Y., Sun, C.C.: Design of a standard LED light source with extremely high optical and color stability. In: *Frontiers in Optics*, Orlando (2013)
8. Liu, W.C., Wu, G.N., Kuo, C.J., Chen, Y.L.: A light standard for FPD dynamic parameter. In: *Secretariat of Asia Display 2011 Conference*. Kanshan, China (2011)

Standard Measurement Geometries

Bor-Jiunn Wen

Department of Mechanical and Mechatronic Engineering, National Taiwan Ocean University, Keelung, Taiwan

Synonyms

[Standard illuminating and viewing conditions](#)

Definition

Reflectance and transmittance measurements are partly dependent upon the geometry of illumination and viewing. Accordingly, the International Commission on Illumination (CIE) has specified illuminating and viewing conditions, namely, $45^\circ:0^\circ\text{a}$, $45^\circ:0^\circ\text{c}$, $45^\circ:0^\circ\text{x}$, $0^\circ:45^\circ$, $\text{d}:0^\circ$, $0^\circ:\text{d}$, $\text{de}:8^\circ$, and $\text{di}:8^\circ$. With the $45^\circ:0^\circ$ geometry, the sample is illuminated at an angle of 45° from the normal to the sample surface and viewed normal to the sample surface or within 10° of the normal. There are three standard configurations of the geometry: $45^\circ:0^\circ\text{a}$, $45^\circ:0^\circ\text{c}$, and $45^\circ:0^\circ\text{x}$. The $45^\circ:0^\circ\text{a}$ has a 10° wide cone of light incident on the sample at 45° to the sample normal. The “a” in $45^\circ:0^\circ\text{a}$ indicated “annular,” meaning that the 10° cone continues uninterrupted all the way around the sample. $45^\circ:0^\circ\text{c}$ indicates circumferential illumination, which is similar to annular except that it is distributed in discrete locations around the ring. $45^\circ:0^\circ\text{x}$ does not have ring illumination, but instead uses a single beam at a single azimuth angle. The $0^\circ:45^\circ$ geometry is the reverse of $45^\circ:0^\circ$. The $0^\circ:\text{d}$ (normal: diffuse) geometry is where the sample is illuminated by a beam whose axis is at an angle not exceeding 10° from the normal to the sample surface. The reflected flux is collected by an integrating sphere and viewing is accomplished by a receiver being directed toward an inner wall of the sphere. In all geometries the viewing cones should not exceed 5° . The $\text{d}:0^\circ$ geometry is the reverse of $0^\circ:\text{d}$. The $\text{d}:8^\circ$ hemispherical geometries present the sample with diffuse illumination and approximately normal detection. Diffuse illumination is accomplished by the use of an integrating sphere. Light introduced into the sphere is sent through a series of reflections off the diffuse white sphere wall, quickly resulting in the diffuse illumination incident on the sample. Two standard configurations of diffuse geometry are $\text{de}:8^\circ$ and $\text{di}:8^\circ$. The former excludes the specular component and the latter includes it.

Example of Measurement Geometry

Diffuse and directional illumination geometries are used to determine the color performance of

the device under test (DUT). This article introduces the examples of measurement geometries for flat panel displays. In practice, configurations for implementing these illumination geometries are represented as follows [1–6].

Diffuse Illumination: Hemispherical Illumination

Use a light source with an integrating sphere, as a uniform diffuse light source, to illuminate the DUT. Place the light-measuring device (LMD) at the measurement port with an inclination angle of θ_D off the normal to the center of the DUT (Fig. 1). When θ_D is 8° , the measurement geometry is $\text{d}:8^\circ$. The $8^\circ:\text{d}$ geometry is the reverse of $\text{d}:8^\circ$. The $\text{d}:8^\circ$ and $8^\circ:\text{d}$ geometries are specified by the CIE as standard viewing/illuminating geometries. The measurement area covers no less than 500 pixels and the measurement field angle is $\leq 2^\circ$. For the measurement excluding specular reflection, open the cover of the specular port during the test. For the measurement including specular reflection, close the cover of the specular port during the test.

Directional Illumination

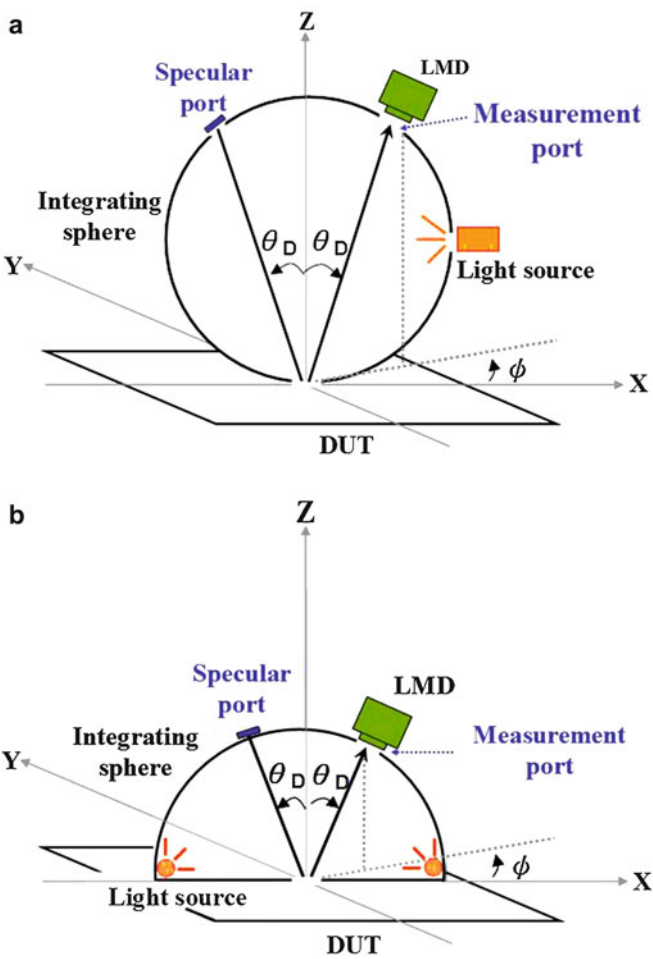
Ring-light Illumination

Use a ring-light source to provide uniform directional illumination of the DUT over all azimuth angles at an inclination angle of θ_S (recommended angle 45°). Place the LMD normal to the center of the DUT (Fig. 2). The measurement area covers no less than 500 pixels and the measurement field angle is $\leq 2^\circ$.

Collimated Light Illumination

Use a collimated light source to provide directional illumination of the DUT at an inclination angle of θ_S . Place the LMD at a viewing angle of θ_D (Fig. 3). When θ_S is 0° and θ_D is 45° , the measurement geometry is $0:45$. The $45:0$ geometry is the reverse of $0:45$. The $0:45$ and $45:0$ geometries are also specified by the CIE standard (viewing/illuminating geometries.) The measurement area covers no less than 500 pixels and the measurement field angle is $\leq 2^\circ$.

Standard Measurement Geometries, Fig. 1 (a) Type I and (b) Type II



Application 1

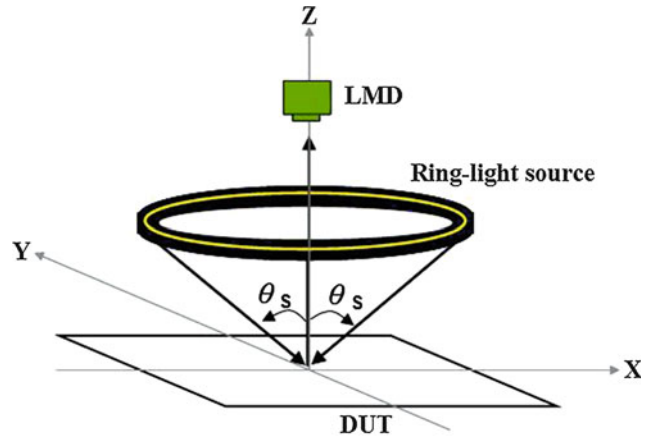
Users usually view liquid crystal displays (LCDs) in a bright environment, such as in an office or living room. Evaluating the contrast of display devices under ambient light is more practical than in a dark environment. The results of ambient contrast measurements are affected by different illuminance, viewing angles, etc. Therefore, the measurement geometry of the ambient illuminance is important for the measurement of ambient contrast for LCDs [7].

For diffuse illumination, a uniform diffuse-ambient light is provided by a light source with an integrating sphere to illuminate the screen of the display (Fig. 1). The integrating sphere has

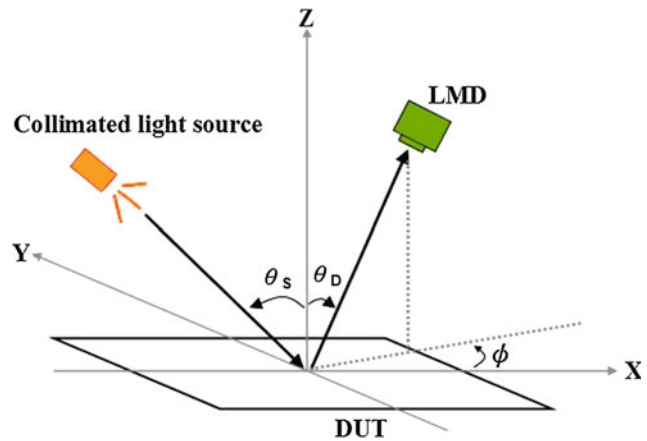
plural ports. The DUT is placed against the sampling port of the integrating sphere. The light measurement display (LMD) is arranged to view the screen through the measurement port. Only one measurement port is in use during the measurement. Measurement ports not in use have covers on them. The inner surfaces of covers are coated with standard white, which should have the same reflection properties as the inner surface of the sphere. In addition, the directional illumination can also be provided to illuminate the screen of the display for the measurement of the ambient contrast of LCDs.

After setting up the measurement geometry, the measurement procedure for the ambient contrast of LCDs using diffuse illumination is as follows.

Standard Measurement Geometries, Fig. 2 Ring-light illumination measurement geometry



Standard Measurement Geometries, Fig. 3 Collimated illumination measurement geometry



For the measurement with the specular port closed, first warm up the DUT for at least 40 min. Second, measure the luminance of full-screen white (L_W)_{dark} and full-screen black (L_K)_{dark} under darkroom conditions. Third, set up the integrating sphere and turn on its lamp (warm up for at least 30 min). Fourth, measure the illuminance on the surface of the DUT. The illuminance is adjusted to the required levels (200 and 500 lx are typical for a bright living room and an office environment respectively) by controlling the intensity of the light source. It is important to keep the correlated color temperature (CCT) of the light source stable during the illuminance adjustment. Finally, measure the luminance of full-screen white (L_W) and full-screen black (L_K). The illuminance should be kept stable during the measurement process.

For measurement with the specular port open the procedure is identical to that with it closed, except for the cover of the specular port being taken out at the end of the third step.

After measurement, the ambient contrast of the LCD can be directly calculated by using Eq. 1, where C_A is the symbol of ambient contrast, L_W and L_K are the luminance of full-screen white and full-screen black respectively.

$$C_A = \frac{L_W}{L_K} \quad (1)$$

The ambient contrast can also be calculated by using the reflectance. The reflectance of full-screen white ρ_W and full-screen black ρ_K is derived using Eq. 2, and the ambient contrast C_A is calculated using Eq. 3. It is recommended that

the reflectance is measured at the illuminance E_{01} not smaller than 1000 lx. In Eq. 3, E_0 is the desired illuminance?

$$\begin{cases} \rho_W = [L_W - (L_W)_{\text{dark}}] \times \frac{\pi}{E_{01}} \\ \rho_K = [L_K - (L_K)_{\text{dark}}] \times \frac{\pi}{E_{01}} \end{cases} \quad (2)$$

$$C_A = \frac{(L_W)_{\text{dark}} + \frac{\rho_W}{\pi} E_0}{(L_K)_{\text{dark}} + \frac{\rho_K}{\pi} E_0} \quad (3)$$

Application 2

For solid state lighting luminaries (SSL), the standards are described in the framework of the ENERGY STAR program. One of the standards therein is the IES LM-79 on testing and measuring methods. Hereunder, the measurement of the luminous flux and luminous colors using integrating sphere photometers and their related requirements are described. For measurement geometries, the LM-79 describes two different structures: 2π and 4π . In the 2π version, light sources are operated outside the sphere. The luminous flux emitted from the light source enters the sphere through a measuring aperture. The application range is limited to spot-lights of a maximum 2π (hemisphere). In the 4π version, the light sources are operated within the integrating sphere with which all directional radiation can be measured. In the LM-79, it is recommended to measure 2π and 4π sources in the inner part of the sphere. The use of 2π sphere geometry is only recommended in exceptional cases, i.e., when required because of the dimensions and form of the light source. Therefore, further key concerns of the LM-79 with the integrating sphere are:

- Reflectivity of the sphere coating
- Field of view function of the detectors
- Baffle implementation
- Ability to measure temperature

It is very important to check the key concerns for SSL measurements before using the LM-79 standard.

Cross-References

- [Correlated Color Temperatures](#)
- [Illuminance Meter](#)
- [Luminance Meter](#)

References

1. Kuhn, R.G.: Color: An Introduction to Practice and Principles. 3rd edn. John Wiley & Sons, Inc. (2012)
2. Fairchild, M.D.: Color Appearance Models. 3rd edn. John Wiley & Sons, Inc. (2013)
3. Wyszecki, G., Stiles, W.S.: Color Science: Concepts and Methods, Quantitative Data and Formulae. 2nd edn. John Wiley & Sons, Inc. (2000)
4. Hunt, R.G.W.: The Reproduction of Colour, 6th edn. John Wiley & Sons, Inc. (2004)
5. SID IDMS Information Display Measurements Standard, v1.03, 1 June 2012
6. SEMI D68-0512: Test Methods for Optical Properties of Electronic Paper Displays (2012)
7. SEMI D56-0310: Measurement Method for Ambient Contrast of Liquid Crystal Displays (2010)

Stanziola, Ralph A

Michael H. Brill
Datacolor, Lawrenceville, NJ, USA



Ralph A. Stanziola (1931–2007), a consultant and teacher of color technology, was born in Philadelphia, received his B.S. degree in Chemistry from

the Philadelphia College of Textiles and Science (now Philadelphia University), and resided most of his life in New Jersey. In his early years, at the Research and Technical Service for the Dyes Department of the American Cyanamid Company, Ralph learned color technology from such pioneers as Orrin W. Pineo and Edward I. Stearns. He next served for 9 years as Technical Representative and General Sales Manager for the Davidson & Hemmendinger Company, and later became a sales manager for the Kollmorgen Corporation Color Systems division, which had acquired Davidson and Hemmendinger.

In 1970, Ralph cofounded Applied Color Systems, Inc. where he eventually became the Executive Vice President and Technical Director. In May, 1985 Ralph founded Industrial Color Technology, a consulting company under whose auspices he solved many industrial problems involving color control. For the rest of his life, Ralph remained a consultant for ACS (later Datacolor), instructed their Color Technology seminars, and helped to create the first video training series on color technology.

Ralph held five US patents [1–5] and authored a large number of technical papers and presentations. He developed the Color Curve System for color communication, and helped the Glad Products Company to develop the now famous “Glad Difference,” *Yellow and Blue Make Green*® seal. Less familiar will be his development (with Bob Swain of Chroma Corp.) of a method of using a colored sample to test the wear of a metal piece. The test consists of extruding a molded plastic piece made of component materials of two different colors, and monitoring the color of the mixture. Earlier, Ralph developed the first colorant-dispenser system driven by computer color matching (1979) and patented a Maxwell-disk-based color simulator (1980). He also codeveloped an asymmetrical color tolerancing system using artificial intelligence algorithms (1991), a photonic visual color simulator using LEDs (late 1990s), and a new version of the Color Rule for testing observers and light booths for metamerism (2006).

Ralph joined the Inter Society Color Council (ISCC) in 1962 and was an active member for the

rest of his life. He cochaired several ISCC meetings: the Annual Meeting in Princeton in 1992 that celebrated the 25th anniversary of the AIC; the 1994 joint meeting with the Detroit Color Council (DCC); and the 2003 Williamsburg Conference on Industrial Color Problems held at Philadelphia University. He was instrumental in setting up the Education Interest Group, and was the first to chair the Industrial & Applied Color Interest Group. In recognition of these contributions, Ralph received the 2004 Nickerson Service Award, and in 2005 became an Honorary Member of the ISCC. He also served as *Color Research and Application's* Special Editor for Industrial Applications for 10 years.

In 1995, the Technical Association of the Pulp and Paper Industry presented Ralph with their Finest Faculty Award. As a member of the Federation of Societies for Coatings Technology, he received the Armin J. Bruning Award for his outstanding contribution to the science of color in the field of coatings technology. He was also a member of the American Association of Textile Chemists and Colorists and the Detroit Colour Council.

Ralph's numerous lectures included computer color matching seminars at the Rensselaer Color Measurement Laboratory (Rensselaer Polytechnic Institute) and at the Munsell Color Science Laboratory (Rochester Institute of Technology). The many personal anecdotes he told in his color courses showed a wealth of experience not available in any book. Some urged him to write a book, but he preferred the humbler route of interacting directly with the world and with other people.

References

1. Worn, P.R., Stanziola, R.A.: DR Hall Reflected color simulator. US Patent 4,310,314 (1982)
2. Stanziola, R.: Objective color notation system. US Patent 5,012,431 (1991)
3. Swain, R.D., Stanziola, R.A.: Method of determining wear. US Patent 6,306,319 (2001)
4. Swain, R.D., Stanziola, R.A.: Method of determining wear. US Patent 6,797,204 (2004)
5. Riddle, G.H.N., Reitmeier, G., Steinmetz, C., Stanziola, R.A., Burstyn, H.C.: Color display device. US Patent 6,985,163 (2006)

Stereo and Anaglyph Images

Krzysztof Templin

Computer Graphics, Max Planck Institute for Informatics, Saarbrücken, Germany

Synonyms

[Stereo pairs](#); [Stereograms](#)

Definition

Stereo images are pairs of conventional images depicting a scene as seen from two different viewpoints, which correspond to the eye positions of an observer in the scene. When each image of a stereo pair is presented separately to the corresponding eye, the impression of three-dimensionality known as stereopsis arises. Stereo images find its use in arts, entertainment, and scientific visualization.

An anaglyph image is a method of presentation of a stereo image, in which the images composing the stereo pair are colored with different hues and superimposed. The resulting image is viewed using glasses with matching color filters, so that each eye sees only one of the images and stereopsis can appear. Usually, red and cyan filters are used.

Stereo and Anaglyph Images

Stereo Images

Since human eyes have different vantage points from which they view the world, a point in space may be projected to different positions in the two eyes. The difference between the points of projection is called “binocular disparity” and it is a basis to the impression of depth called “stereopsis” [1]. Stereopsis can be artificially triggered when two images, in which corresponding points have been displaced according to their depth, are presented dichoptically, i.e., one to the left eye and the other to the right eye. The images (called

half images) are fused into a percept of a single, three-dimensional image. An important tool for studying stereopsis is a random dot stereogram, where the depth impression is evoked solely by the binocular disparity in isolation from any other sources of depth information [2].

The simplest method of stereo image presentation is free fusion. The half images are shown side by side, and the observer is expected to simply converge or diverge their eyes, until the images overlap. The invention of the first device for watching stereograms is attributed to Wheatstone and dates back to 1838 [1, 3]. The modern presentation methods range from various setups making use of special glasses (color, polarizing or interference filters, time multiplexing) to glasses-free, autostereoscopic displays. Each method has its advantages and limitations, important factors being usability, viewing comfort, and faithful color reproduction. A serious flaw of a stereo system is cross talk, i.e., leakage of the signal intended for one eye to the other eye. Cross talk hinders fusion and in extreme cases can spoil the 3D effect completely.

An important issue in stereoscopy is viewing comfort [4]. Since stereo images are presented in a single plane, they require decoupling the focusing and vergence mechanisms of the observer’s eyes. This restricts the range of depths possible to reproduce to a certain volume around the display, because excessive mismatch between the focusing and vergence points causes significant eye-strain in the observer [5]. Since more distant objects require less refocusing, the comfort zone can be greatly extended by increasing the distance to the display or by using lenses which move back the focusing point. Other sources of discomfort include, but are not limited to, various types of inconsistencies between the half images, e.g., in vertical position, focus, or color balance.

When capturing a stereo image, the distance between the viewpoints does not need to equal the human interocular distance (ca. 64 mm). One situation when this is useful is capturing very distant or very close objects. For the former, there would be hardly any stereopsis if natural separation was used, while for the latter the resulting image could become impossible to

Stereo and Anaglyph

Images, Fig. 1 A pair of paper anaglyph glasses with red and cyan filters



fuse. Altering the separation between the viewpoints may be also necessary to maintain plausible proportions of the objects or to fit the content within the comfort zone of the display. Sometimes, the background is captured using different viewpoint separation than the foreground and the outcome is composed into one image [6], or the disparities are adjusted throughout the whole range of depths using image processing techniques [7].

Anaglyph Images

Anaglyph images, first described in 1853 [8], are probably the most widespread presentation method of stereo images. The half images are colored with different colors and superimposed using additive color mixing. Each filter of the glasses lets in the light from the corresponding half image while blocking the light from the other one. Various color combinations are used, such as green-magenta or red-blue; however, the most common one is red-cyan, with the left filter transmitting red light and the right filter transmitting blue and green. Apart from the glasses, this method does not require any special equipment, such as a dedicated printer or display, which makes it very inexpensive and convenient (see Fig. 1). However, since the colors of the image need to be modified, this method suffers from poor color reproduction. Moreover, each half image is manipulated in a different way, so the

outcome may cause retinal rivalry. Slight blurriness due to chromatic aberration may be visible in one of the eyes, and some high-end glasses include corrective lenses to counteract this effect. Finally, since the filters often do not separate the two half images perfectly, cross talk may appear.

In the most basic form of anaglyph 3D for RGB displays, one filter transmits one of the primaries and the other filter the remaining two primaries. Hence, three types of anaglyph images are possible: red-cyan, green-magenta, and blue-yellow. The final image is formed by selecting one of the RGB channels from one of the half images and the remaining two channels from the other (see Fig. 2). The yellow filter transmits most of the perceived colors; however, there is a significant luminance imbalance between the yellow and blue filters. The green-magenta pair is well balanced in terms of luminance but gives rise to noticeable color sheen over a wide range of hues [9]. The colors can be traded off for viewing comfort by converting one or both half images to grayscale or by ignoring one of the channels completely [10].

Several algorithms have been proposed to enhance the quality of the anaglyph image given a pair of filters. If the spectral absorption curves of the glasses and the density functions of the display primaries are known, the creation of an anaglyph image as close as possible to the desired stereo pair can be posed as an optimization problem [11].



Stereo and Anaglyph Images, Fig. 2 A grayscale stereo image presented as an anaglyph image; from left to right: red-cyan, green-magenta, yellow-blue

By performing a few color matching tasks, anaglyph image generation process can be adjusted to reduce ghosting [12]. Proprietary systems exist, such as ColorCode 3-D or Inficolor 3D, that combine advanced image processing techniques with careful filter pair selection.

Further improvement in the color reproduction can be achieved by using wavelength multiplexing. Each filter transmits all three primary colors but in slightly different wavelength sub-bands for each eye. The half images have different gamuts, one leaning toward green and the other toward magenta, but the visual system compensates for the difference, and after a short time it is not visible anymore. Even better color fidelity can be obtained with systems based on light polarization. The disadvantage of both systems in comparison to anaglyph 3D is that in addition to the glasses, they require special displays [6, 10].

Cross-References

- [Chromostereopsis](#)
- [Trichromacy](#)

References

1. Howard, I.P., Rogers, B.J.: *Perceiving in Depth*. Oxford University Press, New York (2012)
2. Julesz, B.: *Foundations of Cyclopean Perception*. MIT Press, Cambridge, MA (2006)
3. Wheatstone, C.: Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Phil. Trans. R. Soc. London* **128**, 371–94 (1838)
4. Lambooi, M., Fortuin, M., Heynderickx, I., IJsselstein, W.: Visual discomfort and visual fatigue of stereoscopic displays: a review. *J. Imaging Sci. Technol.* **53**(3), 1–12 (2009)
5. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.* **8**(3), 33.1–30 (2008)
6. Mendiburu, B.: *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press, Boston (2012)
7. Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., Gross, M.: Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph.* **29**(4) (Proc. ACM SIGGRAPH), 75:1–10 (2010)
8. Rollmann, W.: Zwei neue stereoskopische Methoden. *Ann. Phys.* **166**, 186–187 (1853)
9. Sorensen, S.E.B., Hansen, P.S., Sorensen, N.L.: Method for recording and viewing stereoscopic images in color using multichrome filters. US Patent No. 6,687,003, 2004
10. Hainich, R.R., Bimber, O.: *Displays: Fundamentals and Applications*. A K Peters/CRC Press, Boca Raton (2011)
11. Dubois, E.: A projection method to generate anaglyph stereo images. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 1661–1664 (2001)
12. Sanftmann, H., Weiskopf, D.: Anaglyph stereo without ghosting. *Comput. Graphics Forum* **30**(4) (Proc. EGSR), 1251–1259 (2011)

Stereo Pairs

- [Stereo and Anaglyph Images](#)

Stereograms

► Stereo and Anaglyph Images

Stevens, Stanley Smith

Mark D. Fairchild
College of Science, Rochester Institute of
Technology, Rochester, NY, USA



S.S. Stevens was an American experimental psychologist perhaps best known in the world of color science for introducing the psychophysical power law and for collecting data on brightness perception as a function of adaptation [2]. He was even more well known in the general field of experimental psychology and psychophysics for many accomplishments in acoustical psychophysics, his important publications, and founding of the Psycho-Acoustic Laboratory at Harvard University.

Among Stevens' literary contributions to the field are the extensive, and immensely praised, *Handbook of Experimental Psychology* (1951) [3] that provides much information that remains useful more than half a century later and the text *Psychophysics* (1975) [4] that was published posthumously after final editing by his wife. It, too, remains a

useful reference on the topic. Among the topics described is the hierarchy of mathematical scales (nominal, ordinal, interval, and ratio) in psychophysics, something else developed by Stevens.

Stevens was an undergraduate student at the University of Utah and then transferred to Stanford where he completed his degree in an undetermined field since the variety of courses he took was so extensive. He then was accepted at the Harvard Medical School, but chose to enroll in Harvard's School of Education to avoid the medical school's \$50 fee and organic chemistry requirement and still have access to Harvard's resources. At Harvard he was transformed by a course on perception taught by E.G. Boring, became Boring's unpaid research assistant, and earned a Ph.D. in Philosophy by the end of his second year there. After some more studies in physiology and physics, he was appointed an instructor in the Psychology Department, where he remained until his death. In 1962, Harvard University agreed to appoint Stevens as Professor of Psychophysics in honor of the field founded by G.T. Fechner.

One of Stevens' most notable papers was entitled "To honor Fechner and repeal his law," published in *Science* in 1961 [1]. In it, Stevens summarized psychophysical scaling data for a variety of perceptual stimuli (e.g., brightness of light, loudness of sound, hotness of heat, pain of electric shock, etc.) and illustrated that they all could not possibly follow the logarithmic relationship between stimulus and perception predicted by Fechner's Law. Instead, all of his data could be well described by power functions with various exponents, which depend on the perception being scaled. This new relationship is known as the psychophysical Power Law, or the Stevens' Power Law. One technique used to establish the Power Law that Stevens developed is known as cross-modal matching. In these experiments, observers match perceived magnitude of one stimulus (e.g., brightness of light) with another (e.g., loudness of sound).

Stevens' work on brightness is well known in color science and led to the description of a color appearance phenomenon as the Stevens' Effect. This work, published in 1963 by Stevens (no relation) and Stevens, examined the effects of light and dark adaptation on the perception of

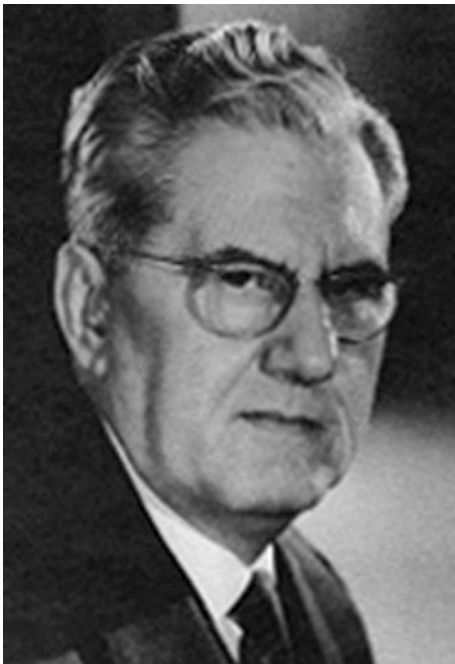
brightness. Scales of brightness were obtained using magnitude estimations (a Steven's hall-mark) and showed that the contrast of brightness increased with increasing adapting luminance (exponent in the power increases). So, as Stevens showed us, when the level of lighting increases, dark colors look darker and light colors lighter.

References

1. Stevens, S.S.: To honor Fechner and repeal his law. *Science* **133**, 80–86 (1961)
2. Stevens, J.C., Stevens, S.S.: Brightness functions: effects of adaptation. *J. Opt. Soc. Am.* **53**, 375–385 (1963)
3. Stevens, S.S. (ed.): *Handbook of Experimental Psychology*. Wiley, New York (1951)
4. Stevens, S.S.: *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. Wiley, New York (1975)

Stiles, Walter Stanley

Robert W. G. Hunt
Department of Colour Science, University of
Leeds, Leeds, UK



Dr. W.S. Stiles, OBE, FRS, trained in physics at University College, London, and in mathematics at St. John's College, Cambridge. He spent almost all his working life at the National Physical Laboratory of Great Britain in Teddington.

He was created OBE (Officer of the British Empire) in 1946 for his wartime work on visibility, visual search, and the defensive use of dazzle. He was elected to the (British) Royal Society in 1957 and was awarded the Tillyer Medal of the Optical Society of America in 1965. He served as General Secretary of the Commission Internationale de l'Éclairage from 1928 to 1931, was Chairman of the Colour Group of Great Britain from 1949 to 1951, and was President of the (British) Illuminating Engineering Society from 1960 to 1961.

His earlier work with B. H. Crawford introduced the concept of veiling glare, a subject of particular importance in street lighting, and other applications of illuminating engineering.

His name is perhaps best known in the Stiles-Crawford effect, the reduction in sensitivity of the retina as the angle of incidence of the light becomes increasingly different from normal.

In the late 1950s, he spent much time and effort in constructing a visual colorimeter for redetermining the color-matching functions of the normal observer. This work, involving measurements on over 50 observers, provided the major basis for what became the CIE 1964 supplementary standard colorimetric observer for field sizes of 10° ; the study also included experiments with a 2° field which confirmed the validity of the CIE 1931 standard colorimetric observer apart from the well-known deficiency in sensitivity at the extreme short-wave end of the spectrum. This work also addressed the phenomenon of rod intrusion: the fact that in fields of 10° size, some input from the rods can be added to that of the cones.

Much of his later work was devoted to the study of increment thresholds: the magnitude of the just-noticeable amount of a stimulus of one wavelength when superimposed on a uniform field of another wavelength. From these studies he identified a series of basic visual mechanisms which he identified as π mechanisms, the

significance of which have been a matter for considerable discussion.

Toward the end of his life, he collaborated with Gunter Wyszecki in the production in 1967 of *Color Science: Concepts and Methods, Quantitative Data and Formulas*, published by Wiley, with a second edition in 1982 [1]. This work still provides an enormous amount of useful data and information on color science.

He was dedicated to traditional psychophysical experimental methods and regarded it as very necessary to avoid arguments founded on introspective descriptions of sensations, which he regarded as notoriously difficult to interpret correctly. He therefore had no contact with the area of magnitude estimation pioneered by S.S. Stevens and developed subsequently by other workers in color science.

As a lecturer he was clear and authoritative with a commanding demeanor. Although a man of great intellect and prodigious experience in color science, he was nevertheless always willing to help others in the field who approached him for help. His interests included mathematics, reading, and painting.

References

1. Wyszecki, G., Stiles, W.: *Color Science, Concepts and Methods, Quantitative Data and Formulae*, 2nd edn, Wiley, New York (2000)

Further Reading

- Stiles, W.S.: The directional sensitivity of the retina and the spectral sensitivities of the rods and cones. *Proc. Roy. Soc. Lond.* **B127**, 64 (1939)
- Stiles, W.S., Burch, J.M.: Interim report to the Commission Internationale de l'Eclairage, Zurich (1955), on the National Physical Laboratory's investigation of colour matching. *Optica. Acta.* **2**, 168 (1955)
- Stiles, W.S., Crawford, B.H.: The luminous efficiency of rays entering the eye pupil at different points. *Proc. R. Soc. Lond B* **112**, 428–450 (1933)

Street Lighting

- [Road Lighting](#)

Stroop Effect

Colin M. MacLeod

Department of Psychology, University of Waterloo, Waterloo, ON, Canada

Synonyms

[Color-word interference](#); [Stroop interference](#)

Definition

The Stroop effect is one of the best known phenomena in all of cognitive science and indeed in psychology more broadly. It is also one of the most long standing, having been reported by John Ridley Stroop in the published version of his dissertation in 1935 [1]. In its basic form, the task is to name the color in which a word is printed, ignoring the word itself. When the word is a color word printed in a mismatched ink color, this is very difficult to do and results in slow, error-prone responding. To illustrate, consider **GREEN**: To say “red” to the ink color is difficult relative to a variety of comparison or control conditions such as naming the color of **XXXXX** or the word **TABLE** or even the word **RED**. The performance cost in the mismatch condition – usually referred to as the incongruent condition – relative to the controls is called the Stroop effect or Stroop interference. Figure 1 provides an illustration of the phenomenon.

Background

Since the very beginning of experimental psychology, it has been clear that words are faster to read than objects or their properties are to name. In his dissertation in 1886, Cattell [2] even went so far as to suggest that word reading is automatic due to extensive practice, introducing the concept of automaticity to cognitive science. Automatic processes can be thought of as unintentional, uncontrolled, unconscious, and fast [3]. Under

1	2	3	4
red	red	xxx	yellow
yellow	yellow	xxxxxx	blue
green	green	xxxxx	red
red	red	xxx	yellow
blue	blue	xxxx	red
green	green	xxxxx	blue
yellow	yellow	xxxxxx	green
blue	blue	xxxx	green
green	green	xxxxx	red
red	red	xxx	blue
blue	blue	xxxx	yellow
yellow	yellow	xxxxxx	green

Stroop Effect, Fig. 1 An illustration of the Stroop effect. In columns 1 and 2, the task is to read each word in the column aloud, ignoring its print color, and to do so as quickly as possible. This represents Stroop’s (1935) first experiment, where he found little difference in reading time between the experimental condition (column 2) and the control condition (column 1). In columns 3 and 4, the task is to name the print color of each word in the column aloud, ignoring the word itself, again doing so as quickly as possible. This represents Stroop’s (1935) second experiment, where he found dramatic interference in that color naming time in the experimental condition (column 4) was much slower and more error prone than was the case in the control condition (column 3). Note that to make the conditions comparable, the same responses are required for every item in every column

the automaticity account, people cannot comply with an instruction not to read because reading cannot be “turned off”; hence it is guaranteed that incompatible words will cause interference when attempting to name their print colors.

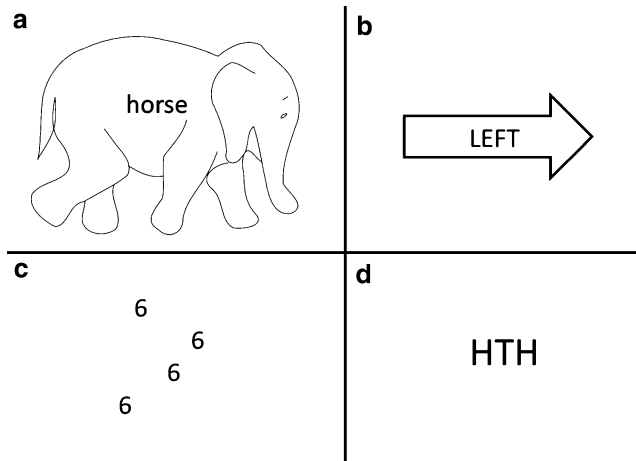
From early on, the other prevalent explanation of Stroop interference was the relative speed of processing account, which in its simplest form argued that faster processes can affect slower processes but not vice versa [4]. Thus, because words are read faster than colors can be named, interference results when the task is to name the colors and ignore the words. This also fits nicely with Stroop’s other finding – that there was no “reverse Stroop” interference when the task was to read the words and ignore the colors: Reading performance for incongruently colored words was

equivalent to that for words printed in standard black ink. Stroop’s results are highly replicable, as MacLeod [5] showed over a half century later.

Since Stroop’s landmark study, many hundreds of studies have sought to understand this superficially simple phenomenon, and many more have used his method to explore key aspects of attention, learning, memory, reading, language, and other cognitive skills [5]. More recently, the Stroop task has also been extended to investigate neural mechanisms [6] and clinical disorders [7], among other issues. Interest in Stroop’s method shows no signs of abating; indeed, it is one of the rare phenomena/tasks where interest seems to be growing rather than diminishing with the passage of time.

The Gradient of Interference

An intriguing feature of the Stroop literature, though, is that there was virtually no follow-up to his work for about 30 years; it is only in the 1960s that research on this phenomenon resumed and then with a vengeance. The simplest explanation of this is that the advent of computer-controlled experiments, and especially the resulting ability to time individual trial stimuli, opened up a rich new realm of investigation for which the Stroop task was ideally suited. But the study that actually relaunched research on color-word interference was reported by Klein in 1964 [8] using multiple-stimulus cards in much the same format as Stroop had used. Klein sought to understand what aspects of the words interfered with naming colors. To do so, he incorporated several new conditions. In addition to the baseline “colors-alone” card, there were six interference cards. As always, the standard incongruent condition – using words incongruent with the print colors – showed large interference. When color words that were not the names of the print colors were substituted, interference was cut in half, implicating a large role for response set in interference. Interference then declined more slowly across color-related words (e.g., *lemon*, *sky*); common, unassociated words (e.g., *put*, *heart*); and rare words (e.g., *sol*, *abjure*) and fell



Stroop Effect, Fig. 2 Illustrations of some interference tasks derived from the Stroop “template.” (a) The picture-word task, where naming the picture while trying to ignore the word results in interference (e.g., compared to a picture containing a row of Xs); (b) the “directional Stroop” task, where stating the direction in which the *arrow* is pointing suffers interference from the incompatible word (e.g., compared to no word or a row of Xs); (c) the “counting Stroop”

task, where counting the number of digits suffers interference from the identity of the digits (e.g., compared to using letters in place of the digits); and (d) the flanker task, where identifying the center letter suffers interference from peripheral (flanker) letters that sometimes also appear as central targets (e.g., compared to flankers that are never central targets)

to its smallest – but still reliable – level for unpronounceable nonsense syllables (e.g., *hjh*, *gsxrq*). Clearly, despite the instruction to ignore the word, subjects cannot do so, and the greater the relevance of the word, the greater is the resulting interference with color naming. MacLeod [5] reviews the many studies that have taken up where Klein left off.

Variants of the Classic Stroop Task

In essence, Stroop’s paradigm provides a template for studying interference, and investigators have often mined that template to create Stroop-like tasks suited to their particular research purposes. Figure 2 illustrates some of the many alternate versions in the literature. The best known is the picture-word interference task [9; A in Fig. 2], in which a conflicting word is embedded in a picture. As with the classic Stroop task, interference is largely unidirectional: Naming the picture shows interference from the word, but reading the word is hardly influenced by the picture. Other common variants include the directional version [10; B in

Fig. 2], in which again reading the word is quite unhampered by the mismatched arrow, but there is substantial interference from the embedded word when identifying the direction of the arrow, and the digit version [11; C in Fig. 2], in which counting the number of digits is impaired when the digits themselves are incompatible with their numerosity. Further afield, but clearly related, is the flanker task [12, D in Fig. 2], wherein identification of the central letter is impaired by flankers that sometimes also serve as central targets (relative to flankers that are never targets). The list goes on, and MacLeod [5] reviews some of the more common illustrations, although more have appeared in the ensuing quarter century since that review.

Evidence Against the Speed of Processing Account

Stroop himself interpreted the interference phenomenon that now bears his name as follows: “it seems reasonable to conclude that the difference in speed in reading names of colors and in naming

colors may be satisfactorily accounted for by the difference in training in the two activities.” Indeed, the third experiment in his paper actually manipulated training and provided support for this explanation, and more recent studies investigating training have confirmed the importance of experience. Notably, MacLeod and Dunbar [13] trained subjects over multiple days to apply color names to initially unfamiliar white shapes. They then were able to track the growth of interference in naming the print colors of incongruent shapes (e.g., the shape called “red” printed in blue) with extended practice on naming just the shapes in white. As shape naming became more practiced (more automatic), shapes produced more interference with naming colors, and print colors produced less interference with naming shapes.

One striking element to the MacLeod and Dunbar pattern was that there was a point in shape training where shape interference in color naming coexisted equally with color interference in shape naming. Such a result is entirely incompatible with a speed of processing account in that the shapes, although still slower to name than colors, nevertheless interfered with color naming. Two converging results were reported around the same time. First, Glaser and Glaser [14] separated the word and color, allowing one dimension to be presented at varying lags before the other. Most critically, even when the color preceded the word by a time sufficient to allow processing of the color before processing of the word, there was still no “reverse Stroop” interference: The color never interfered with the word. Second, Dunbar and MacLeod [15] transformed the words by rotating them in various ways, such as upside down and backwards. This dramatically slowed word reading time, such that it became slower than color naming time. Yet when transformed words were presented in incongruent colors, they still caused as much interference as did upright words. If relative speed of processing were all that mattered for interference, then interference should have been apparent in the Glaser and Glaser studies when color was sped up but not in the MacLeod and Dunbar studies when words were slowed down: In fact, the converse was true.

Although compellingly intuitive and widely accepted [4], then, the speed of processing account does not provide a sufficient explanation of Stroop interference.

Alternative Explanations of Stroop Interference

So how *are* we to explain this apparently powerful and simple effect? In the past 25 years, three major explanations have emerged. The first of these was Cohen, Dunbar, and McClelland’s parallel distributed processing, or connectionist, model, proposed in 1991 [16]. At its core, their theory is a strength theory, designed as it was to capture the training data reported by MacLeod and Dunbar [13]. Processing pathways gain strength with practice, and relative strength determines likelihood and degree of interference. Thus, given our extensive experience with reading, color-word pathways ordinarily are much more strongly connected to color name responses than are color pathways. When Cohen, Dunbar, and McClelland trained the model in various ways, they were able to reproduce many of the major empirical results in the Stroop literature identified by MacLeod [5].

In 2003, Melara and Algom [17], coming from a fundamental perception perspective, proposed that two factors underlie Stroop interference: dimensional imbalance and dimensional uncertainty. Dimensional imbalance reflects how correlated the two dimensions of a stimulus are and how surprising a stimulus is and determines the ease of recovery of a stimulus representation from memory. Dimensional uncertainty reflects how salient a stimulus is, notably how likely or unlikely it is in the context of other (recently presented) stimuli. Together, these two factors determine the success of attentional selection by focusing on salient, surprising, and/or correlated information contained within each dimension and across the two dimensions of a Stroop stimulus. Each influences excitation of targets and inhibition of distractors. Stroop interference occurs both because there is more uncertainty in the colors than in the words and because the words are more salient than the colors. Their experiments

manipulating these two factors were able to produce many of the major empirical results.

Also in 2003, Roelofs [18] proposed his model of Stroop interference, a model situated in an already implemented model of word production (WEAVER++) from the psycholinguistic literature. This also can be viewed as a two-factor model, with processing interactions occurring in the system that carries out language production, modulated by a supervisory attentional system that maintains task control. Roelofs posited that different architectures underlie color naming and word reading, with color naming, because it is conceptually driven, requiring an extra step due to colors not being directly connected to their names, unlike words. His experiments derived from this model also reproduced many of the major results in the Stroop literature.

The Big Picture

Anyone who has tried doing the Stroop experiment themselves knows that the interference caused by an incongruent word in naming its print color is powerful. As soon as we can read, we start to show this interference. Although we think of colors as being processed with ease, the existence of Stroop interference is an indication that there is computation involved in processing color names, a computation that can be disrupted. Some of this interference in the standard Stroop paradigm is response interference (i.e., the word and the color name are different responses within the small set), which reveals little about color processing. But the fact that color words not in the response set and words simply related to colors also produce interference demonstrates the semantic contribution to the Stroop effect.

This remarkable phenomenon has been used in recent years for many purposes, including to help to identify the function of certain brain regions, notably the anterior cingulate cortex and dorsolateral prefrontal cortex, both of which are active when resolving conflict. In the context of the Stroop task, the dorsolateral prefrontal cortex appears to be involved in the relevant executive functions, particularly in maintaining response set

(to name the color, not to read the word), whereas the anterior cingulate cortex plays a central role in selecting the appropriate response and evaluating its accuracy [19, 20].

One thing is clear after 80 years: This simple and robust paradigm and the interference phenomenon that it demonstrates continue to provide valuable insights into the operation of cognitive processes. It would have been hard to imagine in 1935 that merely pitting a few colors against their names could provide such an important tool.

References

1. Stroop, J.R.: Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643–662 (1935)
2. Cattell, J.M.: The time it takes to see and name objects. *Mind* **11**, 63–65 (1886)
3. Moors, A., De Houwer, J.: Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* **132**, 297–326 (2006)
4. Dyer, F.N.: The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Mem. Cogn.* **1**, 106–120 (1973)
5. MacLeod, C.M.: Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* **109**, 163–203 (1991)
6. MacLeod, C.M., MacDonald, P.A.: Inter-dimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends Cogn. Sci.* **4**, 383–391 (2000)
7. Williams, J.M.G., Mathews, A., MacLeod, C.: The emotional Stroop task and psychopathology. *Psychol. Bull.* **120**, 3–24 (1996)
8. Klein, G.S.: Semantic power measured through the interference of words with color-naming. *Am. J. Psychol.* **77**, 576–588 (1964)
9. Rosinski, R.R., Golinkoff, R.M., Kukish, K.S.: Automatic semantic processing in a picture-word interference task. *Child Dev.* **46**, 247–253 (1975)
10. Shor, R.E., Hatch, R.E., Hudson, L.J., Landrigan, D. T., Shaffer, H.J.: Effect of practice on a Stroop-like spatial directions task. *J. Exp. Psychol.* **94**, 168–172 (1972)
11. Windes, J.D.: Reaction time for numerical coding and naming of numerals. *J. Exp. Psychol.* **78**, 318–322 (1968)
12. Eriksen, B.A., Eriksen, C.W.: Effects of noise letters upon identification of a target letter in a non- search task. *Percept. Psychophys.* **16**, 143–149 (1974)
13. MacLeod, C.M., Dunbar, K.: Training and Stroop-like interference: evidence for a continuum of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 126–135 (1988)

14. Glaser, M.O., Glaser, W.R.: Time course analysis of the Stroop phenomenon. *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 875–894 (1982)
15. Dunbar, K.N., MacLeod, C.M.: A horse race of a different color: Stroop interference patterns with transformed words. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 622–639 (1984)
16. Cohen, J.D., Dunbar, K., McClelland, J.L.: On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* **97**, 332–361 (1990)
17. Melara, R.D., Algom, D.: Driven by information: a tectonic theory of Stroop effects. *Psychol. Rev.* **110**, 422–471 (2003)
18. Roelofs, A.: Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychol. Rev.* **110**, 88–125 (2003)
19. Banich, M.T., et al.: fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *J. Cogn. Neurosci.* **12**, 988–1000 (2000)
20. Milham, M.P., Banich, M.T., Claus, E.D., Cohen, N.J.: Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. *Neuroimage* **18**, 483–493 (2003)

Stroop Interference

- [Stroop Effect](#)

Structure of Color Preference

- [Comparative \(Cross-Cultural\) Color Preference and Its Structure](#)

Structure Property Relationships

- [Colorant, Environmental Aspects](#)

Subjective Colors

- [Fechner's Colors and Behnam's Top](#)

Subsampling

- [Interpolation of Spectral Data](#)

Surface Plasmon Polariton

- [Surface Plasmons](#)

Surface Plasmon Resonance

- [Surface Plasmons](#)

Surface Plasmons

Li-Lin Tay

Measurement Science and Standards, National Research Council Canada, Ottawa, ON, Canada

Synonyms

[Plasmonics](#); [Surface plasmon polariton](#); [Surface plasmon resonance](#)

Definitions

Surface plasmons are collective excitations of conduction electrons at the interface between a metal and dielectric that are stimulated by electromagnetic radiation.

A *surface plasmon polariton* (SPP) occurs when surface plasmons interact strongly with electromagnetic radiation.

Surface plasmon resonance (SPR) is another synonym for SPP and refers to the coherent (resonant) oscillation of the surface conduction electrons excited by electromagnetic radiation.

A *localized surface plasmon resonance* (LSPR) is a nonpropagating SPP confined to nanostructured surfaces.

The term *plasmonics* is often used to cover the various phenomena sustained by SPP and SPR types of light-matter interactions.

Overview

The electron charge density and its electromagnetic fields propagate as a surface wave along a metal–dielectric interface. The electromagnetic field intensity decays exponentially away from the interface. Material systems with free or almost free electrons will sustain SPP excitation. More generally, an SPP occurs when the real part of the dielectric function of the metal is negative and its magnitude is larger than that of the complex dielectric function. This also means that an SPP can only exist for transverse magnetic polarization since no surface modes exist at the metal surface for transverse electric polarization. For common noble metals, such as gold and aluminum, this condition is met in the visible–NIR wavelength region at the air–metal interface. As an SPP propagates along the metal–dielectric boundary, it is very sensitive to changes that modulate this interface. SPP can be excited by electrons or photons. A planar surface of a good conductor that sustains SPP with given frequency ω and parallel momentum obeys the following dispersion relation [1]:

$$k_{\parallel}^2 = \left(\frac{\omega}{c}\right)^2 \text{Re} \left[\frac{\varepsilon_0 \varepsilon}{\varepsilon_0 + \varepsilon} \right]$$

where k_{\parallel} is the parallel component of the wavevector, $\varepsilon(\omega)$ is the dielectric function of the conductor, c is the speed of light in vacuum, and ε_0 is the dielectric constant of the ambient medium (air or vacuum), also known as the permittivity.

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}$$

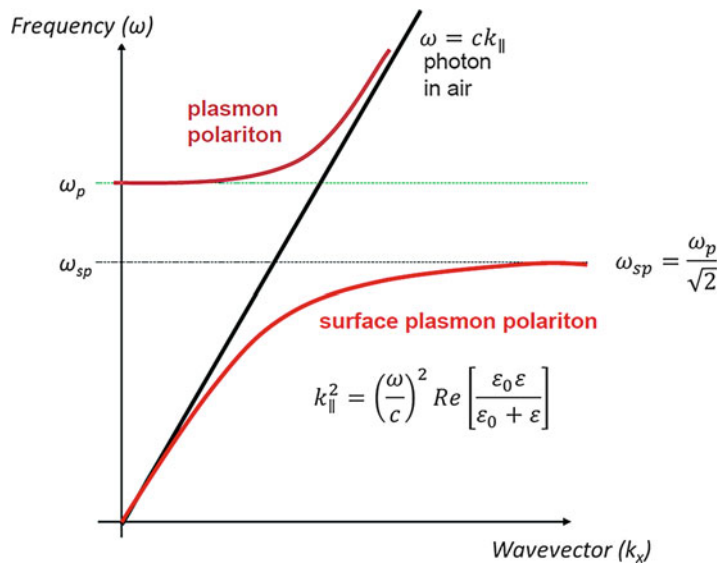
where bulk plasma frequency ω_p is

$$\omega_p = \sqrt{\frac{ne^2}{\epsilon m_{\text{eff}}}}$$

with n and e being the free electron density and electron charge, respectively, and m_{eff} the electron effective mass; ϵ here is the vacuum permittivity. Figure 1 depicts the SPP dispersion relation. At lower wavenumber (mid-IR or lower frequency), SPP exhibits photon characteristics, and the waves extend over many wavelengths into the dielectric space. In this regime, SPP behaves like a grazing-incidence light field. As the wave vector k increases, the dispersion relation tends toward the asymptotic limit of the surface plasma

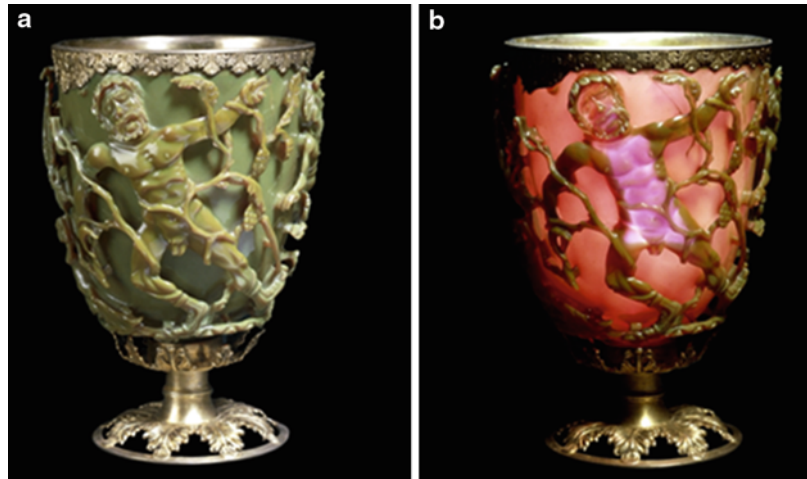
Surface Plasmons,

Fig. 1 Dispersion relation of plasmon polariton and surface plasmon polariton



Surface Plasmons,

Fig. 2 The Lycurgus cup from the Roman Empire, fourth century AD. The cup displays a *green colour* (a) when viewed with reflected light and a *brilliant red colour* when viewed with transmitted light (b)



frequency, ω_{sp} . Due to their bound nature, the SPP dispersion relation lies entirely to the right of the light line (photon in air, with $\epsilon_0 = 1$). In this regime, the out-of-plane component of the SPP wave vector (k_{\perp}) is purely imaginary and decays evanescently away from the metal surface. Thus, special phase-matching techniques such as grating or prism coupling are required to excite SPP.

For an incident electromagnetic wave to excite the SPP, both the frequency and parallel momentum must be conserved. This condition cannot be achieved with air (or vacuum) as the ambient medium at the interface due to wave vector mismatch. For the same reason, an SPP from a smooth flat surface does not radiate into the ambient. It is confined to the metal surface with its amplitude decaying exponentially away from the metal surface. SPP propagates on the metal surface and eventually dissipates its energy as heat through Ohmic loss and electron core interaction. For a relatively absorbing metal such as aluminum, SPP propagation lengths can reach $2\ \mu\text{m}$ when excited with $500\ \text{nm}$ (visible, green) radiation. For a low-loss metal such as silver where there is less damping, excitation at the same wavelength increases the SPP propagation length to $20\ \mu\text{m}$. Moving to a longer wavelength where there is less confinement, for example an excitation wavelength of $1.55\ \mu\text{m}$ typically used for optical fiber-based telecommunication, the SPP propagation length can increase into the mm range. The propagation length of SPP sets the

upper size limit for the SPP-based photonic circuitry. However, its strength lies in its ability to transmit information in subdiffraction limited circuits as shown in the section on examples and applications below.

Examples and Applications

Gold, silver, copper, and aluminum are among the many metals that exhibit strong plasmonic responses. These metals have long been known to have optical properties distinct from standard dielectrics. In fact, long before scientists began to investigate the optical properties of metal nanostructures, plasmonic nanoparticles were used to produce brilliant colors in ancient artifacts. One of the best-known examples is the famous Lycurgus cup (Fig. 2) that is on exhibit at the British Museum. This ancient Roman relic was constructed with one of the few dichroic glasses that appear opaque and green in reflected light and a brilliant translucent red in transmitted light. The difference in color is due to the distinct plasmonic properties of colloidal gold and silver which are present in trace amounts in the glass of the cup. Studies from a small fragment of the relic have revealed trace amounts of $20\text{--}30\ \text{nm}$ colloidal gold and silver nanoparticles which give the glass its unique dichroic properties. Similar inclusions of colloidal metal nanoparticles have been found in brilliantly colored stained glass windows in churches.

Surface plasmon resonance (SPR), also known as SPS, is a popular analytical tool for the study of biomolecular interaction, the determination of kinetic reaction rates, and the evaluation of binding affinity of measured molecular binding events.

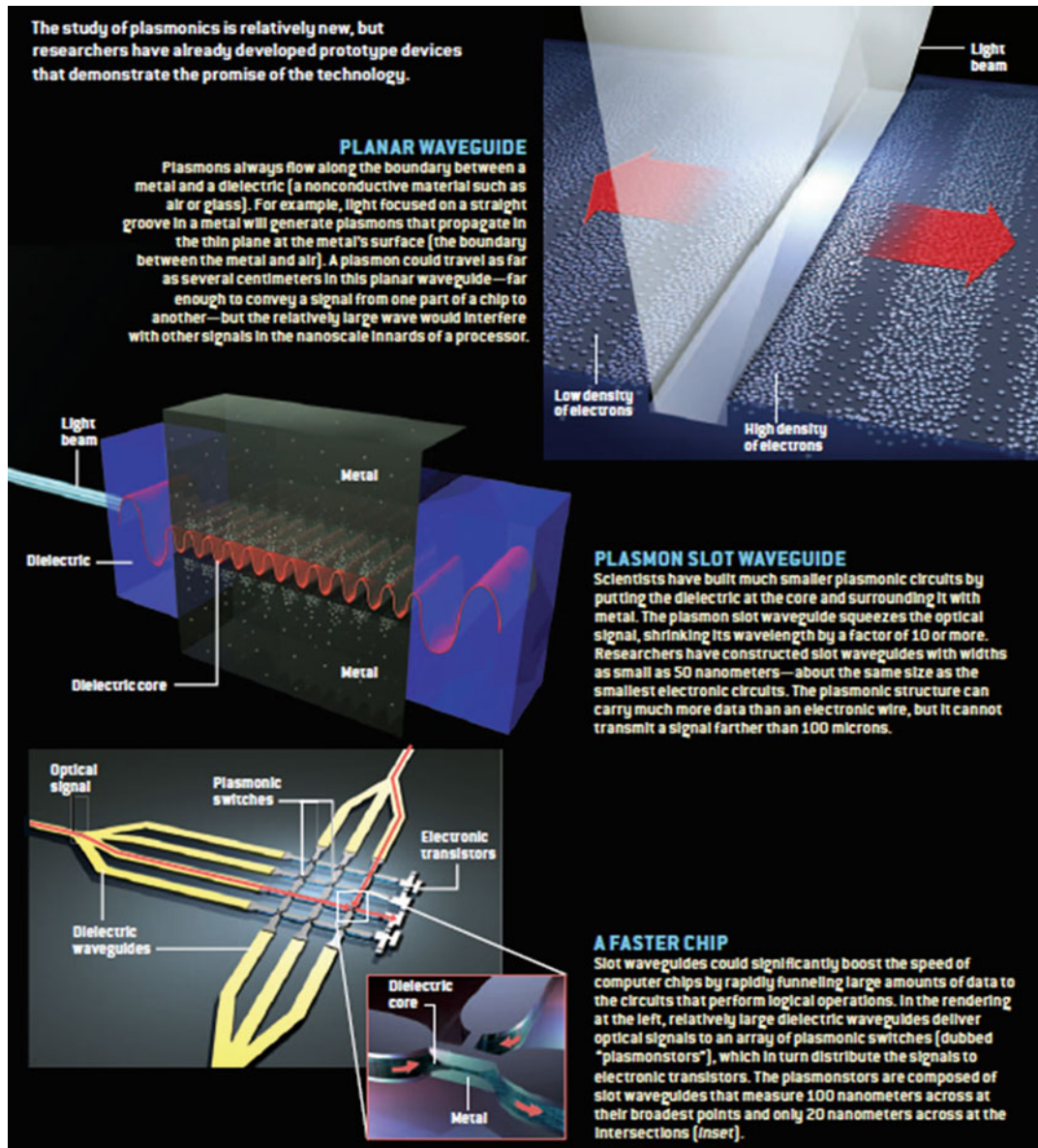
More recently, applications of SPP involve the design and fabrication of metal nanostructures to tune their plasmonic properties. This requires delicate chemical synthesis with protocols designed to disperse metallic nanostructures tagged with desired molecular recognition units or nanofabrication of specifically structured metal nanostructures to enable the confinement and/or propagation of SPP to transmit information. While it may seem impractical to transmit light signals with metal structures, the confinement of SPP to the metallic nanostructures enables the propagation of information in subdiffraction limited circuitry. Because planar plasmonic structures act as waveguides confining SPP along the metal–dielectric boundary, they can be used to route signals on a chip. Metal circuits can be defined lithographically and mass produced as miniature-sized plasmonic devices enabling very high density and fast chips as shown in Fig. 3 below [2].

Another striking application of SPP comes in the form of the localized surface plasmon resonance (LSPR) which is essentially SPP that is confined to small metal nanostructures, most often in the form of metal nanoparticle aggregates. Excitation of an LSPR can be tuned by material composition, size, shape and coupling between nanostructured entities. Similar to SPR, LSPR is widely used in biosensors for the study of small molecule interactions. Localization of electromagnetic fields at certain regions of the nanostructure due to resonant excitation of SPP is the foundation of many forms of surface enhanced spectroscopy and colorimetry based bioassay.

The detailed mechanism of LSPR is described in the section on surface enhanced Raman spectroscopy below, as it is the dominant mechanism that enables this form of spectroscopy. Here we will simply describe the photothermal effect and its theranostic application. Gold nanoparticles (NPs) are one of the most studied nanomaterials for applications in the biomedical sciences. Gold

NPs are potential drug carriers, photothermal treatment agents, radiosensitizers, optical imaging contrast agents and have shown great promise for cancer therapy. There are a few different types of gold NPs: solid Au NPs, Au nanoshells, and Au nanorods that are commonly used in biomedicine. These different types of NPs are differentiated by their characteristic localized surface plasmon resonances. Among them, Au nanoshell and Au nanorods are known to exhibit broad LSPR tunability, covering much of the visible and NIR spectral window. Synthetic protocols and surface functionalization methods are well publicized through peer reviewed scientific literatures [3]. The nanoparticles are typically stabilized and guided to the targeted tissues by coating them with polyethylene glycol and antibodies. As mentioned in the earlier discussion, surface plasmon resonance interaction with metal nanostructures is a lossy process with most of its energy being dissipated as heat through Ohmic loss and electron-core interaction. The heat generated through optical excitation of LSPR provides gold NPs with a unique photophysical property to act as an anti-cancer agent. Tumor-specific ligands (e.g., transferrin, folic acid, antibodies) have been attached to the surface of gold nanoparticles for specific targeting of and delivery to tumor sites.

There are currently a few different versions of gold-NP based cancer treatment therapy that are undergoing various phases of FDA (Food and Drug Administration) approval [4]. One version of such anti-cancer treatment relies on the plasmonic properties of a gold nanoshell. The gold nanoshell can be engineered to resonantly interact with different wavelengths of the light used to excite the LSPR. This is done by tuning the shell thickness of the gold coating over the silica nanosphere. With the LSPR of Au nanoshell tuned to a near-infrared absorption window, it can be delivered to cancer sites with the help of targeting agents as mentioned above. Upon arriving at the tumor site, a near infrared laser is used to excite the Au nanoshell and trigger the photothermal effect. This turns the optical radiation into intense but localized heating sources thereby enabling selective and rapid tumor



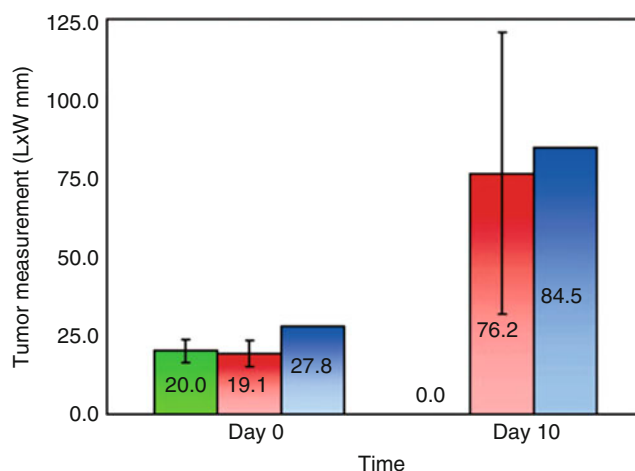
Surface Plasmons, Fig. 3 Examples of SPP waveguides and an example of plasmonic circuitry (Sources: Scientific America) [2]

destruction with minimal damage to surrounding tissues. Photothermal therapy of head and neck cancer is currently under phase I of FDA approval. Phase I of clinical trial is aimed to evaluate safety, determine safe dosage range and identify side-effects of new drugs. Photothermal therapy may also be used in combination with

standard chemotherapy and radiation. This concept was pioneered by Professors Naomi Halas and Jennifer West of Rice University in 2004 [5]. Figure 4 below shows therapeutic efficacy and tumor growth over a period of time. The figure showed the original mean tumor size at the beginning of the treatment (day 0) and day

Surface Plasmons,

Fig. 4 Animal study showing gold nanoshell efficacy in cancer treatments [5]. Mean tumor size at the beginning of the treatment and on day 10 for the treatment group (*green*), control group (*red*), and sham treatment (*blue*, no nanoshell but treated with same dose of NIR irradiation)



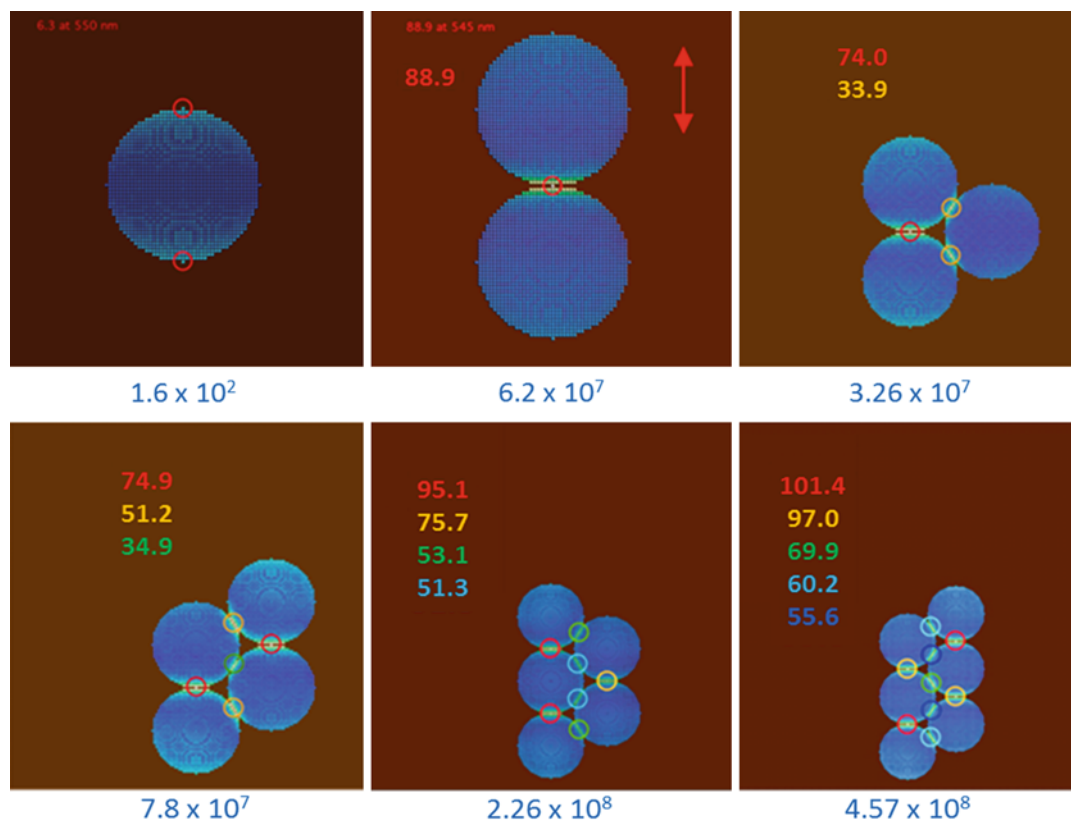
10 of the photothermal therapy. The treatment group (with Au nanoshell injected to tumor site) is shown in green, control group (receives no treatments) in red and sham treatment group (which receives same laser treatment as treatment group with no nanoparticle) in blue.

Surface Enhanced Spectroscopies

Raman spectroscopy is a vibrational technique which provides a wealth of information on the molecular species under investigation. Its many fields of application include identification and structural characterization of semiconductor materials, authentication of artifacts and legal documents, quality control in pharmaceutical manufacturing and the development of new biophysical tools for disease diagnosis and health monitoring. A major disadvantage of Raman spectroscopy is the inherently small scattering cross-sections of the Raman process, which are typically 12–14 orders of magnitude smaller than fluorescence cross sections. This was why the observation of an unexpectedly large Raman signal from pyridine adsorbed on roughened Ag electrode, reported in 1974 by Fleischmann et al., surprised researchers in the field [6]. The Ag electrode had been roughened through repeated oxidation-reduction cycles in an attempt to develop an in situ monitoring tool that is chemically specific [6]. The authors attributed the large

pyridine Raman signal to the increase in number of molecules adsorbed on the increased surface area of the roughened Ag electrode. In 1977, Van Duyne et al. recognized that the large observed intensity could not be accounted for by the increase in the number of scatterers and proposed an electric field enhancement model [7]. Ultimately, it was the 1978 analysis by Moskovits [8] which attributed the predominant contributing factor of SERS (surface Enhanced Raman Scattering) to be the electromagnetic excitation of localized surface plasmon resonance (SPR) sustained by properly designed metal nanostructures. His interpretation cements the electromagnetic enhancement theory for SERS and paved the way for the development of the Plasmonics field. It is important to also note that LSPR is not the sole contributing factor to the inordinate large Raman intensity observed from molecules associated with metal nanostructures. An alternate theory proposed by Albrecht and Creighton in their 1977 report proposed that resonance Raman scattering from molecular electronic states, broadened by their interaction with metal surface, could be responsible for the increased intensity [9]. It is understood that both factors contribute to the observed SERS signature with the LSPR being the dominating enhancement mechanism.

The plasmonic interpretation of SERS led to the understanding and prediction of many characteristic SERS effects. One of them is the fourth

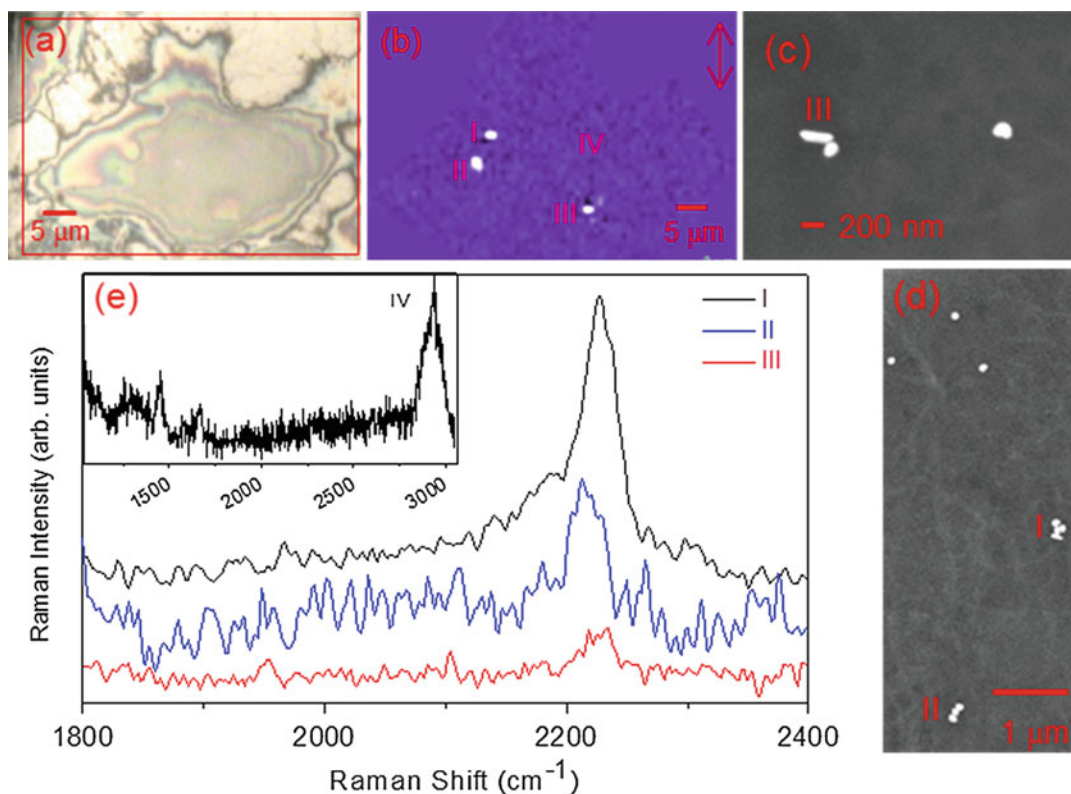


Surface Plasmons, Fig. 5 Electric field distribution in different 2D geometric gold nanoclusters

power dependency of the SERS enhancement factor (G) upon the local incident electric field ($|E_0|$); $G = |E_0|^4$. Another well-known characteristic of the plasmonic nature of SERS is the coupling of LSPR between adjacent nanostructures which results in extraordinary local field strength thereby enabling single molecule detection from tightly coupled nanostructures. Figure 5 below shows the false coloured electric field distribution of a single gold nanoparticle and a variety of coupled gold nanostructures.

LSPR is extremely sensitive to composition, morphology, geometry and degrees of aggregation and dielectric properties of the ambient environment as partly demonstrated in Fig. 5 [10]. All these factors, while affecting the overall strength of the observed SERS signal, also enable flexibility in the design of nanostructures to sustain stable SERS.

SERS is a powerful spectroscopic effect that leverages the enormous electromagnetic field enhancement caused by the excitation of an intense, sharp and localized surface plasmon resonance of metal nanostructures. The effect can provide rich structural and chemical information for molecules adsorbed on or bonded to SERS active nanometer-sized metal structures. Recent advances in nanotechnology and nanofabrication techniques have enabled researchers to design SERS active colloidal NP systems with optimized electromagnetic enhancements engendering a large variety of SERS-based bioanalytical and bioimaging applications. The advantages of SERS over standard fluorescent labeling techniques lie not only in its multiplexing capability and superior photostability, but also in the potential multimodal functionality provided by the other intrinsic properties of the metal NPs, such



Surface Plasmons, Fig. 6 (a) Optical image of a HeLa cell labeled with nitrile-reporter and target-specific molecules. (b) SERS intensity map of the nitrile vibrational band with three SERS hot spots I, II and III identified. (c) and (d) SEM (scanning electron microscopy) image of the

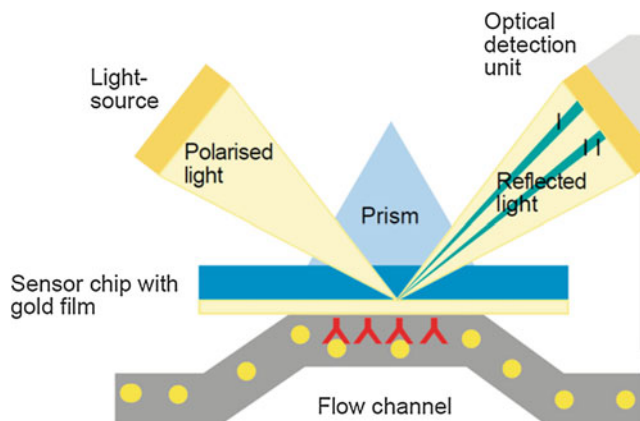
corresponding area showed SERS activity. (e) SERS spectra of the three spots identified in the intensity map. *Inset* of (e) shows the native cell Raman spectrum. *Inset* of (e) shows the cell Raman spectrum taken from the nucleus region labeled as spot IV

as the photothermal response which enables its therapeutic role in the treatment of cancer as outlined earlier [11]. In addition to its cancer fighting ability, the intense and unique signal from SERS nanoparticles also enables diagnostic imaging. SERS labels rival that of fluorescent contrasting agents and can be adapted to most of the fluorescent detection schemes. Figure 6 below shows an example SERS imaging of cell surface receptors in mammalian cells [12]. By the similar targeting mechanism, nanoparticles can be guided towards diseased tissue sites and optical spectroscopy-microscopy can be used as a diagnostic means to pin point the disease sites.

In addition to biomedical applications, SERS is also very popular in security applications and trace chemical detection owing to its extraordinary sensitivity and information rich spectra.

SPR Sensors

The surface plasmon resonance (SPR) enables many forms of enhanced spectroscopy but the simplest form of SPR sensor relies on very simple SPR reflectivity to detect molecular absorption, binding dynamics such as association and dissociation constants of a multitude of molecules from simple organic molecules, polymers, and DNA to complex proteins and antibodies. In its most common form, a flat thin Au film is deposited or index matched to a prism as shown in Fig. 7 below. Users measure the angle of minimum reflection which corresponds to the maximum absorption of the Au film. The reflection angle changes as film thickness changes. This is due to the adsorbed molecules modulating the effective index of refraction of the film/adsorbate system and hence

Surface Plasmons,**Fig. 7** Typical SPR sensor setup

the plasmon resonance condition. The SPR sensors on the Au film can be pre-functionalized with a capture molecule. When the molecule of interest is present in the test solution, binding events will occur at the Au-water interface, changing the SPP condition of the Au thin film. This change is then observed as a change in resonance angle. Conversely, when the molecule of interest detaches from the Au surfaces, it also changes SPP resonance condition. SPR sensors are often used to measure association and dissociation constants in addition to the simple binding assay in biological systems. This measurement can also be achieved by varying the incident wavelength to match the effective index of refraction. However, such a setup requires a high accuracy tunable laser source and so poses more technical challenges as compared to the simple angle measurement SPR systems.

The same SPR principle can be applied to localized surface plasmon resonances made available by confining the SPR to a small nanostructure. It is possible to monitor the change in the LSPR resonance condition using a very simple dark-field microscopy setup coupled to a simple desk-top spectrometer and achieve very high density parallel detection of molecular adsorption and desorption events.

Summary

With recent advancements in nanotechnology and materials science, SPP based devices are finding

more applications. In addition to its broad applications in biomedical agents, sensors and enhanced spectroscopies, plasmonic devices can now be found in narcotic sensors, solar energy converters, field portable autoclaves and an expanding array of microelectronic devices. While most of these devices are still in early development stages, there is no doubt that the impact of SPP will expand as we gain better understanding of its physics and a better handle on nanofabrication techniques.

References

1. Moskovits, M.: Surface enhanced spectroscopy. *Rev. Mod. Phys.* **57**, 42 (1985)
2. Atwater, H.: The promise of plasmonics. *Sci. Am.* **17**(13), 56–63 (2007)
3. Haslett, T.L., Tay, L., Moskovits, M.: Can surface-enhanced Raman scattering serve as a channel for strong optical pumping? *J. Chem. Phys.* **113**(4), 1641–1646 (2000)
4. Pillai, G.: Nanomedicines for cancer therapy: an update of FDA approved and those under various stages of development. *SOJ Pharm. Pharm. Sci.* **1**(2), 13 (2014)
5. O'Neal, D.P., Hirsch, L.R., Halas, N.J., Payne, J.D., West, J.L.: Photo-thermal tumor ablation in mice using near infrared-absorbing nanoparticles. *Cancer Lett. (Amsterdam, Neth.)* **209**(2), 171–176 (2004)
6. Fleischman, M., Hendra, P.J., McQuillan, A.J.: Raman spectra of pyridine adsorbed at a silver electrode. *Chem. Phys. Lett.* **26**(2), 4 (1974)
7. Jeanmaire, D.L., Duynne, R.P.V.: Surface raman spectroelectrochemistry: Part I. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode. *J. Electroanal. Chem.* **84**(1), 1 (1977)

8. Moskovits, M.: Surface roughness and the enhanced intensity of Raman scattering by molecules adsorbed on metals. *J. Chem. Phys.* **69**(9), 4159 (1978)
9. Albrecht, M.G., Creighton, J.A.: Anomalously intense Raman spectra of pyridine at a silver electrode. *J. Am. Chem. Soc.* **99**(15), 5215 (1977)
10. Tay, L.-L., Hulse, J.: Surface-enhanced Raman and optical scattering in coupled plasmonic nanoclusters. *J. Mod. Opt.* **60**(14), 1107–1114 (2013)
11. Hirsch, L.R., Stafford, R.J., Bankson, J.A., Sershen, S. R., Rivera, B., Price, R.E., Hazle, J.D., Halas, N.J., West, J.L.: Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance. *Proc. Natl. Acad. Sci. U. S. A.* **100**(23), 13549–13554 (2003)
12. Hu, Q.Y., Tay, L.L., Noestheden, M., Pezacki, J.P.: Mammalian cell surface imaging with nitrile-functionalized nanoprobe: biophysical characterization of aggregation and polarization anisotropy in SERS imaging. *J. Am. Chem. Soc.* **129**(1), 14–15 (2007)

T

Tactile Painting and Haute Couture

- [Art and Fashion Color Design](#)

Temporal Collective Color Preferences

- [Color Trends](#)

Tendencies

- [Color Trends](#)

Tesselations

- [Mosaics](#)

Tesserae

- [Mosaics](#)

Tetrachromatic Vision

Gabriele Jordan¹ and John D. Mollon²

¹Institute of Neuroscience, Newcastle University,
Newcastle upon Tyne, UK

²Department of Experimental Psychology,
Cambridge University, Cambridge, UK

Synonyms

[Four-dimensional color vision](#)

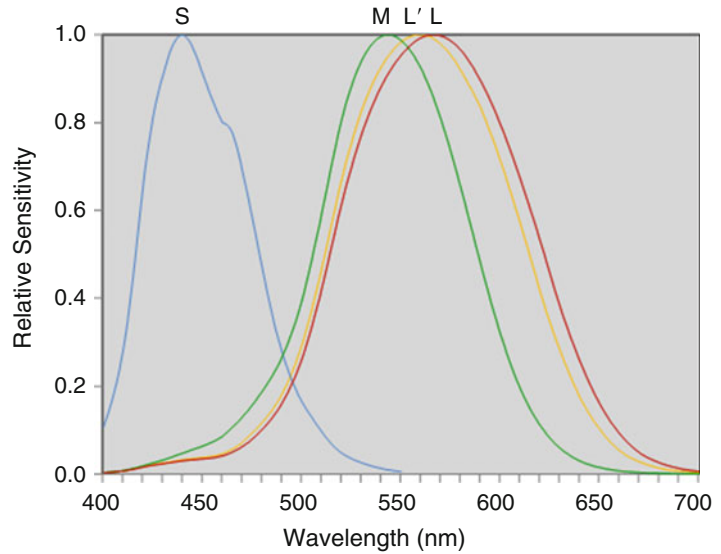
Definition

Tetrachromatic color vision here refers to human color vision that relies on the presence of four types of retinal cone photopigment whose signals are processed independently. By analogy with the classical, psychophysical definition of human trichromacy [1], a tetrachromatic observer will require four primaries in a color-matching experiment to match any other color. In comparative studies of color vision, the term tetrachromacy usually refers to the presence of four distinct types of photoreceptor.

A distinction has been made between strong and weak tetrachromacy [2]. In both cases, there are four types of cone in the retina, but only the former denotes behavioral tetrachromacy, where the cortex gains independent access to a fourth

Tetrachromatic Vision,

Fig. 1 Normalized spectral sensitivity curves for the S, M, L', and L cones of a hypothetical carrier of deuteranomaly (cDa)



signal allowing the individual to perceive colors along a dimension denied to color-normal people.

Conceptual Background

Normal human color vision is trichromatic in that three variables or primary lights are needed to match any given color [1]. Since 1802 [3], it has been thought that the basis of trichromacy is the presence in the retina of three types of “sensitive filament,” now commonly known as cone photoreceptors. According to the wavelength of maximum sensitivity, these cones are referred to as S (short-wave sensitive), M (middle-wave sensitive), and L (long-wave sensitive) cones.

To explain the existence of additional types of cone with different spectral sensitivities (M' or L'), one needs to consider the genetics of the M and L cone photopigments. Nathans et al. [4] sequenced the M and L protein genes that are located in a tandem arrangement on the X chromosome and found a 98 % homology of their DNA. This has been taken as evidence for an evolutionarily recent duplication of an ancestral gene, followed by some degree of divergence of the sequences of the two original copies. Highly similar genes tend to misalign during meiosis and recombine in different ways. The most common

outcome of this process is a gene product that consists of part L and part M gene. Importantly, such a hybrid gene will express a photopigment with a spectral sensitivity intermediate to the normal L and M cone photopigments ([5]; for review see [6]).

Some 6 % of men have such a hybrid gene on their single X chromosome and, as a consequence, will be classified as anomalous trichromats. Specifically, 5 % of these men are deuteranomalous (Da), and their color vision relies on inputs from S, L, and L' cones, while another 1 % of men are protanomalous (Pa) having S, M, and M' cones in their retinæ. Of interest here are the first-degree female relatives (mothers, daughters) of anomalous trichromats. These women are heterozygous for either deutan or protan anomaly and are candidates for tetrachromacy: On one of their X chromosomes, they have the genes for the normal L and M cone photopigments, but on the other they carry a hybrid gene that encodes a spectrally shifted cone photopigment. Figure 1 shows the four normalized spectral sensitivity curves for the S, M, L', and L cones of a hypothetical carrier of deuteranomaly (cDa).

A seminal paper by Mollon and colleagues [7] provides a useful analogy to human tetrachromacy. They studied variations of color vision in a basically dichromatic species of New

World primates, the squirrel monkey. They discovered in the population pool a range of cone pigments with different spectral sensitivities and, more importantly, behavioral trichromacy in a subset of females. Though no genetic analyses were available at the time, the authors hypothesized correctly that the origin of these animals' good color discrimination was genetic and afforded to them by heterozygosity at the X-linked locus.

X-Chromosome Inactivation and the Retinal Cone Mosaic

In females, during embryonic development, one of their two X chromosomes is silenced [8]. This process is random with respect to the parental origin of the X chromosome, but once inactivation has taken place in a given cell, it is preserved in all daughter cells. Crucially for tetrachromacy, the different types of expressed photopigment will be segregated in different cone cells so that the retina of a carrier for, say, deuteranomaly will be a mosaic of normal and anomalous cones (S, M, L', and L), and the size of the retinal cone patches will depend on the time of onset of X inactivation and the migration of cones in the developing retina. The analogous case exists for carriers of protanomaly (cPa) though they are much rarer in the population.

What Are the Necessary Conditions for Tetrachromacy?

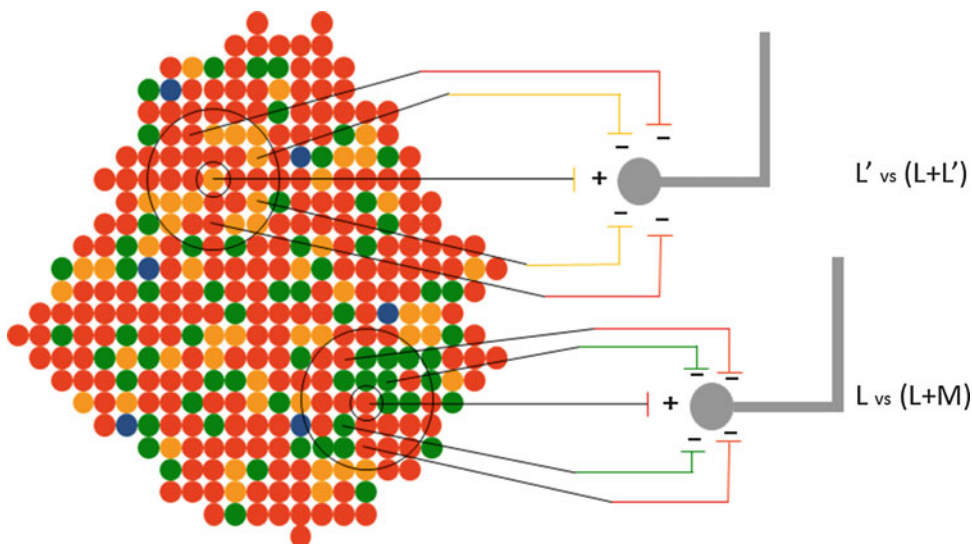
To allow tetrachromacy, it would seem that two essential conditions should be met: Firstly, in the middle- to long-wave region of the spectrum, the heterozygote must have three photopigments that are sufficiently different in their spectral absorption curves. Secondly, there must be independent neural channels that preserve the ratios of excitations of these photopigments.

- (i) It is known that there is a range of L pigments with different peak sensitivities and similarly

a range of M pigments [9, 10]. In particular, a common polymorphism at site 180 in the L cone photopigment gives a ~ 4 nm spectral difference in sensitivity [4, 11]. In addition, some hybrid pigments may be present in lower optical density and thus have narrower absorption curves. How big a difference in absorption curve between, say, L and L' is sufficient for tetrachromatic color vision? It may be significant that the strongest candidate for tetrachromacy identified so far (see below, [16]) has L and L' cone photopigments with an estimated spectral difference of 12 nm.

- (ii) Color vision does not simply depend on the number and type of retinal cone photopigment. In order to perceive colors and be able to discriminate between them, the inputs from different classes of cone need to be compared and two chromatic post-receptoral channels have been identified for this task (for discussion see [12]).

Of particular interest for tetrachromatic color vision is the phylogenetically younger, "red-green" opponent channel. Cells of the parvocellular pathway have receptive fields that are divided into spatially and chromatically antagonistic center and surround regions: In the fovea, midget ganglion cells draw their center input from a single cone of one type (e.g., L or M), whereas the surround input is likely to be drawn indiscriminately from L and M cones (mixed surround) (for review see [13]). Such a channel lends itself to tetrachromatic color vision as its chromatic specificity is determined primarily by the type of cone feeding into the center. Thus, if the signal from a single L' cone is compared to the pooled input of other cones, say L + L', as is the case in Da, a cDa individual too will have the neural basis of making similar comparisons. Figure 2 shows a schematic representation of a tetrachromatic cone mosaic with a deuteranomalous (L' vs (L + L')) as well as a normal (L vs (L + M)) red-green channel. If both channels are capable of providing salient input to cortical mechanisms, then strong tetrachromacy should result.



Tetrachromatic Vision, Fig. 2 Schematic representation of a tetrachromatic cone mosaic with a “deuteranomalous” and a normal red-green channel. For explanation see text

Historical Considerations

Prior to the 1980s the majority of scientists concerned with individual differences in human color vision and possible phenotypic manifestations of X-linked heterozygosity believed that female carriers of red-green deficiencies either shared a little in their sons’ disability or did not differ in any significant way from color-normal observers (for review see [1]). A different view was held by the Dutch physicist De Vries [14] who, in 1948, explicitly stated that the daughters of a Da father “must be tetrachromatic,” and that it should be possible to demonstrate (strong) tetrachromacy using color mixing experiments. That the existence of an additional cone type in the retina indeed has an effect on color matches was found by Nagy and colleagues 32 years later by demonstrating failure of the additivity law [15]. According to this fundamental law of color matching for trichromats, the addition of a monochromatic light to an existing color match (here a match of a 546 nm and a 660 nm mixture ratio to a 588 nm reference light) should not upset the match even though the overall appearance might be different. Four carriers of an unidentified type, but no color-normal male or female control, changed their matches after chromatic adaptation. Since the

carriers were nevertheless capable of making trichromatic matches, they would be classified as weak tetrachromats.

Methodological Issues and Results for Strong Tetrachromacy

Explicit tests for strong tetrachromacy must ensure that a fourth chromatic primary is provided in the middle- to long-wave range of the spectrum on which the tetrachromat can base her discriminations. Thus no ordinary RGB color monitor can be used for any test of this kind.

Ideal is a specifically designed colorimeter which allows inclusion and manipulation of the wavelength of n primary lights with independent control of luminance. It would also permit the precise specifications of spatial and temporal stimulus parameters. Such a system was used by Jordan et al. [16] for the identification of tetrachromats in a Rayleigh-type discrimination test. The stimuli used were in the spectral range from 556 nm to 670 nm, where S cones are insensitive. A triplet of three successive lights was presented. One light was a red-green mixture (546 + 670 nm) that could vary in the proportion of red, and the other two were a monochromatic

yellow (590 nm). Observers were asked to identify the red-green mixture. Since all normal trichromats can make a dichromatic color match in this spectral range, they were expected to fail to detect the mixture light for some red-green mixture and some intensity of the monochromatic yellow. In contrast, a strong tetrachromat was predicted to always detect the mixture with her L' (M') cones. Only one out of 18 cDa (known as cDa29) and none of 7 cPa participants performed according to the prediction. As expected, all control subjects failed to discriminate. Since cDa29 could not make a trichromatic color match with only two primaries available in the Rayleigh region, she was classified as strong tetrachromat.

Another approach to explicitly test for strong tetrachromacy is to ask candidates to discriminate surface colors. The challenge here is that one needs to find stimuli that are metameric (indistinguishable) for normal observers, but look distinct to a tetrachromat. Such a test was designed to demonstrate that Da individuals are capable of making discriminations not achievable by normal observers [17] and cDa individuals ought to be able to make the same discriminations if their L vs ($L + L'$) channel is salient enough. Participants were asked to rate how distinct pairs of surface colors (two sets of different blue-yellow acrylic mixtures) looked [12]. Multidimensional scaling (MDS) was then used to reconstruct the subjective color space for each observer. Four out of nine cDa observers were found to be able to make discriminations on this test, and their MDS solution correlated significantly with the theoretical L vs ($L + L'$) channel. One of these four carriers was cDa29. Molecular genetic analyses of saliva samples taken from these observers confirmed the existence of a hybrid gene for an L' cone [16].

Future Directions

To provide the most conclusive evidence for strong tetrachromacy, a full set of color-matching functions is needed for selected candidates. This would allow direct derivation of the four-dimensional color space of four-cone individuals.

Furthermore, the influence of factors such as cone ratios, cone optical density, and absorption by the ocular media must be considered in further tests.

Cross-References

► Color Vision, Opponent Theory

References

1. Brindley, G.S.: Human colour vision. *Progr. Biophys.* **8**, 49–94 (1957)
2. Jordan, G., Mollon, J.D.: A study of women heterozygous for colour deficiencies. *Vision. Res.* **33**, 1495–1508 (1993)
3. Young, Th.: The Bakerian lecture: On the theory of light and colours. *Phil. Trans. Roy. Soc. Lond.* **92**, 12–48 (1802)
4. Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B., Hogness, D.S.: Molecular genetics of inherited variation in human colour vision. *Science* **232**, 203–210 (1986)
5. Merbs, S.L., Nathans, J.: Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science* **258**, 464–466 (1992)
6. Deeb, S.S.: Genetics of variation in human color vision and the retinal cone mosaic. *Curr. Opin. Genet. Dev.* **16**, 301–307 (2006)
7. Mollon, J.D., Bowmaker, J.K., Jacobs, G.H.: Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proc. R. Soc. Lond. B.* **222**, 373–399 (1984)
8. Lyon, M.F.: X-chromosome inactivation and developmental patterns in mammals. *Biol. Rev.* **47**, 1–35 (1972)
9. Alpern, M., Pugh, E.N.: Variation in the action spectrum of erythrolabe among deuteranopes. *J. Physiol.* **266**, 613–646 (1977)
10. Asenjo, A.B., Rim, J., Oprian, D.D.: Molecular determinants of human red/green color discrimination. *Neuron* **12**, 1131–1138 (1994)
11. Windericks, J., Lindsey, D.T., Sanocki, E., Teller, D. Y., Motulsky, A.G., Deeb, S.S.: Polymorphism in red photopigment underlies variation in colour matching. *Nature* **356**, 431–433 (1992)
12. Mollon, J.D.: Tho' she kneeled in that place where they grew. . . The uses and origins of primate colour vision. *J. Exp. Biol.* **146**, 21–38 (1989)
13. Dacey, D.M.: Primate retina: cell types, circuits and color opponency. *Prog. Retin. Eye Res.* **18**, 737–763 (1999)
14. De Vries, H.: The fundamental response curves of normal and dichromatic and trichromatic eyes. *Physica* **14**, 367–380 (1948)

15. Nagy, A.L., MacLeod, D.I.A., Heyneman, N.E., Eisner, A.: Four cone pigments in women heterozygous for color deficiency. *J. Opt. Soc. Am.* **71**, 719–722 (1981)
16. Jordan, G., Deeb, S.S., Bosten, J.M., Mollon, J.D.: The dimensionality of color vision in carriers of anomalous trichromacy. *J. Vis.* **10**(8), 12, 1–19 (2010)
17. Bosten, J.M., Robinson, J.D., Jordan, G., Mollon, J.D.: Multidimensional scaling reveals a color dimension unique to ‘color-deficient’ observers. *Curr. Biol.* **15**(23), R950–R951 (2006)

Textile Processing

► [Finish, Textile](#)

Textile Wet Processing

► [Finish, Textile](#)

Texture Measurement, Modeling, and Computer Graphics

Eric Kirchner
Color Research, AkzoNobel Performance
Coatings, Sassenheim, The Netherlands

Synonyms

[Coarseness](#); [Glints](#); [Graininess](#); [Sparkle](#)

Definition

Following a definition by ASTM, texture can be defined as the visible surface structure depending on the size and organization of small constituent parts of a material, typically surface structure of a woven fabric. Many other materials that are used in the manufacturing industry also show texture, often of very different types. Examples are various types of wood, satin, paper, foods, and

injection-molded plastics. This review focuses on effect coatings, since the characterization of their texture is advanced as compared to the case for most other materials. Also, effect coatings such as metallic and pearlescent coatings are widely used in the automotive, cosmetics, and packaging industry.

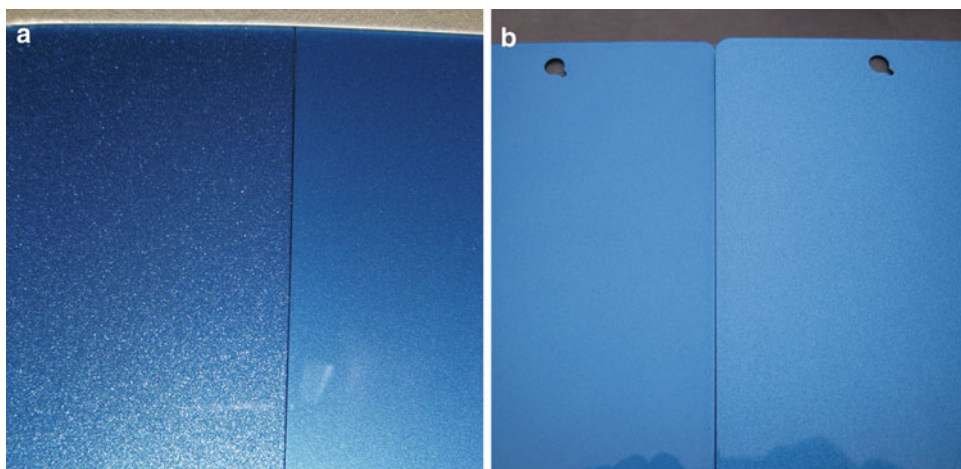
The texture of effect coatings is associated with the spatial variation in appearance caused by nonuniformity of colorant. This texture depends on the type of lighting, as shown in Fig. 1. Under intense unidirectional lighting, glint impression (also known as sparkle) is visible. This is the overall impression of several or many tiny light spots (“glints”) that are strikingly brighter than their surroundings. Under diffuse illumination conditions, the texture of effect coatings changes into diffuse coarseness (also known as graininess). This is the perceived contrast in the light/dark irregular pattern that is visible under diffuse illumination [1].

Some other parameters for describing the texture of effect coatings have been introduced as well, and some of them will be discussed further below. However, those parameters generally lack strict definitions.

Overview

Conventional colorimetric results such as the modern color difference equations assume that the so-called reference conditions hold. The Commission Internationale de l’Éclairage (CIE) recommended the investigation of nonreference conditions. According to one of the reference conditions, samples should be spatially uniform. Since effect coatings clearly do have a texture, their industrial importance explains the large body of research that has been conducted over the past decade on the texture of effect coatings.

Below, the progress is summarized that has been made on different subtopics in this area. Only recently, it became possible to objectively measure the texture of effect coatings. These developments are highlighted, since they also enabled the first quantitative studies on the



Texture Measurement, Modeling, and Computer Graphics, Fig. 1 (a) Two blue metallic paint samples differing in texture (sparkle, glint impression) as viewed

under intense directional light. (b) Two blue metallic paint samples differing in texture (graininess, coarseness) as viewed under diffuse light

influence of texture on perceived color differences between effect coatings. Because of these developments, it now has become possible to set tolerances on texture parameters. For color formulation and recipe prediction, models are needed that are able to predict texture values based on concentrations of colorants, similar to, e.g., the Kubelka-Munk theory for color prediction. Finally, it is discussed how the appearance of texture can be accounted for in computer graphics and 3D rendering of objects.

Measurement and Observation of Texture

As long as no clear definitions of texture parameters were introduced and an understanding of the relation between illumination conditions and the type of texture of effect coatings was lacking, no commercial instrument for measuring texture could be developed. After having derived clear definitions and constraints on illumination conditions, a large amount of visual data was collected on different types of texture: diffuse coarseness for diffuse illumination conditions and glint impression for intense directional lighting. This body of data was obtained for hundreds of

different effect coatings observed under different angles and illumination conditions. The visual data was correlated with texture parameters that were obtained from image analysis applied on CCD images of the same samples. In this way, algorithms were obtained to train and test prototypes of an instrument [1, 2].

The resulting instrument is the BYK-mac™, which is commercially available from BYK-Gardner since 2007. Similar studies are being conducted by other manufacturers of optical instruments as well [3], whereas also the development of the underlying image analysis algorithms has been taken up by the academic world [2].

Based on feedback from the automotive OEM industry, the set of texture parameters considered to be best applicable for the OEM industry has gradually evolved over the years. A graininess parameter G_r was introduced that has a high correlation with the diffuse coarseness parameter mentioned before. Apart from that, glint impression (sparkle) is considered to have two separate dimensions, called sparkle intensity S_i and sparkle area S_a [4]. An overall parameter sparkle grade (S_G) is no longer thought to be important.

Recently, also lighting booths became available that focus on observing the different types of texture under various geometries [5].

Influence of Texture on Color Perception

Using artificially generated images of textures, several studies showed that the presence of texture has an influence on perceived color differences [6]. Images exhibiting textured lightness variations have been shown to increase the visual tolerance for lightness deviations. This explains why, for effect coatings, a color difference equation like dE_{CMC} is best used with $l = 1.5$; $c = 1.0$; for solid (untextured) colors, best results are typically obtained for $l = 1.5$; $c = 1.0$.

However, quantitative studies on effect coatings only became possible after 2007 as outlined in the previous section. Based on a large study on 405 effect sample pairs, an expression was derived that quantitatively describes how differences in texture and color combine into a total appearance difference. The validity of this approach was confirmed in later research [7].

Setting Tolerances on Texture

Being able for the first time to measure texture parameters of effect coatings, manufacturing industries decided to include consistency in texture in their criteria for quality control. Among the first industries to investigate how to set industrial pass-fail tolerances on texture are the automotive industry and the plastics industry, where texture matching used to be a long-standing problem for quality control [8].

Most of the global automotive manufacturers are currently setting texture tolerances, as evidenced, for example, by multiparty meetings of the Detroit Color Council in the USA and BYK-Gardner user meetings in Europe. However, no specific tolerances on texture parameters have been published yet.

Based on perception thresholds of texture parameters, a proposal for realistic tolerances was recently published for different sets of texture parameters: diffuse coarseness, glint impression, graininess, and sparkle grade [9]. Only for the parameters sparkle area and sparkle intensity, the lack of published data makes it not yet possible to derive tolerance values.

Predictive Models for Texture

For color matching, optical models are widely used to optimize color formulations. In the 1940s, Kubelka and Munk developed a model that is still widely used in the paint and textile industry. In order to optimize the concentrations of colorants such that also the texture of an effect coating is matched, a similar model for texture is needed. Studies show that texture parameters like sparkle intensity and graininess can vary greatly with paint composition [10]. However, an accurate optical model for predicting texture based on a physical model like Kubelka-Munk has not been developed yet. The first example of an accurate model for texture uses a very different derivation [11]. It is based on a statistical approach for finding those mathematical terms that contribute significantly to the texture properties that are to be predicted.

Rendering of Texture in Computer Graphics

Digital rendering of three-dimensional objects has become very common in the cinematic and computer game industries. However, for accurate rendering of effect coatings also, the texture aspect needs to be taken into account. Most current approaches either completely ignore texture aspects or utilize a very crude and nonquantitative approach by adding sparkles to the image afterward [12].

Recently, two different approaches were published that do make accurate texture rendering possible. Using a few measured texture values as input as well as reflection measurements from an ordinary multiangle spectrophotometer, it was shown that accurate renderings can be produced [13]. No full BRDF and BTF measurements are needed. The core algorithm was derived from procedures used for determining texture parameter values based on images captured from effect coatings. By reversing these procedures, an algorithm was obtained that produces images resembling effect coatings, based on measured texture parameter values. The algorithm does not aim at

simulating the reflection from each and every flake in a scene, but it accounts for their combined effect in an effective way.

In contrast to conventional methods for rendering, the derivation of the algorithms and the verification of the accuracy of the visualizations were both determined using visual tests in which rendered images were directly compared to physical objects. Thus, a truly appearance-based method for rendering color and texture was obtained, with greatly reduced demands on computation power and disk storage.

A different approach was published very recently by Ferrero et al. [14]. It uses an analytical model to visualize flake reflectance, in which, for example, the aperture angle of the light source and the diameter of the eye pupil of the observer are used as input parameters. In this way, a model is obtained that is able to generate images varying from purely directional until purely diffuse lighting and any illumination condition in between.

Future Directions

Despite all the work that has been done on texture research over the past decade, this area is still in development. There is a clear need for more accurate image analysis algorithms to extract parameters from CCD images that correlate well with visually perceived texture. For manufacturers of optical instruments, this is an absolute necessity for developing instruments that can form a next generation after the commercial launch of the BYK-mac™ in 2007.

Perception studies on the effect of texture on perceived color, or on perceived color differences, is also expected to be an active area of research for the coming decade [15]. Currently, color difference equations are used that are derived from very different types of samples, such as solid colors or textiles. At present, texture can only be accounted for through parametric factors.

Setting tolerances on texture parameters and predicting texture properties based on paint composition are both areas of great industrial importance. Yet, these areas have hardly been explored yet.

For accurate rendering of textured paints, there are a few first publications, but also here the potential for improvement is enormous, and there are many obvious industrial applications in, e.g., the movie and the game maker industry.

Cross-References

► Color Categorical Perception

References

1. Kitaguchi, S., Westland, S., Owens, H., Luo, M.R., Pointer, M.R.: Surface texture – a review. NPL Report DQL-OR 006 (2004); Kirchner, E.J.J., van den Kieboom, G.J., Njo, S.L., Supèr, R., Gottenbos, R.: The appearance of metallic and pearlescent materials. *Col. Res. Appl.* **32**, 256–266 (2007)
2. Kitaguchi, S., Westland, S., Kirchner, E.J.J., van den Kieboom G.J., Luo, M.R.: Computational model for perceptual coarseness prediction. In: *Proceedings Conference on Graphics, Imaging and Vision*, pp. 267–282 (2006); Huang, Z., Xu, H., Luo, M.R.: Camera-based model to predict the total difference between effect coatings under directional illumination, Leeds (UK), *Chin. Opt. Lett.* **9**, 093301–093305 (2011)
3. Kigle-Böckler, G., Weixel, S.: New measurement system for characterizing the total color impression of effect coatings. In: *Proceeding Conference of AIC*, Newcastle (UK), pp. 209–212 (2013); Ellens, M.S., Lamy, F.: From color to appearance in the real world. In: *Proceeding of Human Vision and Electronic Imaging XVII, SPIE, Burlingame (USA)*, vol. 8291 (2012)
4. Klein, G.A.: *Industrial Color Physics*. Springer, New York (2010); Kigle-Böckler, G.: New and innovative testing technologies for effect finishes. In: *Proceeding Conference of AIC*, Sydney (2009)
5. Cramer, W.R.: Die neue Version der Gonio-Vision Box. *Phänomen Farbe*, 38–39 (2012); Martínez-Verdú, F., Perales, E., Viqueira, V., Chorro, E., Burgos, F.J., Pujol, J.: Comparison of colorimetric features of some current lighting booths for obtaining a right visual and instrumental correlation for gonio-apparent coatings and plastics. In: *Proceeding of CIE x037:2012 Lighting Quality and Energy Efficiency*, Hangzhou (2012)
6. Huertas, R., Melgosa, M., Hita, E.: Influence of random-dot textures on perception of suprathreshold color differences. *J. Opt. Soc. Am. A* **23**, 2067–2076 (2006); Xin, J.H., Shen, H.L., Lam, C.C.: Investigation of texture effect on visual colour difference evaluation. *Col. Res. Appl.* **30**, 341–347 (2005); Montag, E.D., Berns, R.S.: Lightness dependencies and the effect of texture on suprathreshold lightness tolerances. *Col. Res. Appl.* **25**, 241–249 (2000)

7. Dekker, N., Kirchner, E.J.J., Supèr, R., van den Kieboom, G.J., Gottenbos, R.: Total appearance differences for metallic and pearlescent materials: contributions from color and texture. *Col. Res. Appl.* **36**, 4–14 (2011); Huang, Z., Xu, H., Luo, M.R., Cui, G., Feng, H.: Assessing total differences for effective samples having variations in color, coarseness and glint. *Chin. Opt. Lett.* **8**, 717–720 (2010)
8. Longley W.V.: Automotive color certification. *Col. Res. Appl.* **20**, 50–54 (1995); Ariño, I., Johansson, S., Kleist, U., Liljenström-Leander, E., Rigdahl, M.: The effect of texture on the pass/fail colour tolerances of injection-molded plastics. *Col. Res. Appl.* **32**, 47–54 (2007)
9. Kirchner, E.J.J., Ravi, J.: Setting tolerances on color and texture for automotive coatings. *Col. Res. Appl.* **39**, 88–98 (2014)
10. Rentschler, T.: Measuring sparkling blues without blues. *Eur. Coat. J.* **11**, 78–83 (2011)
11. E. Kirchner, E.J.J., Ravi, J.: Predicting and measuring the perceived texture of car paints. In: *Proceeding 3rd International Conference on Appearance. “Predicting Perceptions”*, Edinburgh (UK) (2012)
12. Meyer, G., Shimizu, C., Eggly, A., Fischer, D., King, J., Rodrigues, A.: Computer aided design of automotive finishes. In: *Proceeding of AIC Colour*, pp. 685–688 (2005); Ershov, S., Kolchin, K. and Myszkowski, K.: Rendering pearlescent appearance based on paint-composition modeling. *Eurographics 2001*, Granada (Spain), *Comp. Graph. For.* **20**, C221–C238 (2001)
13. Van der Lans, I.B.N., Kirchner, E., Half, A.: Accurate appearance-based visualization of car paints. In: *Proceedings of the European Conference on Colour in Graphics, Imaging and Vision*, Amsterdam, pp. 17–23 (2012)
14. Ferrero, A., Campos, J., Rabal, A.M., Pons, A.: A single analytical model for sparkle and graininess patterns in texture of effect coatings. *Opt. Exp.* **21**, 26812–26819 (2013)
15. Kandi, S.G., Tehran, M.A.: Investigating the effect of texture on the performance of color difference formulae. *Col. Res. Appl.* **35**, 94–100 (2010)

Tincture

Paul Green-Armytage
School of Design and Art, Curtin University,
Perth, Australia

Definition

The word “tincture” has more than one meaning in English. *The Concise Oxford Dictionary* has the

following: “**1** a slight flavor or trace. **2** a tinge (of a color). **3** a medicinal solution (of a drug) in alcohol (*tincture of quinine*). **4** *Heraldry* an inclusive term for the metals, colors, and furs used in coats of arms.”

The word is used by specialists in particular fields and is uncommon in general conversation. However, some of the concepts associated with the word, if not the word itself, have potential value for those who are concerned with color and ► [appearance](#). In meaning 2, the word refers to very subtle colors. A reporter might describe a bride’s dress as having a beautiful tincture, white with a tinge of pink.

The word has a particular meaning in heraldry (meaning 4), and this draws attention to the lack of a word in general English usage that could embrace several different aspects of appearance in a single word. In heraldry, tincture includes color, luster, and texture. In coats of arms, the designs can be rendered in colors, “metals” (luster), and “furs” (texture). If tincture could be understood to mean all three aspects of appearance, the tincture of the bride’s dress could mean that the dress fabric also had a smooth texture and a lustrous sheen as well as being white with a tinge of pink.

In heraldry, the names used to identify the tinctures are Norman French. There are five “colors”: *gules* (red), *azure* (blue), *sable* (black), *vert* (green), and *purpure* (purple). Three other colors are found in rare cases and are sometimes referred to as “stains”: *sanguine* (blood red), *tenné* (tawny orange), and *murrey* (mulberry). The “metals” are *or* (gold) and *argent* (silver) – often represented by yellow and white. The “furs” are variations on the theme of ermine and vair (squirrel skins) and are represented by stylized patterns.

Where a figure, such as a cross, is set over a background, the “rule of tincture” is applied: the figure must not be from the same group of tinctures as the background. So a red (color) cross can be placed on a gold (metal) background but not on blue (color). This is to ensure maximum light–dark contrast and, therefore, maximum legibility and ease of identification.

There are lessons in heraldry for graphic designers today. The principles embodied in the rule of tincture are as relevant today as they were in the Middle Ages, and the concept of heraldic

tincture is a reminder that there are other aspects of ► [appearance](#), besides color, that can be exploited in design.

It has been suggested that the word “tincture” might be adopted and promoted with something like its heraldic meaning. “Tincture” could mean the totality of appearance characteristics, not only color, texture, and luster, as in heraldry, but also those other aspects of appearance that have been collected under the term “cesia.” “Cesia” is a term coined by César Jannello. It refers to ► [transparency](#), translucency, opacity, gloss, and specular reflection, each of which can also be represented in scales as from matte via semigloss to full gloss. And if the word “tincture” itself is problematic, with its meanings already established, perhaps another new word could be coined that would embrace all aspects of appearance: color, texture, and cesia.

Cross-References

- [Appearance](#)
- [Transparency](#)

Tint

- [Pigment, Ceramic](#)

Total Appearance

- [Appearance](#)

Transition-Metal Ion Colors

Richard J. D. Tilley
Queen's Buildings, Cardiff University,
Cardiff, UK

Synonyms

[d-block elements](#)

Definition

Transition metals are d-block elements with partially filled 3d, 4d, and 5d orbitals.

Color Production

The transition metals are (somewhat imprecisely) described as being colored because when cations of these elements are incorporated into colorless solids or liquids, the material frequently takes on a characteristic hue (Fig. 1, Table 1). The color arises from electronic transitions between the ionic ground state and energy levels lying between 1.77 and 3.10 eV above it, giving absorption maxima in the visible wavelength range (400–700 nm). The low-lying energy levels that give rise to color arise from interactions of the d orbitals on the cation with neighboring atoms in a material and are a function of the symmetry of the surroundings [1–3]. (These ions in the gaseous state are not colored). The energy levels that occur with the 4d and 5d transition metals are higher in energy than those of the 3d series and are of lesser importance in color terms. In this entry only the 3d transition-metal ions are covered and, in addition, only single dopants are considered. (When two or more transition-metal ion impurities are present, an alternative mechanism, charge transfer, can also lead to coloration).



Transition-Metal Ion Colors, Fig. 1 Synthetic ruby spheres consisting of normally colorless Al_2O_3 doped with trace amounts of Cr^{3+}

Transition-Metal Ion Colors, Table 1 Typical 3d transition-metal ion colors

Ion	Color	Symmetry	Host matrix
Ti ³⁺ d ¹	Purple	Tetrahedral	Silicate glass
V ⁴⁺ d ¹	Red	Tetrahedral	Silicate glass
V ³⁺ d ²	Green	Tetrahedral	Silicate glass
	Blue	Octahedral	Al ₂ O ₃
Cr ³⁺ d ³	Green	Tetrahedral	Silicate glass
	Green	Octahedral	Be ₃ Al ₂ Si ₆ O ₁₈ (emerald)
	Red	Octahedral	Al ₂ O ₃ (ruby), TiO ₂ , MgAl ₂ O ₄
	Violet	Octahedral	KCr(SO ₄) ₃ ·12H ₂ O (chrome alum)
Mn ³⁺ d ⁴	Purple	Tetrahedral	Silicate glass
Mn ²⁺ d ⁵	Yellow	Tetrahedral	Silicate glass
	Red	Octahedral	MnCO ₃
	Green	Octahedral	MnO
	Pink	Octahedral	MnSiO ₃
Fe ³⁺ d ⁵	Green	Tetrahedral	Silicate glass
	Yellow	Octahedral	Al ₂ SiO ₄ (OH) (topaz)
Fe ²⁺ d ⁶	Green	Octahedral	Fe(H ₂ O) ₆ ²⁺ in solution and hydrates
	Red	Cubic	Ca ₃ Al ₂ Si ₃ O ₁₂ (garnet)
Co ²⁺ d ⁷	Blue	Tetrahedral	CoAl ₂ O ₄ , silicate glass
	Pink	Octahedral	Co(H ₂ O) ₆ ²⁺ in solution and hydrates
Ni ²⁺ d ⁸	Green	Octahedral	Ni(H ₂ O) ₆ ²⁺ in solution and hydrates
	Yellow	Octahedral	Al ₂ O ₃ , NiCl ₂
Cu ²⁺ d ⁹	Green	Octahedral	Cu ₂ (OH) ₂ (CO ₃) (malachite)
	Blue	Octahedral	Cu(H ₂ O) ₆ ²⁺ in solution and hydrates

Transition-Metal Ion Colors, Table 2 Low-energy terms of free 3d transition-metal cations

Ion	Free-ion terms ^a
Ti ³⁺ V ⁴⁺ d ¹	²D
V ³⁺ d ²	³F ³ P ¹ D ¹ G
Cr ³⁺ d ³	⁴F ⁴ P ² G ² P
Mn ³⁺ Cr ²⁺ d ⁴	⁵D
Mn ²⁺ Fe ³⁺ d ⁵	⁶S ⁴ G ⁴ P ⁴ D ⁴ F
Fe ²⁺ d ⁶	⁵D
Co ²⁺ d ⁷	⁴F ⁴ P ² G ² P
Ni ²⁺ d ⁸	³F ³ P ¹ D ¹ G
Cu ²⁺ d ⁹	²D

^aGround state term **bold**

designation, each term is written as ^{2S+1}L where L is a many-electron quantum number describing the total orbital angular momentum of the electrons surrounding the atomic nucleus and S is a many-electron quantum number representing the total electron spin. The superscript (2S+1) is called the *multiplicity* of the term. The total angular momentum quantum number L is given a *letter symbol*: S (L = 0), P (L = 1), D (L = 2), F (L = 3), and thereafter alphabetically, omitting J. The energies of the terms must be determined by quantum mechanical calculations, except for that of the ground state, which is given by Hund's second rule: the ground state is the term with the highest multiplicity and, if more than one term of the same multiplicity is present, by that with the highest L value. The lower-energy terms of free 3d transition-metal cations are given in Table 2.

Crystal Field Splitting

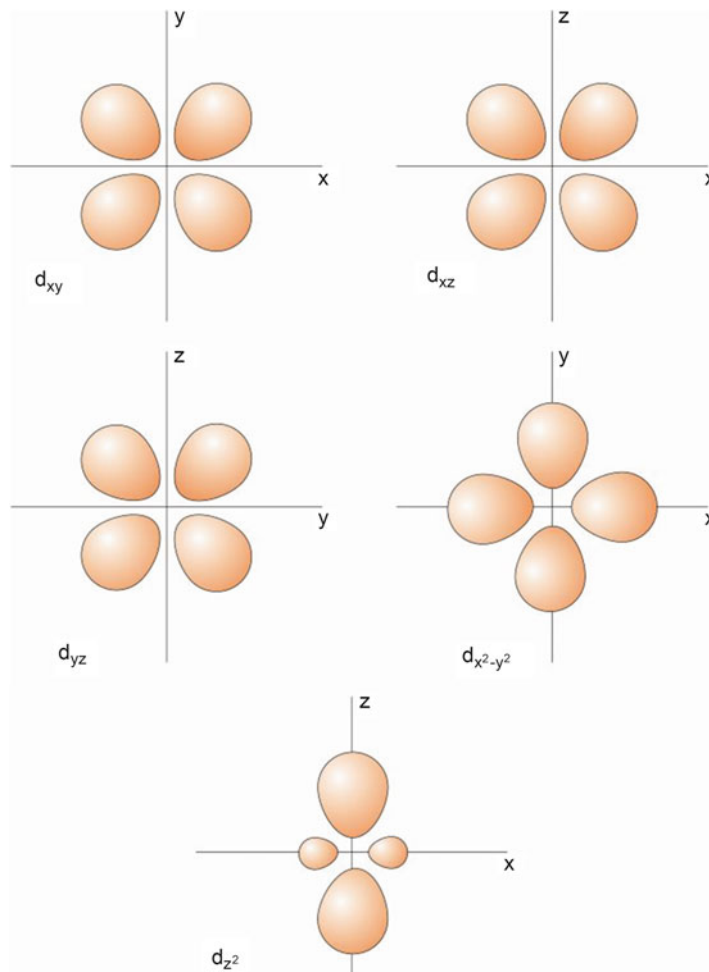
When a cation is placed into a crystal, the d electrons interact with the surrounding anions and give rise to new energy states. The simplest conceptual model that describes the formation of the color-producing energy levels is called *crystal field theory* [5, 6]. In this approach the electrostatic effect of neighboring anions upon the energies of the cationic d orbitals is considered in terms of surrounding point charges. More comprehensive treatments, which include the effects of covalent bonding, fall into the domain of *ligand field theory* [7, 8], although the two expressions are often used interchangeably.

Electronic Energy Levels

Energy Levels and Terms

The energy levels of a free transition-metal cation are dominated by electron–electron repulsion and are described by *term symbols*. A *term* is a set of states which are very similar in energy. Transitions between terms, or, more precisely, the energy levels specified by the term, give rise to the observed line spectrum of the free ion. The terms of an ion can be derived by Russell–Saunders (LS) coupling [2, 4]. In this

Transition-Metal Ion Colors, Fig. 2 Electron density lobes of d orbitals



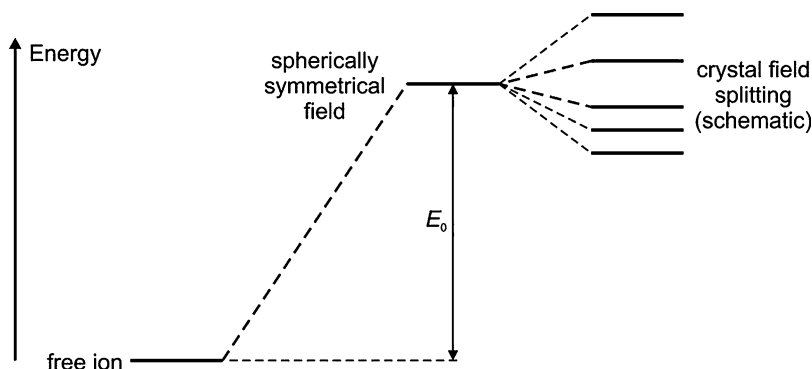
The d orbitals on a 3d transition-metal cation can be described as pointing along or between a set of x-, y-, and z-axes (Fig. 2). The orbitals directed between the axes are the d_{xy} , d_{yz} , and d_{xz} set and those pointing along the axes are the $d_{x^2-y^2}$ and d_{z^2} pair. *If electron–electron interactions are ignored*, all of these orbitals are degenerate (have the same energy). However, this is not true when the ion is placed into a crystal because of the interaction (most easily imagined as repulsion) between the electrons on the surrounding ions and the electrons occupying d orbitals. If these surrounding charges were distributed evenly over the surface of a sphere, the five d orbitals would still be energetically degenerate, although at a higher energy by an amount E_0 . In a crystal the charges are not smeared out but arranged in a

pattern. This removes the degeneracy and gives rise to a new set of energy levels (Fig. 3). The pattern of this *crystal field splitting* depends upon the symmetry of the surrounding anions.

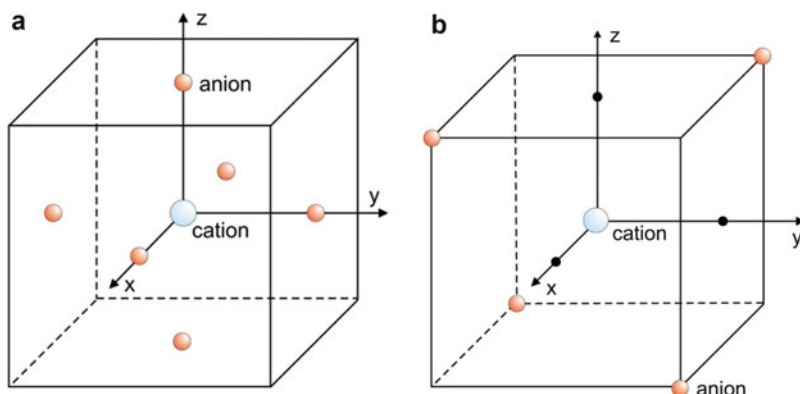
The two most important geometries to consider, especially for oxide pigments and ceramics, are octahedral and tetrahedral coordination (Fig. 4a, b). When a cation is surrounded by six anions arranged at the vertices of a regular octahedron, the $d_{x^2-y^2}$ and d_{z^2} orbitals point directly toward the surrounding anions and are raised in energy more than the d_{xy} , d_{xz} , and d_{yz} trio, which point between the anions. This splits the originally degenerate energy levels into two groups. The higher-energy state, labeled e_g , consists of two equal-energy levels, derived from the $d_{x^2-y^2}$ and d_{z^2} orbitals. The lower-energy state, labeled t_{2g} , consists of three equal-energy levels,

Transition-Metal Ion

Colors, Fig. 3 Change of d-orbital energies when a free ion is introduced into a crystal (schematic)

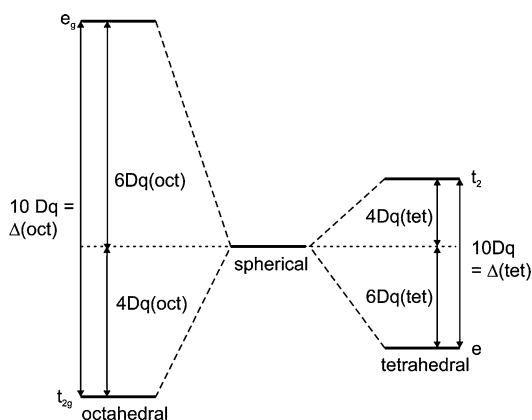
**Transition-Metal Ion**

Colors, Fig. 4 (a) Octahedral and (b) tetrahedral coordination geometry. A cation is at the cube center. In (a), anions are at the center of each face. In (b), four anions occupy half the cube corners



derived from the d_{xy} , d_{xz} , and d_{yz} orbitals (Fig. 5). The energy gap between the t_{2g} orbitals and the e_g orbitals is written Δ or $10\Delta q$, with the t_{2g} set at $-4\Delta q$ and the e_g set at $+6\Delta q$ with respect to the spherically symmetrical situation. When Δ is small, the crystal field is said to be *weak*, and when large it is referred to as *strong*. Strong crystal fields arise from ions with multiple charges and relatively short cation–anion spacing, together with contributions due to covalent bonding.

When a cation is surrounded by a tetrahedron of anions, the crystal field splitting found for octahedral coordination is reversed. In this case the higher-energy group, now labeled t_2 , is derived from the d_{xy} , d_{xz} , and d_{yz} orbitals, and the lower-energy group, now labeled e , is derived from the $d_{x^2-y^2}$ and d_{z^2} orbitals. The t_2 set is at $+4\Delta q$ and the e set is at $-6\Delta q$ with respect to the spherically symmetrical situation (Fig. 5). The magnitude of Δ for tetrahedrally coordinated cations is $4/9$ of that for the same cation when octahedrally coordinated.



Transition-Metal Ion Colors, Fig. 5 Splitting of d-orbital energy levels in a crystal field: octahedral and tetrahedral symmetries

Term Splitting

The electronic energy levels of a free transition-metal ion arise from electron–electron interactions, and the simple model just outlined must be adapted so as to apply to the free-ion terms. It is

Transition-Metal Ion Colors, Table 3 Splitting of terms in fields of cubic symmetry

Free-ion term	Terms in tetrahedral crystal field	Terms in octahedral crystal field
S	A ₁	A _{1g}
P	T ₁	T _{1g}
D	E, T ₂	E _g , T _{2g}
F	A ₂ , T ₁ , T ₂	A _{2g} , T _{1g} , T _{2g}
G	A ₁ , E, T ₁ , T ₂	A _{1g} , E _g , T _{1g} , T _{2g}

found that a term may split into several energy states. The number of states that arise is a function of the value of L and the symmetry of the surrounding neighboring anions (Table 3). The magnitude of the splitting is a function of the anion–cation separation and the charges on the ions involved. The multiplicity of the term is carried over onto the new states. The energy states in the crystal are given labels that describe the degeneracy of the orbitals: A is a singly degenerate state, E represents a doubly degenerate state, and T represents a triply degenerate state. Different configurations of sets of orbitals with the same degeneracy are labeled with a subscript 1, 2, and so on. Thus a P term gives rise to a T₁ term, while a D term gives rise to a T₂ term. The subscript “g” means a center of orbital symmetry exists. The energy of the new states needs to be calculated using quantum mechanical methods or determined experimentally from spectra.

Selection Rules

Electron transitions are governed by selection rules that give the *probability* that a transition will occur. Transitions between d orbitals are forbidden by the Laporte selection rule. However, this rule may break down for ions in compounds. The main reason for this is a degree of mixing between s, p, and d orbitals can occur when an ion is not located at a center of symmetry. As s or p to d transitions are allowed, transitions giving rise to color are also allowed, to a degree corresponding to the amount of orbital mixing achieved. Thus ions situated in tetrahedral coordination, which are not at a center of symmetry, show quite strong colors. Ions at the center of a perfect octahedron are at a center of symmetry and color-producing

transitions are forbidden, but in most solids and liquids, thermal agitation of the surroundings and crystal distortions remove the precise symmetry, making such transitions weakly allowed. As a consequence, they are often less intense than those from similar ions in tetrahedral sites.

In addition, transitions are only allowed between states of the same multiplicity, called *spin-allowed transitions*. Transitions between states of differing spin can be weakly allowed in some circumstances, but in general these do not give rise to strong colors.

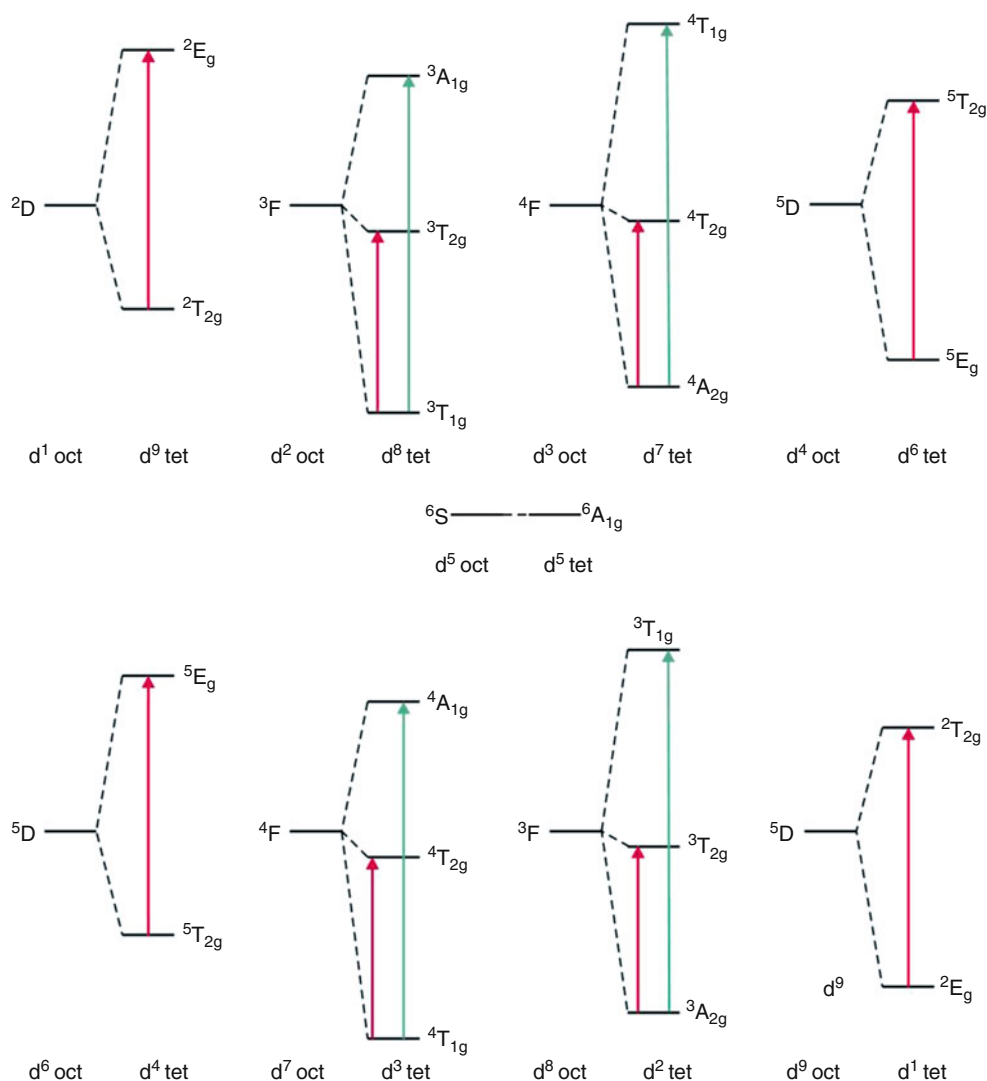
The Color of 3d Transition-Metal Ions

Color and Term Splitting

Using Tables 2 and 3, a qualitative energy level diagram can be readily constructed for the splitting of the ground state terms (Fig. 6). It is seen that the arrangement of the energy levels is symmetrical about the d⁵(⁶S) term and the arrangement for tetrahedral coordination the inverse of that for octahedral coordination. The colors exhibited by many transition-metal ions can now be understood.

Figure 6 indicates that octahedrally coordinated d⁵ ions (Fe³⁺, Mn²⁺) exhibit no ground state crystal field splitting and are not expected to show any colors. This is so, and compounds of these ions are at best weakly colored due to other factors. The free-ion ground state of octahedrally coordinated d¹ ions (Ti³⁺, V⁴⁺), d⁴ ions (Cr²⁺, Mn³⁺), d⁶ ions (Fe²⁺), and d⁹ ions (Cu²⁺) splits into two, and these ions will show one absorption band, corresponding to a transition between these two energy levels. If all or part of this band is in the visible, compounds containing these cations will show color (Table 4).

The situation with the ions d² (V³⁺), d⁷ (Co²⁺), and d⁸ (Ni²⁺) is not so simple. Figure 6 indicates that these should show two transitions, but the experimental data (Table 4) makes it clear that three transitions are observed. As transitions giving rise to color are between states of the same multiplicity, for these ions it is necessary to take into account the existence of other terms with the same multiplicity as the ground state. The ions d²



Transition-Metal Ion Colors, Fig. 6 Splitting of ground state free-ion terms in crystal fields of octahedral and tetrahedral symmetry. Transitions that give rise to color

shown as *arrows* (subscripts g should be omitted for tetrahedral coordination)

(V^{3+} , d^3 (Cr^{3+}), d^7 (Co^{2+}), and d^8 (Ni^{2+}) have two free-ion terms of the same multiplicity, 3F 3P for d^2 and d^8 and 4F 4P for d^3 and d^7 (Table 2). In the crystal field, the higher-energy 3P or 4P states give rise to 3T and 4T states (Fig. 7). Spin-allowed transitions to these additional energy levels give rise to three absorption bands in the case of V^{3+} , Co^{2+} , and Ni^{2+} (Table 4). In Cr^{3+} ions the transition to the 4T state is at a high energy and is not recorded in the visible spectrum.

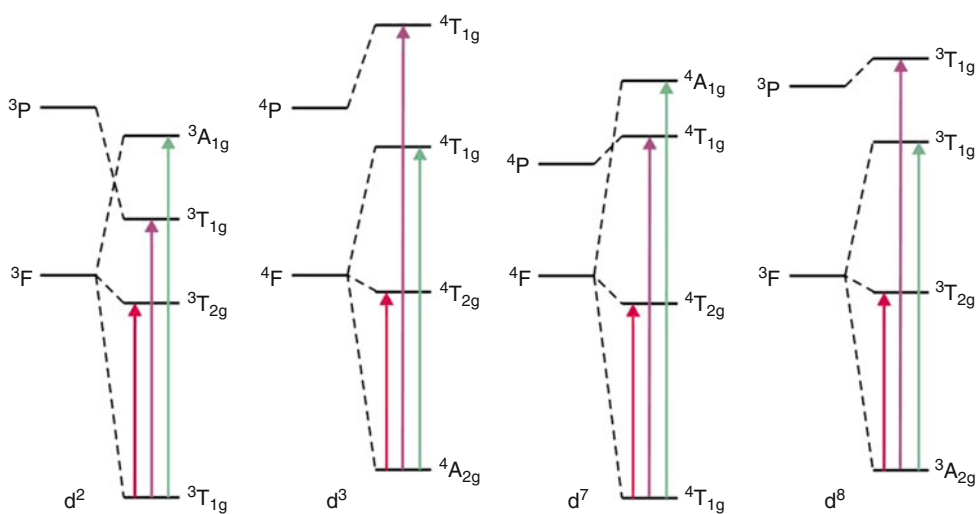
Effect of Crystal Field Strength and Other Factors

Although crystal field splitting gives an accurate picture of the number of absorption peaks observed, the actual position of these peaks depends upon the crystal chemistry of the surroundings, which alters the effective strength of the crystal field. Moreover, the absorption peaks are frequently broad or show shoulders, indicating that the

Transition-Metal Ion Colors, Table 4 Octahedral crystal field absorption peaks

Ion	Crystal field terms ^a	Absorption	Position/nm	Comments
		Peaks		
Ti ³⁺ d ¹	² T _{2g} ² E _g	² E _g → ² T _{2g}	493	In [Ti(H ₂ O) ₆] ³⁺
V ³⁺ d ²	³ T _{1g} ³ T _{2g} ³ A _{2g}	³ T _{2g} → ³ T _{1g}	575	In Al ₂ O ₃ , ³ T _{1g} → ³ T _{1g} from ³ P
			397	
		³ A _{2g} → ³ T _{1g}	294	
Cr ³⁺ d ³	⁴ A _{2g} ⁴ T _{2g} ⁴ T _{1g}	⁴ T _{2g} → ⁴ A _{2g}	556/575	In Al ₂ O ₃ /[Cr(H ₂ O) ₆] ³⁺
		⁴ T _{1g} → ⁴ A _{2g}	400/408	
Mn ³⁺ Cr ²⁺ d ⁴	⁵ E _g ⁵ T _{2g}	⁵ T _{2g} → ⁵ E _g	461/476	In MnF ₆ /[Mn(H ₂ O) ₆] ³⁺
Mn ²⁺ Fe ³⁺ d ⁵	⁶ A _{1g}	0		
Fe ²⁺ d ⁶	⁵ T _{2g} ⁵ E _g	⁵ E _g → ⁵ T _{2g}	1,000	In [Fe(H ₂ O) ₆] ²⁺
Co ²⁺ d ⁷	⁴ T _{1g} ⁴ T _{2g} ⁴ A _{2g}	⁴ T _{2g} → ⁴ T _{1g}	1,176	In [Co(H ₂ O) ₆] ²⁺
			500	
		⁴ A _{2g} → ⁴ T _{1g}	560	
Ni ²⁺ d ⁸	³ A _{2g} ³ T _{2g} ³ T _{1g}	³ T _{2g} → ³ A _{2g}	1,176/935	In [Ni(H ₂ O) ₆] ²⁺ /[Ni(NH ₃) ₆] ²⁺
		³ T _{1g} → ³ A _{2g}	741/571	
			429/355	
Cu ²⁺ d ⁹	² E _g ² T _{2g}	² T _{2g} → ² E _g	780	In [Cu(H ₂ O) ₆] ²⁺

^aDerived from ground state term for a free ion in an octahedral field, listed in ascending order



Transition-Metal Ion Colors, Fig. 7 Splitting of ground state and higher free-ion terms with the same multiplicity in a crystal field of octahedral symmetry, schematic. Transitions that give rise to color shown as arrows

simple description given above is in need of refinement. Here just one example will be given. Octahedrally coordinated Cr³⁺ ions can induce a variety of colors including ruby red, emerald green, and violet. In all cases the color arises from transitions between the ⁴A_{2g} ground state and the higher ⁴T_{2g} and ⁴T_{1g} levels. Ruby consists

of colorless corundum (Al₂O₃) crystals with ~1 % of Cr³⁺ ions occupying octahedral Al³⁺ sites. In this material the absorption peaks are the following:

- ⁴T_{2g} → ⁴A_{2g}; maximum 556 nm; green yellow
- ⁴T_{1g} → ⁴A_{2g}; maximum 400 nm; violet

The absorption at 556 nm removes green yellow and the absorption at 400 nm removes violet. Between the absorption curves there is a relatively small transmission window at approximately 486 nm and a red transmission window is present at wavelengths greater than 650 nm. This means that the color transmitted by the ruby will be red with something of a blue undertone.

Emerald consists of the colorless mineral beryl with Cr^{3+} impurity ions occupying octahedral Al^{3+} sites. In beryl the octahedra surrounding the Cr^{3+} ions are slightly larger than in corundum, and so the crystal field experienced by the Cr^{3+} in emerald is weaker than in ruby, causing a shift in the ${}^4\text{T}_{1g}$ and ${}^4\text{T}_{2g}$ levels toward the ground state with the result that the transitions move slightly toward the lower-energy end of the spectrum:

${}^4\text{T}_{2g} \rightarrow {}^4\text{A}_{2g}$; absorption maximum 650 nm; orange red

${}^4\text{T}_{1g} \rightarrow {}^4\text{A}_{2g}$; absorption maximum 450 nm; blue

The band that absorbs green yellow in ruby now absorbs orange red in emerald, and the violet absorbing band in ruby now absorbs blue. Between the absorption curves, there is a blue-green transmission window at 500 nm. The result of these small shifts is to transform the ruby red into emerald green.

In crystalline chrome alum and water solution, the Cr^{3+} ions are at the center of an octahedron of O^{2-} ions in the complex $[\text{Cr}(\text{H}_2\text{O})_6]^{3+}$. The crystal field now gives rise to the absorption bands:

${}^4\text{T}_{2g} \rightarrow {}^4\text{A}_{2g}$; absorption maximum 575 nm; yellow

${}^4\text{T}_{1g} \rightarrow {}^4\text{A}_{2g}$; absorption maximum 408 nm; violet

These bands are positioned so as to emphasize purple violet, the dominant tone of chrome alum.

Although the overall color of transition-metal ions in solids is explained by crystal field effects, the eye can detect subtle variations that require more complex explanations. Again ruby serves as an example. In this gemstone the octahedra that are occupied by the Cr^{3+} ions are distorted,

causing the ${}^4\text{T}_{2g}$ and ${}^4\text{T}_{1g}$ levels to split due to a change of local symmetry. The separation of the new levels is small but does give a color change that adds to the value of the stones and also produces polarization effects. In addition an energy level ${}^2\text{E}_g$ (derived from splitting of the free-ion ${}^2\text{G}$ term) falls between the ground state and the ${}^4\text{T}_{2g}$ level. This level is also split in the distorted octahedral sites in these crystals. Although direct excitation to and from the ground state to the ${}^2\text{E}_g$ levels is forbidden by the multiplicity selection rule, it is populated indirectly, and that gives rise to two closely spaced emission lines R_1 at 693.5 nm and R_2 at 692.3 nm in the ruby spectrum. This radiation enhances the color of the best rubies and is made to dominate light emission during ruby laser action.

Color as Structural Probe

The absorption spectrum of a transition-metal ion in a solid can give structural and chemical information that is not easily obtained via other techniques. This is because the absorption spectrum depends upon the local geometry of the site occupied and the valence state of the colored ion. Although this information is sometimes available from X-ray diffraction, in crystals where several cations with more than one valence state may statistically occupy the same site, X-ray diffraction data can be ambiguous. Moreover, the measurement of absorption spectra is diagnostic of the oxidation state of transition-metal ion and can be used to infer the local conditions prevailing during formation.

For example, the spinel structure is adopted by many compounds with a formula AB_2O_4 , where A and B are medium-sized cations. In this structure the cations sit in both octahedral and tetrahedral sites. The absorption spectra of transition-metal ions are quite different for these geometries, and so the site occupancy can be easily and unambiguously determined. The spinel NiAl_2O_4 is a case in point. The absorption spectrum of this material reveals that the Ni^{2+} ions are found in both positions. In the related spinel NiGa_2O_4 , the Ni^{2+} ions exclusively occupy octahedral sites.

The problem of site occupancy is more acute in amorphous or disordered materials such as glasses because these structures cannot be determined by X-ray crystallography. However, it is often possible to incorporate a small amount of a transition metal into the structure as a probe of local geometry. For example, silicate glasses containing a small quantity of Co^{2+} are colored blue, typical of tetrahedral coordination, indicating that these ions replace Si^{4+} in the glass network. Similarly, small amounts of Mn^{2+} , Co^{2+} , and Fe^{2+} incorporated into a ZnCl_2 glass show that in each case the absorption spectrum corresponds to that expected from tetrahedrally coordinated ions and indicates that the structure is formed from a random network of linked ZnCl_4 tetrahedral units.

Cross-References

- [Lanthanoid Ion Color](#)
- [Pigment, Ceramic](#)
- [Pigment, Inorganic](#)

References

1. Smith, D.W.: Ligand field theory and spectra. In: King, R.B. (ed.) *Wiley Encyclopedia of Inorganic Chemistry*, 2nd edn. Wiley, Chichester (2005)
2. Tilley, R.J.D.: Chapter 7. In: *Colour and the optical properties of materials*, 2nd edn. Wiley, Chichester (2011)
3. Nassau, K.: Chapter 4. In: *The physics and chemistry of colour*, 2nd edn. Wiley, New York (2001)
4. Schriver, D.F. Atkins, P.W. Langford, C.H.: Chapters 6, 14. In: *Inorganic chemistry*, 2nd edn. Oxford University Press, Oxford (1994)
5. Newman, D. J., Ng, B. (eds.): *Crystal Field handbook*. Cambridge University Press, Cambridge (2001)
6. Henderson, B., Bartram, R.H.: *Crystal-Field Engineering of Solid State Laser Materials*. Cambridge University Press, Cambridge (2000)
7. Schäfer, H.L., Gliemann, G.: *Basic Principles of Ligand Field Theory*. Wiley, New York (1969)
8. Figgis, B.N., Hitchman, M.A.: *Ligand Field Theory and Its Applications*. Wiley-VCH, New York (2000)

Transparency

Osvaldo Da Pos

Department of Psychology, University of Padua, Padua, Italy

Transparency/translucency: a. the degree of visibility of an object through a medium; b. the property of a material or substance by which objects may be seen through that material or substance. Transparency occurs when objects are clearly perceived through the medium, while translucency occurs when they appear hazy. Very often the term transparency is used instead of translucency.

The episotister model is suitable to describe transparency percept as it involves different areas separated by clear margins, while variations of this model are required to deal with translucency, due to blurred edges which add some degree of opacity to the impression of transparency.

First of all figural and topological conditions, not handled here, must be met: they essentially consist of figural unity of the transparent layer, continuity of the boundary line, and adequate stratification. Another strong factor which strengthens the figural organization is the movement, either of the background or of the transparent medium; this in turn makes transparency more evident. Transparency can also disambiguate objects (plaids) moving in different directions which otherwise would appear as a structure moving in another direction. Devices which induce strong depth perception, like natural or simulated stereopsis, also increase the evidence of the transparency effect.

Transparency involves the distinct perception of the rear and front colors along the same direction of sight well differentiated also where they are superimposed. This part of the definition is very restrictive as there are cases in which one sees objects behind and through another object, but the colors perceived in the overlapping area amazingly differ from those in plain view. In these

cases no color constancy is achieved, and no model seems presently applicable.

History of the Model

Models are known physical phenomena, with characteristics analogous to those to be described, whose explanatory structure is heuristically applied to not yet known phenomena (for instance, the atomic planetary model by Rutherford, the hydraulic analogy of electricity, and so on). A different definition of models includes just the mathematical formulation of relevant relationships which describe the phenomenon.

The episcotister model derives from experiences with rotating wheels, used to produce different colors by mixing two or more sources of radiation. Most times the disk is divided in two or more differently painted parts, and at a fusion speed temporal resolution of human eye cannot distinguish the different parts, but confuses them and sees only one color. This method of mixing radiation was made famous by Maxwell from whom it took the name. Sometimes instead of a solid disk, only two symmetrical sectors are spun so to mix rays from the background with those from the solid sectors. Although the laws governing these kinds of mixtures are well known, the actual perception does not always conform the prediction of seeing one resulting color, as sometimes two colors are instead seen in the overlapping area. This outcome happens when some figural conditions are fulfilled, and the figure of a transparent disk is visible in front of a background completely or partially covered and stratified in depth. This experience led to theoretical and experimental analysis of the psychological impression of transparency.

Field of Applications

Knowledge about transparency has been mainly applied to food and drink analysis, skin and cosmetics, medicine, dentistry, architecture, textile manufacturer, art, graphics, design, fashion, painting, printing, and others.

Overview

Metelli's Episcotister Model of Perceptual Transparency

Metelli's model [1] consists of a wheel, rotating at fusion speed in front of a background, divided in two parts: an empty sector (totally transmitting area) and a solid sector (reflecting area). It is a kind of special filter, like a riddled surface, which transmits and reflects light at the same time.

Through the open part of the episcotister (or through the holes of the riddled surface), light coming from the back object can reach the observer's eye without being spectrally modified. The size of the open sector (or of the holes in the riddled surface) determines how much light is transmitted to the observer: from a maximum when the whole disk is completely open to zero when the whole disk is solid. Therefore, the episcotister and the riddled surface behave like a neutral filter by decreasing the intensity of the light passing through them by a certain amount without changing its spectral distribution.

In turn the solid part of the episcotister (or the solid part of the riddled surface) can reflect light to the observer and this is mixed with the light from the back. As both the open and the filled sectors of the episcotister (or the holes and the solid part of the riddled surface) together cannot be larger or smaller than the whole disk (or surface), their proportion can vary from 0 to 1: if α is the proportion of the open sector to the whole disk, $(1-\alpha)$ is the proportion of the solid sector (or of the solid riddled surface). As one increases, the other decreases accordingly: for this reason the mixture of the two lights, that coming from the background and that coming from the surface of the episcotister (or riddled surface), is a special kind of mixture called partitive. If \mathbf{a} is the vector describing the light coming from the background not modified by a filter, and \mathbf{t} is the vector describing the light coming from the solid surface of the filled episcotister, the presence in the rotating disk of an open sector of size α determines the partitive mixture described by the following equation:

$$\mathbf{q} = \alpha \mathbf{b} + (1 - \alpha) \mathbf{t} \quad (1)$$

where \mathbf{q} is the vector describing the mixture of the two lights (\mathbf{a} and \mathbf{t}) arriving at the observer's eye.

This equation perceptually means that the back color \mathbf{b} is visible in the color \mathbf{q} proportionally to α , and the color corresponding to \mathbf{t} is visible in \mathbf{q} proportionally to $(1-\alpha)$. The reduced visibility of the two colors due to the transparency factor α means that they are not perceived as completely in plain sight, but one appears transparent and the other seen by transparency. Both the lights coming from the back opaque objects and the front transparent object arrive at the observer reduced in intensity but unchanged in their spectral composition. Therefore α means how much color of the background is visible in the overlapping area: this dimension corresponds to the degree of perceived transparency of the fore object and according to Metelli is a logarithmic function of it, while α is a neutral (achromatic) multiplicative factor. The additive factor \mathbf{t} , which can be interpreted as either the reflectance characteristic of the object or the light (luminance) reflected by it, corresponds to the color of the filter and can assume all values from 0 to 1 (black and white, respectively).

If the episcotister rotates over a uniform background, it appears like an opaque disk, and the same happens when physically transparent sheets are superimposed over a uniform surface. This effect shows that physical transparency is not sufficient condition to perceive transparency. Another particular case of transparency is obtained when a transparent sheet completely covers a background figure without sharing their margins with it: this effect has been little studied [2]. The case of riddled surface can be extended, by analogy, to textured surfaces, in which the density of the texture plays the role of color.

Equation 1 can be interpreted in two different ways, by analogy with a chemical reaction: by reading it from right to left, it means that the partitive mixture of the two colors \mathbf{b} and \mathbf{t} gives rise to the resulting and univocally defined color \mathbf{q} ; but if one reads it from left to right, the equation can describe a color scission, that is, the color \mathbf{q} can partitively split into the two components \mathbf{b} and \mathbf{t} . This scission, described by many researchers, would be in the model the opposite process of fusion and occurs in the perception of two colors in the same area, one in

front and the other in the back. According to a different wording, this scission would be a case of double representation in gestalt terms. In this latter case, however, an indefinite number of \mathbf{b} and \mathbf{t} pairs can be obtained because α is not determined. Therefore, at least two equations are necessary to derive a specific pair of α and \mathbf{t} , and a second background color seen through the transparent object must be present. For instance, in Fig. 1, the episcotister should not only rotate over the green background figure but also over the white background \mathbf{a} , and in this case another partitive mixture would be described by the following equation:

$$\mathbf{p} = \alpha^* \mathbf{a} + (1 - \alpha)^* \mathbf{t} \quad (2)$$

where \mathbf{p} is the vector describing the mixture of the two lights (\mathbf{a} and \mathbf{t}) arriving at the observer's eye.

The determination of α and \mathbf{t} , characteristics not present in the stimulation, can be univocally found, starting from the colors of the display in the bottom row of Fig. 1 according to the following equations:

$$\alpha = (\mathbf{p} - \mathbf{q}) / (\mathbf{a} - \mathbf{b}) \quad (3)$$

$$\mathbf{t} = ((\mathbf{a} \mathbf{q}) - (\mathbf{b} \mathbf{p})) / (\mathbf{a} + \mathbf{q} - \mathbf{p} - \mathbf{b}) \quad (4)$$

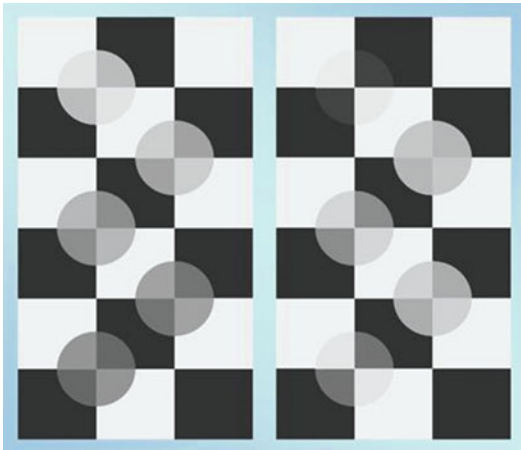
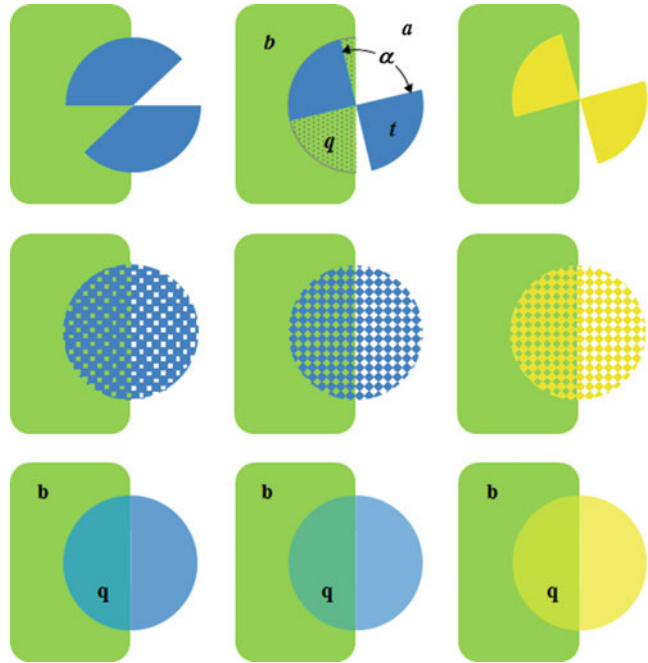
Once the \mathbf{a} (the lightest area which is perceived in the background) and \mathbf{b} background colors and the \mathbf{t} color of the transparent virtual object have been chosen and the degree α of transparency is decided, the regions of the mosaic of Fig. 2 are filled with the colors determined according to Eqs. 1 and 2.

The colors of the transparent disks and their degree of transparency can be computed according to Eqs. 3 and 4 from the colors of Fig. 2. The formal independence of α and \mathbf{t} is a feature of the model. The square wave margins between colors in the overlapping areas are characteristics of the transparency effect, while blurred margins denote translucency [2, 3]. The alpha blending procedure in computer graphics is based on this model [4].

The model of the episcotister applied to transparency phenomena entrains also some constraints and consequences. The most important is that α [4] cannot be higher than 1 (an open sector larger than

Transparency,

Fig. 1 Top row: three different episcotisters over three identical backgrounds. b = vector defining the color of the background; t = vector defining the color of the solid sector of the episcotister; q = vector defining the color of the overlapping area; α = a measure of the open sector of the episcotister, which can vary from 0 to 1. Middle row: three riddled circular surfaces. The ratio holes/total surface (α) can vary from 0 to 1. Bottom row: the perceptual effects



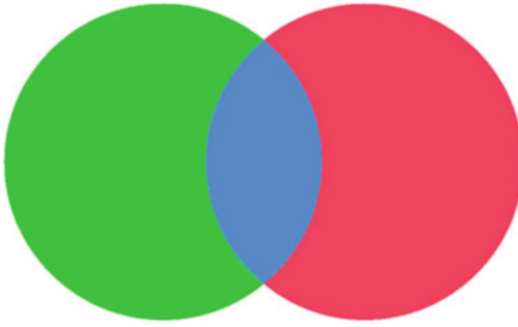
Transparency, Fig. 2 Transparent gray disks of different t color (at left) and different degrees (α) of transparency (at right) over a white and black chessboard

$a-b$: the contrast (here measured as a color difference) between the colors covered by the transparent object is always lower than the contrast between the two background colors in plain view. In other words the transparent object reduces the color contrast in the overlapping area. Secondly, as α must be positive (a negative open sector is a nonsense), the color difference between p and q must go in the same direction of the color difference between a and b . If these conditions are not met, usually transparency cannot be perceived and the model cannot be applied, i.e., nothing can be stated about the possible transparency effect on this basis.

Transparency and Color Constancy

Transparency, like illumination, involves color constancy: first, the colors of the background perceived through the transparent medium should appear the same as those in plain view, and secondly, the color of the transparent medium should also appear the same in all its parts. The analogy with illumination is justified by the often quoted example of a transparent medium as a spectrally selective filter. Models based on simple filters

360° is inconceivable) and lower than 0 (and likewise is inconceivable an open sector smaller than 0°). This involves that, by assuming a as the vector of the lightest color in the background and b of the other background color (both perceived as such), the difference $p-q$ must be smaller than the difference



Transparency, Fig. 3 An example of impression of transparency without color constancy (from Metzger). When the disks are rotating, the impression is much stronger

have been shown not to work because the colors of the overlapping regions can be quite unexpected and sometimes color constancy does not occur at all: for instance, through a red filter a red surface appears white and a green surface appears black. Therefore, filter models usually include either the need of previous experience or some constancy constraints. A model based on the illumination analogy [5] considers the cone excitations as the basis for describing the necessary color conditions of an impression of transparency. This can occur when cone-excitation ratios between the colors in plain view (a, b, \dots) remain the same between the colors (p, q, \dots), i.e., under the transparent medium. Although color constancy is a strong requirement for perceiving transparency, it can be very weak or even null and nevertheless a generic impression of transparency is still possible (Fig. 3). This makes the study or perceptual transparency quite difficult [6].

Transparency and Contrast

The model specifies a, b, t, p, q by the achromatic reflectance characteristics of the surfaces (objects, materials, media) and describes the reduced contrast in the superimposition area by a reflectance difference, while further developments of the model specify colors by lightness Munsell values [7], by luminance [8], or by log luminance [9]. On the other side, there is no unique shared measure of contrast, and sometimes one can use the Michelson contrast

[3] which seems to supply a better fitting model, suitable to explain why usually light filters seem more opaque and dark filters more transparent, although this model too is not general [10]. In any case the filter reduces the contrast of the colors seen behind it because as they get closer, the more dense (or less transparent) is the filter [4]. Apparent contrast is further reduced by blurred edges [2].

Color contrast is relevant in determining the relative stratification of the surfaces, i.e., what is seen in front and what behind, as higher contrast results in larger distance in depth. Contrast does not depend only on color difference but also on the shape of the margins: Craik-O'Brien-Cornsweet edges (blurred versus sharpened) can invert the contrast appearance and as a consequence the relative depth of the surfaces, making transparent what was opaque and the reverse [11]. This effect can be obtained also by simple line drawings, where their contrast with the background determines the different object stratifications and the consequent transparent appearance.

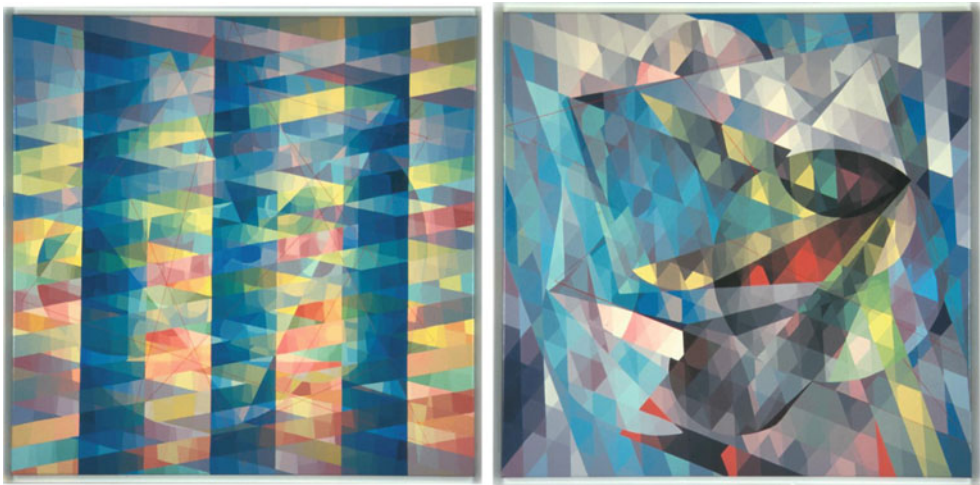
A Qualitative Account of the Model in the Achromatic Domain

The possibility of perceiving transparency can be described as a function of the ordinal position of colors in the regions divided by X junctions [12, 13]. If all colors present in a transparent display are achromatic, and therefore characterized by one dimension, two ordinal conditions of transparency are prescribed by the model: $|a-b| > |p-q|$ and $p > q$ if $a > b$ (or $p < q$ if $a < b$). A better prediction of transparency is achieved with the single ordinal condition: $p \in (a, q)$ (or $q \in (q, b)$) [14].

Another improvement of the model states that, instead of Eq. 3, the apparent density ψ (inverse of transparency) of the filter [15] depends on three color differences:

$$\psi = w_1|a - p| + w_2|b - q| + (1 - w_1 - w_2)(k - |p - q|) \quad (5)$$

where the weight coefficients w_1 and w_2 express a compromise between the three differences.



Transparency, Fig. 4 Attilio Taverna. *Left*: Sentieri interrotti 1992. *Right*: Blue geometric storm 1992. Transparency effects are obtained by painting each overlapping

area with the partitive mixture of the back and front colors. Also perceptual space organization is affected by transparency



Transparency, Fig. 5 Ravenna: Baptistry of Arians, fifth century (detail). Picture by Incola (Wiki) [http://it.wikipedia.org/wiki/File:Soffitto_Battistero_Ariani_Ravenna.jpg#filehistory]

Water and body colors seen from some distance partitively fuse because under spatial discrimination threshold, an impression of transparency arises in the observer

Transparency and Pictorial Art

Since centuries the transparency effect is used both to represent translucent or transparent objects and to enrich the color variety of a painting.

Many methods have been adopted to render transparent objects, all having in common the

purpose of showing two colors, one in front and one in the back, the latter visible through the former. One method is just to put in the overlapping areas the color which would be obtained by partitive mixture of the back and the front colors (Fig. 4), following, not necessarily consciously, the episcotister model.



Transparency, Fig. 6 Hans Memling, fifteenth century. Portrait of a young lady (Sibylla Sambetha). Musea Brugge, Sint-Janshospitaal. Copy purchased from Lukas Art in Flanders, o.n. 1376 (2012)

Another method reaches the same results by using small mosaic tesserae to be looked at a certain distance so to produce spatial color mixtures (Fig. 5).

Another method is the veiling procedure obtained through semitransparent paints, as shown in Fig. 6: the back and the transparent colors are both visible where they overlap. Lastly, a final method consists in accentuating the borders between the opaque and the transparent medium, by simulating Cornsweet edges and introducing a depth displacement between opaque and transparent surfaces.

Improvements of the Model

As the colors in plain view and those covered by the transparent medium, represented in suitable color system, according to the model converge in a point corresponding to the color t , the model has been called convergence model [4]. The convergence point can lie outside the color solid and still perception of transparency can be possible [6], although not always [16]. In this partitive model, α defines the position of the color p

Transparency, Fig. 7 In the upper part the external areas represent, in isolation, the background in plain view and the internal areas the overlapping surfaces, whose colors have been determined in the NCS. The bottom part shows the complete mosaic resulting when the separated areas are joined



between the color *a* and the color *t*. According to current opinion, the colorimetric distance between the two colors *a* and *t* should not be large; otherwise, transparency perception would not occur. The reason is that [17] the color *p* should not be just a colorimetric intermediate between *a* and *t*, but should appear similar both to *a* and *t*, and this would be a strong condition to see both colors in the overlapping area: opponent *a* and *t* would not allow transparency. A development of the model has been proposed [17] in which the Natural Color System [18], based on the pure perceptive similarity of all color to the six Hering prototypical unique colors W K Y R B G, should be used to describe the relevant color relationships in transparency perception (Fig. 7). The notation of colors in this system clearly shows similarity relationships and therefore the possibility of seeing two colors in a not-unique (i.e., perceptually mixed) color; spatial induction can sometimes add the necessary missing shades (especially if *p* is a neutral color).

Cross-References

- [Anchoring Theory of Lightness](#)
- [Color Constancy](#)
- [Color Contrast](#)
- [Color Phenomenology](#)
- [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)
- [Hering, Karl Ewald Konstantin](#)
- [Maxwell, James Clerk](#)
- [Mosaics](#)

References

1. Metelli, F.: The perception of transparency. *Sci. Am.* **23**, 90–96 (1974)
2. Ekroll, V., Faul, F.: Transparency perception: the key to understanding simultaneous color contrast. *J. Opt. Soc. Am. A* **30**, 342–352 (2013)
3. Singh, M., Anderson, B.L.: Photometric determinants of perceived transparency. *Vision Res.* **46**, 879–894 (2006)
4. D'Zmura, M., Rinner, O., Gegenfurtner, K.R.: The colors seen behind transparent filters. *Perception* **29**, 911–926 (2000)
5. Ripamonti, C., Westland, S., da Pos, O.: Conditions for perceptual transparency. *J. Electron. Imaging* **13**, 29–35 (2004)
6. Richards, W., Koenderink, J.J., van Doorn, A.: Transparency and imaginary colors. *J. Opt. Soc. Am. A* **26**, 1119–1128 (2009)
7. Beck, J.: Additive and subtractive color mixture in color transparency. *Percept. Psychophys.* **23**, 256–267 (1978)
8. Gerbino, W., Stultiens, C.I.F.H.J., Troost, J.M., de Weert, C.M.M.: Transparent layer constancy. *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 3–20 (1990)
9. Vladusich, T.: A reinterpretation of transparency perception in terms of gamut relativity. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **30**, 418–426 (2013)
10. Masin, S.C., Tommasi, M., da Pos, O.: Test of the singh-anderson model of transparency. *Percept. Mot. Skills* **104**, 1367–1374 (2007)
11. da Pos, O., Devigili, A., Giaggio, F., Trevisan, G.: Color contrast and stratification of transparent figures. *Jpn. Psychol. Res.* **49**, 68–78 (2007)
12. Adelson, E.H., Anandan, P.: Ordinal Characteristics of Transparency, pp. 77–81. AAAI-90 Workshop on Qualitative Vision, July 20, Boston (1990)
13. da Pos, O., Burigana, L.: Qualitative inference rule for perceptual transparency. In: Albetazzi, L. (ed.) *Handbook of Experimental Phenomenology: Visual Perception of Shape, Space and Appearance*, pp. 369–394. Wiley Blackwell, Chichester (2013)
14. Masin, S.C.: The luminance conditions of transparency. *Perception* **26**, 39–50 (1997)
15. Masin, S.C., Gardonio, G.: The evaluation of the apparent density of a filter on a bicolored background. *Percept. Psychophys.* **37**, 103–108 (1985)
16. Brill, M.H.: The perception of a colored translucent sheet on a background. *Col. Res. Appl.* **19**, 34–36 (1994)
17. da Pos, O.: *Trasparenze. Transparency. Icone*, Milano (1989)
18. NCS Colour Atlas: SIS Standardiserings Kommissionen I Sverige. Stockholm, 2nd ed. (1989) SS 01 91 02

Trichromacy

- [Comparative Color Categories](#)

Tristimulus Colorimeter

- [Instrument: Colorimeter](#)

Tritanopia

David L. Bimler¹ and Galina V. Paramei²

¹School of Psychology, Massey University, Palmerston North, New Zealand

²Department of Psychology, Liverpool Hope University, Liverpool, UK

Synonyms

Autosomal dominant inherited color vision deficiency; “Blue blindness”; Blue-yellow deficiency; Color blindness; Color vision abnormality; Dichromacy

Definition

Tritanopia (from the Greek *tritos*: third + *an*: not + *opia*: a visual condition) is a congenital form of severe color deficiency (dichromacy) affecting the blue-yellow opponent color system. Tritanopes do not distinguish between colors along a specific direction in color space (tritanopic confusion lines) and are able to match all colors using two primaries (unlike trichromats who require three primaries). Tritanopia follows a dominant pedigree pattern and is linked to chromosome 7, where damage to the gene encoding the photopigment opsin results in the absence of functioning S-cones, cone cells containing the short-wavelength (S-) photopigment. Tritanopia is rare, not sex-linked, with great uncertainty as to its incidence, although some estimates place this at 1 in 13,000 or fewer in the Caucasian population. The condition can be partial rather than complete, i.e., tritanopia lies at the extreme of a continuum of tritanomaly in which some capacity remains to discriminate colors along confusion lines.

Retinal Mosaic

At the retinal level, tritanopia is an absence of functional S-cones (► [CIE Physiologically](#)

[Based Colour Matching Functions and Chromaticity Diagrams](#)), i.e., cone cells expressing a photopigment containing an S-opsin component, tuned for selective sensitivity to the short wavelengths of blue light [1, 2]. These cones comprise only about 8 % of the retinal mosaic. Even in normal trichromats, they are absent from the central region of the fovea, resulting in “foveal tritanopia” affecting the perception of sufficiently small stimuli when they are fixated [3, 4].

Congenital Tritanopia: Genetic Background

Functional loss of the S-cone photoreceptor cells is caused by point mutation in the gene on chromosome 7 that codes for the S-opsin component of the cone photopigment (sensitive to shorter wavelengths) (► [Genetics of photoreceptors, genetics and color vision deficiencies, genes of cone photopigments, genes and cones](#)). The resulting amino acid substitution may prevent the S-opsin from functioning, or it may fold into an aberrant conformation, blocking it from being transported to the ciliary segment of S-cone cells [1, 5].

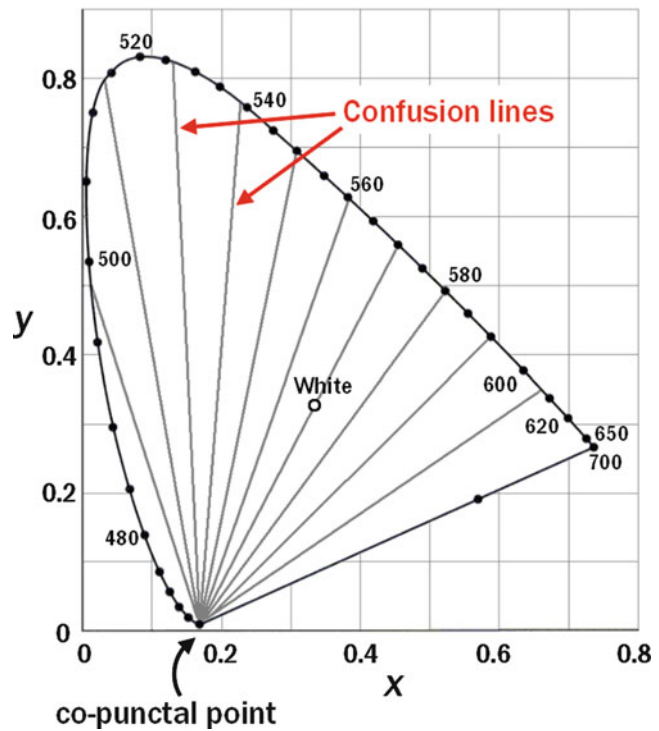
The condition follows a dominant pattern of inheritance: S-cones must contain two undamaged copies of the gene to function properly [6, 7]. However, pedigrees often display variations in penetrance with incomplete tritanopia in some family members who retain some blue-yellow discrimination.

Variation in phenotype may also indicate that loss of S-cones is progressive, and incomplete, in some family members. This is supported by retinal imaging showing S-cone degeneration subsequent to formation of the retina [8].

Disease processes and exposure to environmental toxins can lead to the same end point, i.e., acquired tritanopia (► [Color Perception and Environmentally Based Impairments](#)). In particular, the phototoxicity of prolonged exposure to intense short-wavelength light can cause S-cones to degenerate. In other forms of acquired tritanopia, the damage is localized elsewhere in the processing of the S₀ signal, such as the ganglion cells or the optic nerve.

Tritanopia,

Fig. 1 *Confusion lines* indicate colors that cannot be distinguished by a tritanope. The line going through the *white point* indicates neutral (achromatic) colors. The lines converge at the copunctal point defining the chromaticity of the missing S-cone fundamental. Drawn in the CIE 1931 chromaticity diagram. (After: Benjamin, W.J. (ed.), *Borish's Clinical Refraction* (1998), Fig. 20) (Elsevier Books, Copyright Clearance Center: Licensee: David Bimler; License Number: 3532171403672)

**Confusion Lines**

Tritanopic color processing lacks the S-(L+M) or S_0 signal, produced by bistratified retinal ganglion cells comparing the S-cell signal and the combined L- and M-signals (► [Color Vision, Opponent Theory](#)); hence, there is no difference in appearance among colors which differ only in how much they stimulate S-cones. Examples of colors confused by tritanopes are blue and green or violet, gray, and yellow. Plotted in a chromaticity diagram (► [CIE Chromaticity Diagrams, CIE purity, CIE dominant wavelength](#)), the confused colors lie on straight “confusion lines” that converge at the tritanope copunctal point (Fig. 1).

A related phenomenon is observed in edge perception, where the border between two regions of equiluminant color dissolves if the colors lie on a tritanopic confusion line (e.g., yellow and gray or green and blue). The perception of border distinctness is tritanopic, with no contribution from the S_0 signal, as a consequence of the sparseness of S-cones and chromatic aberration of the lens [9].

Wavelength Discrimination

Because S-cones are insensitive to mid- and long-wavelength lights, tritanopic wavelength discrimination, as a function of wavelength λ , is similar to the normal trichromat function for $\lambda > \text{ca. } 530 \text{ nm}$, with the discrimination threshold $\Delta\lambda$ reaching a minimum at about 570–580 nm near the tritanopic “neutral point” (see below). Tritanopic discrimination deteriorates for blue-green hues where the L- and M-cones are equally sensitive, with $\Delta\lambda$ peaking at ca. 470 nm, then improving at shorter wavelengths [10].

Luminous Efficiency

S-cones normally contribute little or nothing to perceived luminance, but S-cone signals can make a contribution under certain conditions of adaptation and temporal flicker. Specifically, with an intense long-wavelength background, S-cone stimulation exerts a negative but phase-delayed influence on luminance, as if inhibiting the combined L- and M-cone signal [11, 12].

Color Perception

A tritanope recognizes far fewer distinct colors than a trichromat (► [Color categorization and naming in Inherited Color Vision Deficiencies](#)). When naming spectral colors, tritanopes experience one “neutral zone” of subjective gray for yellow light, with midpoint at λ ca. 570 nm, and another for short-wavelength violets, at about $\lambda = 410\text{--}420$ nm [13]. For a tritanope, all equiluminant colors, instead of being arranged in a two-dimensional plane (► [Psychological Color Space and Color Terms](#)), form a single-dimensional line, running from “cool” to “warm” colors between the two extremes of saturated green and saturated red. This single dimension combines the usually separate chromatic qualities of “saturation” and “hue.” Tritanopes distinguish approximately 44 steps along this line, each corresponding to a separate band of interchangeable colors [1]. This tritanopic linear representation of equiluminant colors is part of a broader reduction of a three-dimensional into a two-dimensional color space.

The subjective qualities of tritanopic experience remain unclear. Tritanopes do use the words “blue” and “yellow” when describing lights and surface colors [4, 10]. Dichromacy simulation software, which converts color scenes to show how a tritanope would see these, uses the convention of a red-green scale [14].

Diagnosis of Tritanopia

Acquired blue-yellow deficits can be symptomatic of illness or chemical exposure (► [Color Perception and Environmentally Based Impairments](#)). Thus, standard tests of color abnormality, although targeted primarily at congenital red-green deficiency, do not ignore tritanopia completely. (► [Pseudoisochromatic plates](#)) and arrangement/panel tests both address a tritanope’s perceived similarity of colors along the confusion axis.

The Hardy-Rand-Rittler test (► [Color Vision Testing](#)) contains pseudoisochromatic diagnostic plates to detect tritanopia and gauge its severity, while the Farnsworth F2 plate is designed specifically for the purpose. (The widely used Ishihara series contain no plates targeting tritanopia.) Each

pattern presented in these plates is defined by a color difference that disappears for tritanopes, so that the design vanishes or is supplanted by an alternative design, demarcated by a different chromatic distinction.

In the arrangement tests (Farnsworth-Munsell 100 Hue test, Farnsworth D-15 test), the one dimensionality of a tritanope’s color “plane” collapses a color circle (► [Color Circle](#)) into a line, causing drastic departures from a normal observer’s sequence. Errors, or cap transposition “distances,” are summed to quantify the severity of tritanopia. Plotted graphically, the transpositions in the D-15 manifest as diametrical circle crossings, indicating the tritanopic angle of the confusion axis (► [Color Perception and Environmentally Based Impairments](#)).

The “gold standard” of diagnosis of tritanopia is the Moreland equation, presented in a small viewing field (2°) in the Moreland anomaloscope [15], in which a cyan standard (480 nm with a small admixture of 580 nm) is matched by mixing indigo (436 nm) and green lights (490 nm). Decreasing discrimination along the tritanopic confusion lines increases the range of acceptable mixtures. Notably, when field size is increased beyond 1° , residual blue-yellow discrimination can be detected in some cases; at 8° only complete tritanopes accept the full range of color mixtures as a match to the cyan standard [7].

More recently developed computerized tests for color vision diagnosis, the Cambridge Colour Test and Colour Assessment and Diagnosis test (► [Color Vision Testing](#)), interactively quantify any loss of chromatic sensitivity along the tritanopic confusion line. For a tritanope, in the frame of the CIE ($u'v'$) 1976 chromaticity diagram (► [CIE \$u',v'\$ Uniform Chromaticity Scale Diagram and CIELUV Colour Space](#)), a MacAdam ellipse of indistinguishable colors displays a major axis lengthened to infinity and orientation pointing to the tritanope copunctal point.

Cross-References

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

- [CIE u', v' Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [CIE Chromaticity Diagrams, CIE purity, CIE dominant wavelength](#)
- [Color Circle](#)
- [Color Categorization and Naming in Inherited Color Vision Deficiencies](#)
- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision Testing](#)
- [Color Vision, Opponent Theory](#)
- [Pseudoisochromatic Plates](#)
- [Psychological Color Space and Color Terms](#)

References

1. Sharpe, L.T., Stockman, A., Jägle, H., Nathans, J.: Cone genes, cone pigments, color vision and color blindness. In: Gegenfurtner, K., Sharpe, L.T. (eds.) *Color Vision: From Genes to Perception*, pp. 3–52. Cambridge University Press, Cambridge (1999)
2. Stockman, A., Sharpe, L.T.: Human cone spectral sensitivities and color vision deficiencies. In: Tombran-Tink, J., Barnstable, C.J. (eds.) *Visual Transduction and Non-Visual Light Perception*, pp. 307–327. Humana Press, Totowa (2008)
3. Williams, D.R., MacLeod, D.I.A., Hayhoe, M.M.: Foveal tritanopia. *Vision Res.* **21**, 1341–1356 (1981)
4. Mollon, J.D.: A taxonomy of tritanopias. In: Verriest, G. (ed.) *Color Vision Deficiencies VI. Documenta Ophthalmologica Proceedings Series*, vol. 33, pp. 87–101. Dr. W. Junk Publisher, The Hague (1982)
5. Neitz, J., Neitz, M.: The genetics of normal and defective color vision. *Vision Res.* **51**, 633–651 (2011)
6. Wright, W.D.: The characteristics of tritanopia. *J. Opt. Soc. Am.* **42**, 509–521 (1952)
7. Pokorny, J., Smith, V.C., Went, L.N.: Color matching in autosomal dominant tritan defect. *J. Opt. Soc. Am.* **71**, 1327–1334 (1981)
8. Baraas, R.C., Carroll, J., Gunther, K., Chung, M., Williams, D.R., Foster, D.H., Neitz, M.: Adaptive optics retinal imaging reveals S-cone dystrophy in tritan color-vision deficiency. *J. Opt. Soc. Am. A* **24**, 1438–1447 (2007)
9. Tansley, B.W., Boynton, R.M.: A line, not a space, represents visual distinctness of borders formed by different colors. *Science* **191**, 954–957 (1976)
10. Alpern, M., Kitahara, K., Krantz, D.H.: Classical tritanopia. *J. Physiol.* **335**, 655–681 (1983)
11. Stockman, A., MacLeod, D.I.A., DePriest, D.D.: The temporal properties of the human short-wave photoreceptors and their associated pathways. *Vision Res.* **31**, 189–208 (1991)
12. Ripamonti, C., Woo, W.L., Crowther, E., Stockman, A.: The S-cone contribution to luminance depends on the M- and L-cone adaptation levels: silent surrounds? *J. Vis.* **9**(3), 10 (2009). doi:10.1167/9.3.10
13. Smith, D.M.: Color naming and hue discrimination in congenital tritanopia and tritanomaly. *Vision Res.* **13**, 209–218 (1973)
14. Brettel, H., Viénot, F., Mollon, J.D.: Computerized simulation of color appearance for dichromats. *J. Opt. Soc. Am. A* **14**, 2647–2655 (1997)
15. Moreland, J.D., Young, W.B.: A new anomaloscope employing interference filters. *Mod. Probl. Ophthalmol.* **13**, 47–55 (1974)

Tubular and Compact Fluorescent Lamp

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Low-pressure mercury gas discharge lamps](#)

Definition

Lamp that produces light as a result of an electrical discharge, generated between two electrodes, in a low-pressure mercury vapor that is contained in a transparent tube whose inside is coated with fluorescent powder that converts the ultraviolet part of the emitted radiation from the discharge in visible light.

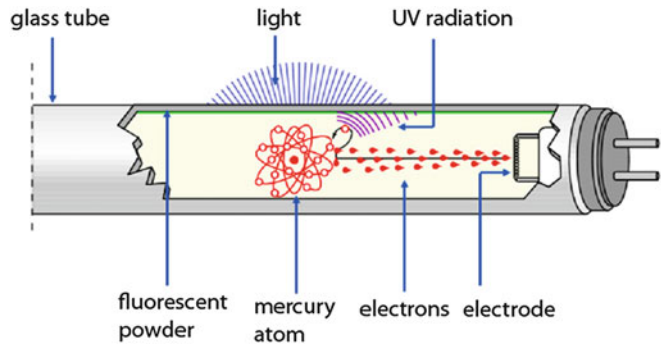
Types of Tubular Fluorescent Lamps

Fluorescent lamps belong to the family of low-pressure mercury gas discharge lamps. They are by far the most widespread discharge lamp types. They are available in both tubular and compact versions. Although the operating principle of

Modified and reproduced from [1] by permission of Philips Lighting, Eindhoven, The Netherlands. © 2012 Koninklijke Philips Electronics N.V.

Tubular and Compact Fluorescent Lamp,

Fig. 1 Main parts and principle of operation of a tubular fluorescent gas discharge lamp (TL) [1]



compact fluorescent lamps is largely the same as that of the tubular version, their construction and performance are in many ways different from that of the tubular version [2, 3].

Working Principle

The discharge tube of a fluorescent lamp is filled with an inert gas and a little mercury and has an electrode sealed into each end (Fig. 1). To facilitate starting, the electrodes of most fluorescent lamps are preheated prior to ignition, which is accomplished by means of a high-voltage pulse, generated by an external device called the igniter or starter. (Igniters are available as simple glow-switch starters or as electronic devices.) When the lamp is switched on, the electrodes begin emitting electrons, and through the collision of these electrons with the gas atoms, the ionization process starts. The inert gas is then heated up and the mercury inside the lamp is completely evaporated to give a mercury vapor pressure of about 0.8 Pa ($8 \cdot 10^{-6}$ atm). The emitted electrons collide with and excite the mercury atoms, resulting in the emission of ultraviolet radiation and a small amount of blue visible light. The inside of the discharge tube is coated with a mixture of fluorescent powders. The ultraviolet radiation is converted to visible light when it passes through the fluorescent powder coating.

Like almost all gas discharge lamps, a fluorescent lamp cannot be operated without some device to limit the current flowing through it. This device, usually in the form of an inductive coil, is called a ballast. Fluorescent lamps also need, again like

most discharge lamps, an igniter for starting the lamp.

Tubular Fluorescent Lamps

Tubular fluorescent lamps are widely used in offices, schools, shops, and low-ceilinged industrial premises. So today their use is mainly indoors.

Materials and Construction

The main parts of a tubular fluorescent lamp are (see Fig. 1):

- Glass tube
- Fill gas
- Electrodes
- Fluorescent powder

Glass Tube

The tube of a normal fluorescent lamp is made of glass that is doped with a special material that blocks that UV radiation from the mercury discharge that is not converted by the fluorescent powder into visible light.

The original fluorescent lamp had a diameter of 38 mm. This tube diameter is usually characterized as T12, where the 12 stands for 12 times one-eighth of an inch (Table 1). This type of fluorescent tube is seldom seen today. During the late 1980s, a smaller and more efficient version was introduced that has a diameter of 26 mm. Today an even thinner and more efficient version

Tubular and Compact Fluorescent Lamp, Table 1 Fluorescent lamps with different tube diameters and their designations

Tube type	Diameter (inch)	Diameter (mm)
T12	12 * 1/8	38
T8	8 * 1/8	26
T5	5 * 1/8	16
T2	2 * 1/8	6
T1	1 * 1/8	2.8

has become the standard: the TL5 with a diameter of 16 mm. Even thinner fluorescent lamps are produced, but these are not for use in general lighting.

Fill Gas

The gas filling in a fluorescent lamp consists of a mixture of mercury vapor and an inert gas. The inert gas has three functions:

- To facilitate ignition, especially at lower temperatures
- To control the speed of the free electrons
- To prolong the life of the electrodes by reducing evaporation of electrode material

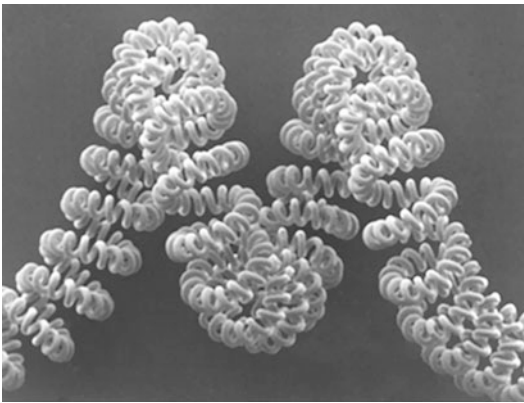
The inert gas usually consists of a mixture of argon and neon, although sometimes krypton is used as well.

Electrodes

The function of the electrodes is to provide free-running electrons, which are necessary to start and maintain the discharge. A fluorescent lamp electrode consists basically of a tungsten filament (Fig. 2) that is coated with the so-called emitter material to facilitate electron emission.

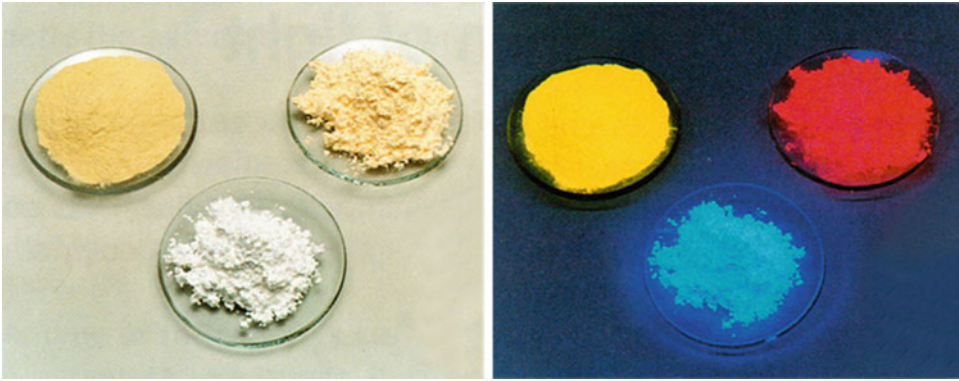
Fluorescent Powder

The small crystals of the fluorescent powder applied to the inside of the discharge tube absorb the UV mercury radiation and convert it into visible light (this physical phenomenon is called luminescence). Different fluorescent powders convert the ultraviolet radiation into visible light of different wavelengths and thus different colors (Fig. 3).



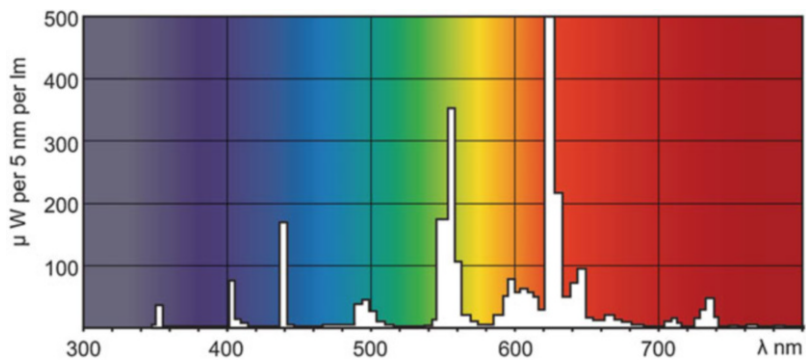
Tubular and Compact Fluorescent Lamp, Fig. 2 Triple-coiled filament electrode [1]

By mixing different fluorescent powders in different proportions, lamps producing different tints of white light can be made. The type and composition of the fluorescent powder is the most important factor determining the light characteristics of a fluorescent lamp, such as color temperature, color-rendering index (R_a), and to a large extent the luminous efficacy of the lamp (lm/W). Some fluorescent powders convert the ultraviolet radiation into wavelengths covering almost the whole visible spectrum. Such powders therefore produce white light when used alone. However, their color rendering and efficacy are poor ($R_a < 70$ and $\text{lm/W} < 80$). Because these powders lose some of their conversion activity relatively quickly, lamp-lumen depreciation is relatively large. Nowadays these lamps are hardly ever produced. Today fluorescent lamps often employ a mixture of three fluorescent powders, each having a very narrow spectrum band in red, green, and blue. In this way, white light is again obtained, but of better color rendering and efficacy ($R_a > 80$, lm/W up to 105). The lamp-lumen depreciation with these powders is very low ($< 10\%$). For cases where extremely good color rendering is required, a mixture of more than three powders is used, resulting in lamps with excellent color rendering ($R_a > 90$), very low lumen depreciation, and high luminous efficacy (slightly smaller than the previous version, viz., up to some 90 lm/W).



Tubular and Compact Fluorescent Lamp, Fig. 3 Different types of fluorescent powder. *Left*, under white light, and *right*, the same fluorescent powders under UV radiation [1]

Tubular and Compact Fluorescent Lamp, Fig. 4 Spectral energy distribution of a fluorescent lamp colortype 827: T_k 2,700 K and R_a 80 [1]



Properties

Energy Balance

Just below 30 % of the input power is converted into visible radiation and a very small part into UV radiation. The rest is lost in the form of heat (at the electrodes, in the discharge itself and as infrared radiation). Compare the figure of 30 % with 8 and 12 % for incandescent and halogen lamps, respectively.

System Luminous Efficacy

As with most lamps, the luminous efficacy of tubular fluorescent lamps is dependent on the wattage of the lamp and the color quality of the light it gives. Types with color rendering better than R_a 80 and with low lumen packages have a system luminous efficacy starting at 50 lm/W, while lamps with higher lumen packages reach efficacies up to some 105 lm/W. Lamps with

extremely good color rendering ($R_a > 90$) have 15 % lower efficacies.

Lumen-Package Range

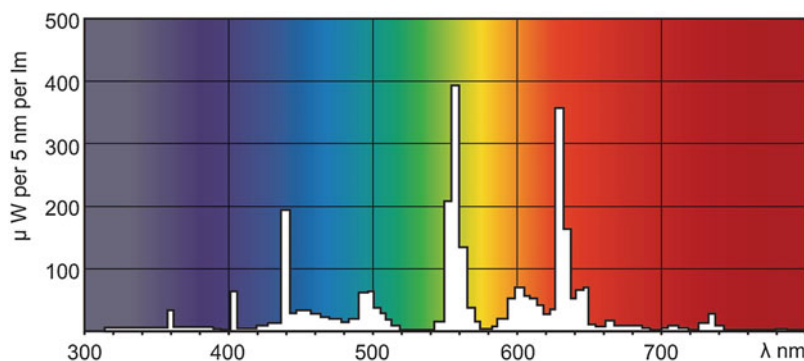
Common tubular fluorescent lamps are produced in the range from some 500–6,000 lm (corresponding wattage range approximately 8–80 W). Special very-high-output lamps are also produced in versions up to some 9,000 lm (120 W).

Color Characteristics

As has been mentioned, by mixing different fluorescent powders in different proportions, lamps with different spectra can be produced. As with all gas discharge lamps, the spectrum is discontinuous. Figures 4, 5, and 6 show the spectra of fluorescent lamps with different color temperatures T_k and color-rendering index R_a .

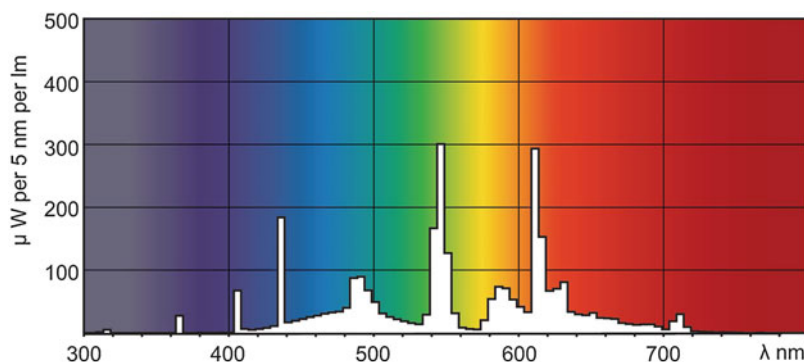
Tubular and Compact Fluorescent Lamp,

Fig. 5 Spectral energy distribution of a fluorescent lamp colortype 840: T_k 4,000 K and R_a 80 [1]



Tubular and Compact Fluorescent Lamp,

Fig. 6 Spectral energy distribution of a fluorescent lamp colortype 940: T_k 4,000 K and R_a 90 [1]



The colortype designation used for fluorescent lamps is standardized, with the first digit standing for the color-rendering index R_a and the last two digits for the color temperature T_k . Thus, the colortype 840 stands for a color-rendering index value R_a in the 80-ties and a color temperature of around 4,000 K.

Present-day quality fluorescent lamps are produced in a whole range of color temperatures varying from 2,700 K (warm-white or incandescent lamp color tint) to 6,000 K (bluish white), most of them in two color-rendering qualities with R_a in the 80-ties (800 series) and 90-ties (900 series), respectively. For special applications, versions are produced with extremely high color temperatures (up to some 17,000 K) with color rendering in the 80-ties. Today, fewer and fewer of those fluorescent lamps with poor color-rendering qualities (viz., R_a around 65 or less) that were produced in the past are produced.

Lamp Life

During the operation of a lamp, the electrodes lose emitter material due to evaporation and as a result

of the bombardment with ions from the discharge. This is the main cause of final lamp failure: the lamp will no longer start, due to a broken electrode, or else it flickers because insufficient emitter material remains. The lifetime of fluorescent lamps and gas discharge lamps in general is thus very much dependent on the construction and materials of the electrodes. Conditions that influence lifetime are principally the type of gear used, the switching frequency, and the ambient temperature. The switching frequency plays a role because the high-voltage peak needed for ignition causes the electrodes to lose some of their material (sputtering). The more accurately the igniter-ballast combination preheats the electrodes, the less severe this effect is. The occurrence of such things as shocks or vibrations, different burning positions, and supply-power variations can play an additional role.

Today's high-quality fluorescent tubes have economic lifetimes (based on 20 % mortality) of 15,000 h to more than 20,000 h if used on a so-called high-frequency or HF ballast and around 12,000 h when used on an electromagnetic ballast.

For situations where the actions needed to replace lamps are expensive, special, more expensive, long-life fluorescent lamp versions are available with lifetimes ranging from some 40,000–65,000 h. This long life is obtained by means of a special electrode design and construction.

Lamp-Lumen Depreciation

The main cause of lamp-lumen depreciation in a fluorescent lamp is that the fluorescent powder slowly becomes less active as a result of chemical attack by mercury ions. There may also be some blackening of the tube wall from the electrodes. For cool-white and cool-daylight lamp types, the lumen depreciation is higher than for lamps with a warmer tint. This is because of the faster depreciation of the blue fluorescent powders. Lamp-lumen depreciation for present-day high-quality fluorescent tubes is around 10 % after some 20,000 h.

Run-Up and Reignition

Fluorescent lamps operated on present-day electronic control gear start very quickly (within 1.5 s) and without flickering. This last is in contrast to lamps used on the older, glow-switch starters. After ignition, the light output increases quickly up to its maximum in about 1 min. When a fluorescent lamp is switched off, the vapor pressure drops so quickly that reignition is instantaneous.

Switching

It has already been mentioned that the switching frequency has an influence on lifetime, the influence being dependent on the type of electrical control gear used. Lamps operated with HF electronic-preheat-start ballasts show in general little sensitivity to the switching cycle. This is because of the well-controlled starting conditions of the lamp (warm start). HF non-preheat starting leads to a relatively short lamp life under frequent-switching conditions.

Dimming

Dimming of modern fluorescent lamps on high-frequency electronic ballasts down to 3 % of the nominal light-output value is easy, possible with

an additional circuit that changes the operating frequency.

Ambient-Temperature Sensitivity

The luminous flux of a fluorescent lamp is determined by the mercury vapor pressure during operation. The mercury vapor pressure is in turn determined by the coldest spot in the tube, normally the tube wall, which of course is also dependent on the ambient temperature. The consequence of this is that the light output of fluorescent lamps is dependent on ambient temperature. The lumen output of fluorescent lamps is generally published for an ambient temperature of 25 °C. T12 and T8 lamps have their maximum light output at 25 °C, while the smaller diameter T5 lamps give their maximum light output at 35 °C (Fig. 7). To obtain a high and near-constant light output over a wide temperature range, the mercury in some fluorescent lamp types is introduced into the tube not as a pure metal but as an amalgam. Such lamps have a lumen output that is near constant over the temperature range of 25–65 °C.

Mains-Voltage Variations

If the mains voltage varies, the power consumed by the lamp changes and with it the vapor pressure and consequently the lumen output. Given an optimized lamp/electromagnetic ballast combination, a 10 % deviation in power will keep the lumen-output decrease usually at less than 10 %. On high-quality high-frequency electronic ballasts, a 10 % deviation in power can keep the lumen-output decrease at less than 3 %. In all cases, with higher mains-voltage decreases, the lumen output decreases rapidly.

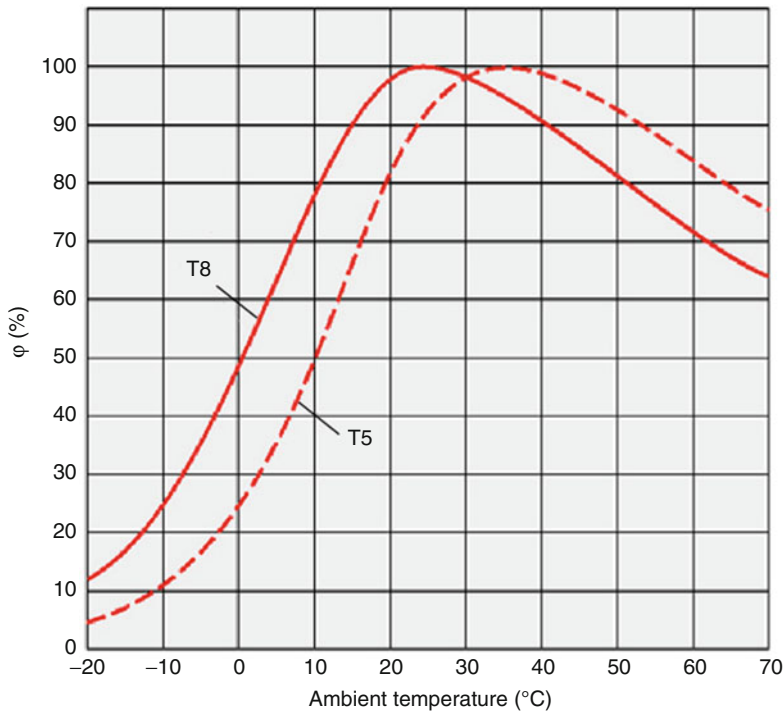
Product Range

Tubular fluorescent lamps are being produced in a wide variety of types with different properties. The more important lamp types are listed in Table 2.

Up until 1990, fluorescent tubes were operated on electromagnetic ballasts. Even today, this way of operating a fluorescent tube is called

Tubular and Compact Fluorescent Lamp,

Fig. 7 Relative light output of a T8 and a T5 fluorescent lamp in relation to the lamp ambient temperature [1]



Tubular and Compact Fluorescent Lamp,
Table 2 Different tubular fluorescent lamp types

Lamp aspect	Range
Tube diameter	38 mm (T12)/26 mm (T8)/16 mm (T5)
Lamp circuit	Electromagnetic/HF electronic
Balance efficacy/lm output	High efficacy (HE)/high output (HO)/very-high output (VO)
Color temperature T_k	Standard range (2,700–6,000 K)/8,000–17,000 K
Color rendering R_a	$R_a > 90$ (900 series)/ $80 < R_a < 90$ (800 series)/ $R_a < 80$
Color	Red, green, blue, etc.
Lifetime	Standard (up to 20,000 h)/very long (up to 65,000 h)
Ignition method	Standard/low temperature ignition
Tube coating	No coating/coating (silicon, protective, or reflector)
Safety of operation	Standard/explosion safe
Tube length	Standard (450–1,800 mm)/mini (150–500 mm)
Tube shape	Tubular/U-shaped/circular

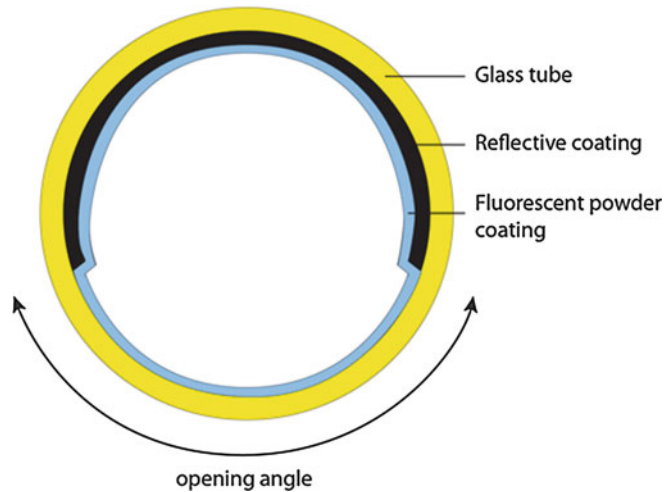
“conventional,” although operation on high-frequency or HF electronic ballasts has since become the standard. T5 lamps can only be operated on HF electronic gear.

T5 lamps are produced in an execution where the design is such that the maximum luminous efficacy is obtained: the high efficiency, HE, series of lamps. Another type of execution is optimized for maximum lumen output at the cost of efficacy: the high output, HO, series (10 % higher lumen output and 10 % lower efficacy). Very-high-output lamps (VHO) are produced as well.

Some tubular fluorescent lamps are given a coating on the outside of the tube. A silicone water-repellent coating is applied to prevent starting problems for types that are used under conditions of high humidity. Types intended for use in food stores can be given a special protective coating that prevents surrounding products from contamination in the event of accidental lamp breakage. Some of these latter types have a spectrum that makes the appearance of food more appealing (especially meat). In reflector

Tubular and Compact Fluorescent Lamp,

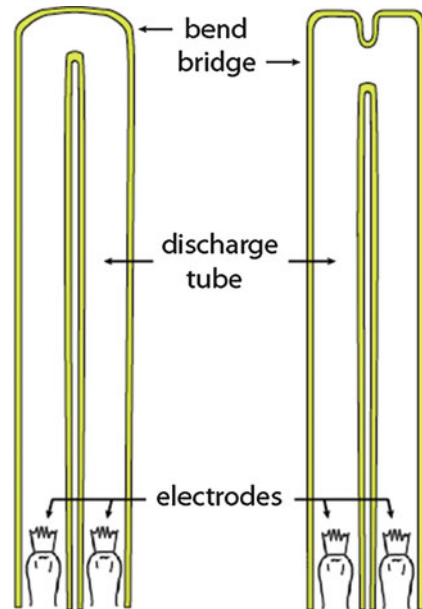
Fig. 8 Reflector fluorescent lamp



fluorescent lamps, a diffuse reflective coating is applied between the upper part of the inner tube wall and the fluorescent powder. The light is concentrated through the uncoated area (or “window”) of the lamp, so increasing the downward component of the light (Fig. 8).

Compact Fluorescent Lamps

Compact fluorescent lamps (CFLs) were originally developed (beginning of the 1980s) for use in those applications where incandescent lamps were traditionally used. Today, the application of compact fluorescent lamps has been widened and includes not only domestic lighting but office and road lighting (residential streets) as well. For the more compact versions, alternatives in the form of solid-state lamps are becoming increasingly more available.



Tubular and Compact Fluorescent Lamp, Fig. 9 Folding one tube (*left*) or connecting two separate tubes (*right*) to form a compact fluorescent lamp with one open pathway between the electrodes

Working Principle

The gas discharge principle employed in compact fluorescent lamps is exactly the same as that in tubular gas discharge lamps (see previous section). Their compactness is achieved by reducing their length. This is done either by folding a longer tube into a shorter one or by joining together two or more parallel tubes so that one open pathway is

obtained where free electrons and ions can move from one electrode to the other, as in a normal straight fluorescent tube (Fig. 9). The folding can be repeated or the interconnection can be done with more than two parallel tubes (always keeping one open pathway) to further increase the size of the lamp.

Tubular and Compact Fluorescent Lamp,

Fig. 10 Nonintegrated two-leg compact fluorescent lamps. From *left* to *right*: two pins with bridge connection, two pins with bend connection, and four pins with bridge connection



Materials and Construction

As far as fill gas, electrodes, and fluorescent powders are concerned, compact fluorescent lamps are essentially the same as conventional tubular fluorescent lamps. Reference is therefore made to the previous section “[Tubular Fluorescent Lamps](#).”

The main difference in construction between compact and tubular fluorescent lamps lies in the tube shape and the lamp caps. But another point of difference is that in those compact lamps that need to work directly from the mains without any external electrical component, the igniter and ballast have to be integrated in the lamp itself. In this case, the lamp foot is used for housing the control gear.

Tube

A great variety of tube shapes are produced. Figure 10 shows versions where tubes are interconnected. All these are examples of compact

fluorescent lamps where the gear is not integrated in the lamp itself (compact fluorescent lamps nonintegrated CFL-NI). Figure 11 shows a set of triple-folded types, while Fig. 12 shows two different folded, single-plane versions.

Figure 13 shows examples of compact lamps where the gear (HF electronic) is integrated in the lamp foot (integrated CFL-I). Since these lamps can directly replace incandescent lamps, it is also desirable that their shape and dimensions should be very close to those of normal incandescent lamps. Figure 14 shows two retrofit versions where an outer bulb with an internal diffusing coating also ensures that the light distribution is close to that of an incandescent lamp.

Lamp Cap

The compact fluorescent lamps with integrated control gear have the same lamp caps as normal incandescent lamps, viz., Edison screw type and bayonet caps.

In most nonintegrated, small, twin-tube versions, the starter is built into the lamp cap itself (integrated starter but nonintegrated ballast). All nonintegrated types are fitted with special caps. A large variety of caps and bases exist in order to ensure that only the correct type of lamp can be used in a given situation (especially defined by the type of gear used in the particular luminaire).

The lamp caps of nonintegrated compact fluorescent lamps are of the push-fit type (they fit by pushing them into the lamp holder). They are

available in both square-shaped and rectangular-shaped versions. Lamp caps for lamps with an integrated starter employ two-pin connectors, while lamps for use with electronic control gear or dimmers have four-pin connectors (Fig. 10 right, Fig. 11 right, Fig. 12 right).

Properties

Energy Balance

Approximately 20 % of the input power of compact fluorescent lamps is emitted in the form of visible radiation. A tubular fluorescent lamp emits some 28 % of visible radiation. The difference is largely due to the fact that in compact lamps the multitude of closely packed tube parts absorb part of the light.

System Luminous Efficacy

The luminous efficacy of a compact fluorescent lamp is partly dependent on the wattage of the lamp and the color quality of the light emitted but more so on how and how many times the compact tube is folded. In low-wattage versions, good color-rendering versions have a system luminous efficacy starting at 45 lm/W, while higher-wattage versions reach efficacies of up to some 70 lm/W.

Lumen-Package Range

Compact fluorescent lamp with integrated control gear is produced in the range from some 250 to 2,000 lm (corresponding wattage range approx. 5–35 W). Nonintegrated types are available in the



Tubular and Compact Fluorescent Lamp,
Fig. 11 Triple-folded nonintegrated compact fluorescent lamps

Tubular and Compact Fluorescent Lamp,

Fig. 12 Single-plane-folded nonintegrated compact fluorescent lamps



Tubular and Compact Fluorescent Lamp,

Fig. 13 Differently shaped integrated compact fluorescent lamps



Tubular and Compact Fluorescent Lamp,
Fig. 14 Retrofit compact fluorescent lamps with shape, dimensions, and light distribution close to that of an incandescent lamp bulb

range from some 250 to 6,000 lm (corresponding wattage range approx. 5–80 W).

Color Characteristics

The color characteristics of compact fluorescent lamps are principally the same as those of normal fluorescent lamps. Reference on this point is therefore made to the previous section “[Tubular Fluorescent Lamps](#).” The range of different color temperatures and color-rendering indices available with compact fluorescents is in practice somewhat limited compared to that found with the very wide range of tubular lamps.

Lamp Life

The lifetime of compact fluorescent lamps is very long compared with that of incandescent lamps, although it is usually shorter than that of tubular fluorescent lamps. Integrated versions (CFL-I) have, depending on type, a rated average lifetime of between 8,000 and more than 15,000 h. Note that the life of retrofit integrated CFLs is usually specified as rated average life (50 % mortality), as is the case with incandescent lamps (which have an average life of 1,000 h). Again, depending on type, nonintegrated versions (CFL-NI) have economic lifetimes (based on 20 % mortality) of 7,000–20,000 h.

Lamp-Lumen Depreciation

The lamp-lumen depreciation of compact fluorescent lamps is similar to that of tubular fluorescent lamps. See previous section “[Tubular Fluorescent Lamps](#).”

Run-Up and Reignition

There are no fundamental differences here compared with normal fluorescent tubes. See previous section “[Tubular Fluorescent Lamps](#).”

Switching

There are no fundamental differences here compared with normal fluorescent tubes. See previous section “[Tubular Fluorescent Lamps](#).”

Dimming

Most integrated compact lamps are not dimmable. However, special, more expensive versions, with

standard Edison or bayonet cap, are produced that can be dimmed to approximately 5 % of full light output. Dimming of nonintegrated, four-pin, compact lamps is possible and is fundamentally the same as with tubular fluorescent lamps.

Ambient-Temperature Sensitivity

With each different lamp shape, the location of the coldest spot is also different, and this has consequences for the mercury pressure and therefore for the ambient-temperature sensitivity. With some shapes, the coldest spot is also influenced by the burning position of the lamp (e.g., base up or base down). The folded, integrated versions often use amalgam instead of pure mercury to stabilize the mercury pressure and thus to minimize the temperature sensitivity.

Mains-Voltage Variations

No fundamental differences compared with tubular fluorescent tubes. See previous section “[Tubular Fluorescent Lamps](#).”

Product Range

Compact fluorescent lamps are being produced in many types with different properties. The more important lamp properties for which different versions are made are: gear integration (integrated or nonintegrated), tube diameter (13 mm and 18 mm), shape, tube length (related to lumen package), lamp circuit, color temperature, and color rendering. Table 3 indicates for all these lamp aspects the range of lamp types available.

Tubular and Compact Fluorescent Lamp, Table 3 Lamp properties versus lamp types

Lamp property	Range of lamp types
Gear integration	Integrated (I)/nonintegrated (NI)
Tube diameter	13 mm/18 mm
Shape	Various
Lamp length	80–200 mm (CFL-I)/ 100–600 mm (CFL-NI)
Lamp circuit	Electromagnetic/HF electronic
Color temperature T_k	Standard range (2,700–6,000 K)
Color rendering R_a	$R_a > 90$ (900 series)/ $80 < R_a < 90$ (900 series)

Cross-References

► [Phosphors and Fluorescent Powders](#)

References

1. Van Bommel, W.J.M., Rouhana, A.: Lighting Hardware: Lamps, Gear, Luminaires, Controls. Course Book. Philips Lighting, Eindhoven (2012)
2. Coaton, J.R., Marsden, A.M.: Lamps and Lighting, 4th edn. Arnold, London (1997)
3. DiLaura, D.L., Houser, K., Mistrick, R., Steffy, G.: IES Handbook. 10th edn. (2011)

Tungsten Filament Lamp

► [Incandescent Lamp](#)

U

UCS Diagram

► [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

Ulbricht Sphere

► [Integrating Sphere](#)

Ultraviolet Radiator

Wout van Bommel
Nuenen, The Netherlands

Synonyms

[Black-light lamps](#); [Wood's lamps](#)

Definition

Radiators that are meant to produce as final outcome ultraviolet radiation as a result of an electrical discharge, generated between two electrodes, in a low- or high-pressure mercury, that is contained in a transparent bulb or tube.

Black-Light Lamps

Many different ultraviolet radiators are produced for industrial applications such as photochemical processes, drying and hardening of materials, lithography, and reprography; for disinfection purposes and insect repellents; and for phototherapy and sun tanning. In these applications, visual effects play no role and are therefore in this encyclopedia not dealt with. Another type of ultraviolet radiators is produced to make fluorescent substances in materials directly visible to the eye. These types are dealt with here. They are usually called black-light lamps and used for artistic and decorative reasons (especially in theaters, discos, and bars) and for making fluorescent dyes visible in bank notes, stamps, and documents to detect counterfeits and for making fluorescent bacterial infections visible for medical purposes.

Since they are not used for general lighting purposes, only the fundamentals of their working principle and of their properties are described in this encyclopedia.

Working Principle and Properties of Black-Light Lamps

Two different types of light sources are used to produce black-light lamps:

- Low-pressure mercury lamps (both in tubular and compact shape)
- High-pressure mercury lamps

The only difference compared with the lamp executions that produce visible light is the type of glass and the type of fluorescent powder used. As discussed in the relevant lamp chapters, these lamps produce ultraviolet radiation as result of the discharge between two electrodes. The glass of the discharge tube exists either of Wood's glass that transmits ultraviolet and infrared radiation but blocks visible radiation (with the exception of purple around 400 nm and deep red around 800 nm) or of normal glass coated with a filter that transmits ultraviolet and blocks visible radiation. The final result is a weak, purple-bluish glowing lamp that emits UV-A radiation with a spectrum depending on the type of fluorescent powder used on the interior of the discharge tube. Either barium disilicate (industrial designation "BSP") with a resulting peak wavelength at 349 nm or strontium tetraborate (industrial designation "SBE") with a peak wavelength at 368 nm is used. The high-pressure mercury type of black-light lamp exists in fluorescent-coated (SBE) and clear versions. The fluorescent version has a higher radiant flux in UV-A, while the clear version has a more concentrated light-emitting area making it more suitable for use in reflector type of luminaires. The high-pressure mercury black-light lamps are far more compact than the low-pressure mercury lamps when comparing a same ultraviolet radiation output. They can therefore be used in floodlight type of luminaires as used for theatrical and concert performances. They are also available in higher output versions.

The low-pressure mercury versions are available in the wattage range from 4 to 36 W with UV-A radiation output power between some 1 and 10 W. The high-pressure versions are available in the wattage range from 100 to 1000 W with UV-A radiation output between some 50 and 500 W.

Cross-References

- [High-Pressure Mercury Lamp](#)
- [Phosphors and Fluorescent Powders](#)
- [Tubular and Compact Fluorescent Lamp](#)

Uniform Chromaticity Scale Diagram

- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)

Unique Hues

Eriko Miyahara Self

California State University, Fullerton, CA, USA

Synonyms

[Focal colors](#)

Definition

Unique hues are unmixed colors without a tint of any other colors. For example, unique yellow is a pure yellow that is not tinged by red or green. There are four unique hues: red, green, blue, and yellow.

Historical View

One can see so many different colors in the environment. And yet there are only four colors that occupy a special place in color perception. They are called unique hues and they originate from the opponent colors theory proposed by Ewald Hering in 1878 [1, 2]. The theory postulates three opponent processes: two chromatic processes of red-green and blue-yellow and one achromatic process of white-black. Unique hues are perceived when one of the two chromatic processes is polarized in one direction and the other is at equilibrium. For instance, one perceives unique red when the red-green process is polarized toward red and the blue-yellow process is at equilibrium. Phenomenologically, one can describe any color he or she sees by a mixture of various ratios of two unique hues. However, one does not perceive two opposing unique hues at the

same and at the same location. Namely, one does not see greenish red or bluish yellow.

The opponent colors theory was at odd with the trichromatic theory that was initially proposed by Thomas Young in 1802 and developed by Hermann von Helmholtz in 1850. The trichromatic theory states that there are three types of receptors that are responsible for conveying color signals. These two theories of color vision were the focus of controversy among scientists until they were integrated into the two-stage model in the 1960s [3]. The trichromatic theory holds at the receptor level in the retina where there are three types of cones. These are called short-, medium-, or long-wavelength-sensitive (S, M, or L) cone receptors, and the names are derived from different spectral region of the peak in their spectral sensitivities. Then the opponent colors theory is at work in the second stage such as retinal ganglion cells and the lateral geniculate nucleus (LGN). Even though this idea of integration of the two major theories appears attractive, recent research has shown that spectrally opponent neurons found in these sites do not really represent red-green and yellow-blue processes of the opponent colors theory. Rather, spectrally opponent neurons in the LGN represent neural signals created by the L-cone inputs antagonized by the M-cone inputs (L-M axis) and those created by the S-cone inputs opposed by a combination of L- and M-cone signals (S-(L+M) axis) [4]. Psychophysical data from humans indicate that unique red closely lies along the L+direction on the L-M axis but that all the other unique hues are clustered along the intermediate directions between the L-M and S-(L+M) axes [5]. In order to address this discrepancy, recent models of color vision usually include a third stage where the further processing of the second-stage mechanisms creates neural signals that correspond to color perception suggested by Hering's opponent colors theory. However, such neural signals in any particular site in the brain have yet to be discovered, leaving unique hues a mystery in color science.

Traditionally, unique hues have been measured with spectral lights that are created by a monochromator to show a light of only one wavelength. More recently, broadband stimuli such as those

printed on paper and those presented on a computer screen have been also used. When the broadband stimuli are used, it is common to calculate their dominant wavelengths to characterize such stimuli. Because the long-wavelength end of the visible spectrum (400–700 nm) appears red with a slight tint of yellow for many observers, there have not been many measurements of unique red with spectral lights. Even though the exact unique hue loci vary between different studies, approximate wavelength ranges for other unique hues are 458–495 nm for unique blue, 490–555 nm for unique green, and 544–594 nm for unique yellow [6].

Individual Differences

Unique hue loci vary greatly among people with normal color vision. For instance, the wavelength for unique green can vary up to 80 nm, which is more than a quarter of the entire visible spectrum [7]. However, the sources of such individual differences remain unknown. Possible physiological sources of individual differences include prereceptor light filtering by the lens and macular pigment, spectral sensitivity and optical density of cones, relative number of different types of cones, and relative sensitivity of the two chromatic mechanisms at the second stage. Among these possibilities, variations in prereceptor filtering and optical density of cones are not large enough to account for observed individual differences in unique hues [8]. Cone spectral sensitivity differences estimated by Rayleigh color match do not correlate with unique green variations [3]. The L/M cone ratio can vary by more than a 30-fold range among individuals, and it is far too large to account for the range of observed unique hue differences [7]. Further, no relationship is found between relative sensitivity of the L-M and S-(L+M) mechanisms and unique hue settings [7].

If all these physiological factors are eliminated as potential sources of individual differences, what else can there be to account for variations in unique hues? Scientists have turned to suggest factors that shape color perception are in the environment, not inside observers. For example, Joel

Pokorny and Vivianne C. Smith in 1977 suggested that unique yellow may correspond to the average illuminant in the observer's environment [9]. This position will make a clear prediction that variations in unique hues among observers would be less when they judge surface colors composed of broadband stimuli than when they judge spectral colors that are monochromatic stimuli [3]. Indeed, meta-analysis of ten different studies from nearly 600 observers confirms this prediction except for unique green [6].

Physiological factors are rejected as sources of individual differences of unique hues among people with normal color vision as described above. However, cone spectral sensitivity seems to play a role when color vision defective observers are considered. Anomalous trichromats have three types of cones like color normal observers, but the spectral sensitivity of one type of cones is shifted. Among them, deuteranomalous observers have M-cones whose spectral sensitivity is shifted toward a longer wavelength, making the difference between M-cone and L-cone spectral sensitivity smaller. This anomaly leads to inefficient red-green color vision and predicts longer wavelength for their unique yellow compared to color normal observers [9]. Multiple studies show that deuteranomalous observers set their unique yellow at substantially longer wavelengths compared to normal observers [7], confirming the prediction.

Cultural Differences

Entwined with individual variations, potential differences in unique hues involve cultural differences. Michael A. Webster and his colleagues measured unique hues from a total of 349 observers in India and the USA in 2002 using color palettes printed on paper [10]. The groups from India included optometry school students, urban workers, and two groups of rural farmers. The observers from the USA were college students. English was used for the students and the urban workers, whereas the separate

native languages were used for the two groups of rural residents. The results showed that there are large differences within groups, again confirming individual differences of unique hues. However, there were also smaller but consistent differences between groups. These differences may reflect many variations in observers' lives such as particular languages spoken, cultural norms and color term usage standards, the environment that causes long-term adaptation, and the kinds of colors encountered frequently on a daily basis. For example, particular shades of color may be more desirable in certain culture, leading to long-term adaptation of the visual system as well as certain usage of color terms. Thus, environmental and cultural factors may be often difficult to separate. Interestingly, unique hue choices made by one of the rural resident groups tended to be more similar to those made by the US students [10]. Overall, these findings seem to indicate that unique hue loci are determined by a combination of various environmental and cultural differences.

Unique Hues and Focal Colors

Separately from the line of research on unique hues in color science, Brent Berlin and Paul Kay published a very influential study on color terms in 1969 in the context of linguistics [11]. In order to test their hypothesis of universality of color terms across languages, they collected color naming data from native speakers of 20 languages. Berlin and Kay presented a palette of Munsell color chips to human observers and asked them to choose a best example of the color term and to draw a boundary of the chips that can be called by the color term. The results showed that the best example chips were clustered in small regions of the palette. Focal colors are the best examples of the color terms as this was the wording used in Berlin and Kay's survey. Though their origins are completely different, unique hues and focal colors for categories red, green, blue, and yellow have been shown to closely correspond with each other empirically [12].

Summary

Unique hues are four primary colors in color perception. Unique hue loci vary dramatically among individuals with normal color vision. However, the sources of such differences are not clearly identified yet. Physiological factors have been excluded as the sources. The emerging picture is that unique hues are shaped by a combination of environmental factors that influence an observer's long-term adaptation and cultural factors that include the language and the customs of color term usage. Focal colors can be considered as a synonym of unique hues red, green, blue, and yellow.

Cross-References

- [Berlin and Kay Theory](#)
- [Environmental Influences on Color Vision](#)
- [Primary Colors](#)
- [Psychologically Pure Colors](#)

References

1. Hering, E.: Outlines of a Theory of the Light Sense (Translated by Hurvich, L.M., Jameson, D.) Harvard University Press, Cambridge (1964)
2. Hurvich, L.M., Jameson, D.: An opponent-process theory of color vision. *Psychol. Rev.* **64**, 384–404 (1957)
3. Mollon, J.D., Jordan, G.: On the nature of unique hues. In: Dickinson, C., Murray, I., Carden, D. (eds.) *John Dalton's Colour Vision Legacy*, pp. 391–403. Taylor and Francis, London (1997)
4. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (1984)
5. Krauskopf, J., Williams, D.R., Heeley, D.W.: Cardinal directions of color space. *Vision Res.* **22**, 1123–1131 (1982)
6. Kuehni, R.G.: Variability in unique hue selections: a surprising phenomenon. *Color. Res. Appl.* **29**, 158–162 (2004)
7. Webster, M.A., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) *Anthropology of Color*, pp. 29–53. John Benjamins, Amsterdam (2007)
8. Pokorny, J., Smith, V.C., Wesner, M.: Variability in cone populations and implications. In: Valberg, A., Lee, B.B. (eds.) *From Pigments to Perception*, pp. 23–34. Plenum, New York (1991)
9. Pokorny, J., Smith, V.C.: Evaluation of single-pigment shift model of anomalous trichromacy. *J. Opt. Soc. Am.* **67**, 1196–1209 (1977)
10. Webster, M.A., Webster, S.M., Bharadwaj, S., Verma, R., Jaikumar, J., Madan, G., Vaithilingham, E.: Variations in normal color vision. III. Unique hues in Indian and United States observers. *J. Opt. Soc. Am. A* **19**, 1951–1962 (2002)
11. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley (1969)
12. Kay, P., McDaniel, C.K.: The linguistic significance of the meanings of basic color terms. *Language* **54**, 610–646 (1978)

Unwanted Light Especially Referring to Unwanted Electric Lighting at Night

- [Light Pollution](#)

USC Diagrams; Uniform Chromaticity Scales; Y_u/v'

Stephen Westland

Colour Science and Technology, University of Leeds, Leeds, UK

Definition

The coordinates of the CIE 1960 UCS diagram are calculated from the 1931 CIE *XYZ* tristimulus value or from the *xy* chromaticity coordinates thus:

$$u = 4X/(X + 15Y + 3Z) = 4x/(-2x + 12y + 3),$$

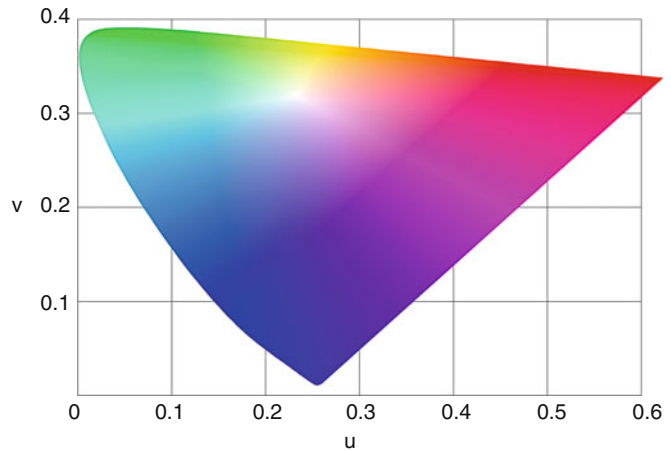
$$v = 6Y/(X + 15Y + 3Z) = 6y/(-2x + 12y + 3).$$

(1)

The 1976 CIE $u'v'$ chromaticity diagram was obtained by stretching the v -axis of the UCS diagram ($u' = u$, $v' = 1.5v$), and the coordinates are calculated thus:

USC Diagrams; Uniform Chromaticity Scales;

Y_u/v' , Fig. 1 CIE 1960 UCS diagram. Figure by Adoniscik (Own Work) [Public Domain] via Wikimedia Commons



$$\begin{aligned} u' &= 4X/(X + 15Y + 3Z) = 4x/(-2x + 12y + 3), \\ v' &= 9Y/(X + 15Y + 3Z) = 9y/(-2x + 12y + 3), \end{aligned} \quad (2)$$

where, alternatively, $U' = 4X/9$, $V' = Y$, and $W' = -X/3 + 2Y/3 + Z/3$ and $u' = U'/(U' + V' + W')$ and $v' = V'/(U' + V' + W')$.

Both the CIE 1960 UCS diagram and the 1976 CIE $u'v'$ chromaticity diagram are associated with the 1931 Y tristimulus value to provide a complete trichromatic specification since $V = V' = Y$.

Historical Development

The 1931 CIE system of colorimetry [1] allowed color stimuli to be defined in terms of tristimulus values XYZ but does not provide a particularly uniform representation of color stimuli in visual terms. If lines are drawn on the CIE 1931 chromaticity diagram that represent equal perceptual steps, then the disparity in the lengths of the lines is as great as 20 times, with the lines in the green region, for example, being much longer than those in the blue region [2].

However, many years earlier it had been suggested, by König, for example, that “it ought not to be difficult in the construction of a chromaticity diagram so to modify the adopted arbitrary assumptions that the separation of two points on it would give a measure for the difference in sensation

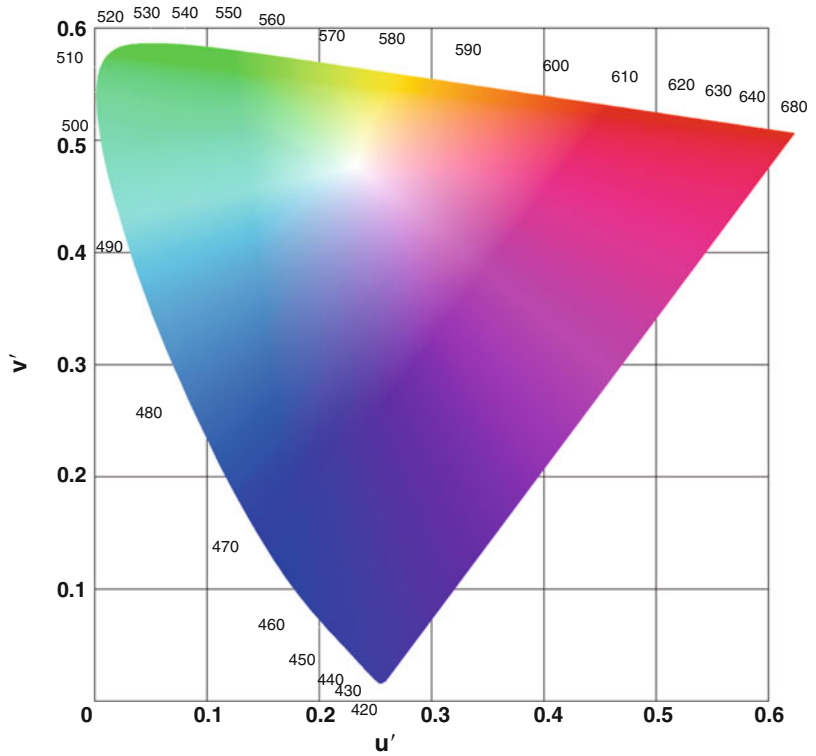
between the colors corresponding to them” [3]. Following work by Judd, in 1935 a Maxwell triangle yielding approximately uniform chromaticity scales (known as the UCS diagram) was defined by reference to the CIE system, and in 1937 MacAdam developed a rectangular coordinate system based on the simple transformation of tristimulus values. In 1959 at a meeting in Brussels, the CIE recommended the 1937 MacAdam uv -diagram for use whenever a projective transformation of the CIE xy -diagram is desired to give uniform chromaticity spacing [4]. This space became known as the CIE 1960 UCS diagram (Fig. 1).

If compared with the 1931 chromaticity diagram, the effect of the linear transformation was to elongate the blue-red portions of the diagram and to relocate the white point to reduce the size of the green area. Although not perfectly uniform, the UCS diagram was thought to be an almost optimal transformation from the 1931 space [2] and was deemed to be sufficient for most practical purposes [4]. However, at a meeting in London in 1975, the CIE proposed modifying the UCS diagram and replaced it with the $u'v'$ -diagram stretching the v -axis ($v' = 1.5v$). The resulting diagram was adopted as the CIE 1976 UCS diagram (Fig. 2).

Properties and Current Status

All chromaticity diagrams, whether xy , uv , or $u'v'$, have the property that additive mixtures of colors

USC Diagrams; Uniform Chromaticity Scales; Y_u/v' , Fig. 2 CIE 1976 UCS diagram. Figure by Adoniscik (Own Work) [Public Domain] via Wikimedia Commons



are represented by points lying on the straight line joining the points representing the constituent colors [5]. However, the CIE 1960 UCS diagram and the 1976 CIE u'/v' diagram represent substantial improvements over the 1931 CIE xy chromaticity diagram in terms of the visual uniformity of the spaces. Whereas lines that represent equal perceptual steps drawn on the CIE 1931 chromaticity diagram differ in length by as much as 20 times, the same lines drawn on the CIE 1976 u'/v' diagram differ in length by about four times (and over much of the diagram, the difference is not greater than two to one) [2].

CIE 1976 UCS is useful for showing the relationships between colors whenever interest lies in their discriminability. However, it is mainly used for the representation of self-emissive colors on display devices or those produced directly from light sources since the diagram assumes that colors are of equal luminance (this condition is rarely met in practice with reflective surface colors). Note, however, that CIELUV, the first approximately uniform three-dimensional space,

is a transformation of the CIE 1976 UCS chromaticity coordinates u' , v' , and Y .

Cross-References

- [CIE Tristimulus Values](#)
- [CIE \$u'\$, \$v'\$ Uniform Chromaticity Scale Diagram and CIELUV Color Space](#)
- [CIELAB](#)

References

1. CIE Pub. No. 15: Colorimetry, Central Bureau of the CIE, Vienna (2004)
2. Hunt, R.W.G.: The Reproduction of Colour, 6th edn. Wiley, Chichester (2004)
3. Judd, D.B., Yonemura, G.T.: CIE 1960 UCS diagram and the muller theory of color vision. J. Res. Natl. Bur. Stand. A. Phys. Chem. **74A**(1), 23–30 (1970)
4. Wright, H.: The Measurement of Colour, 4th edn. Adam Hilger, London (1969)
5. Hunt, R.W.G., Pointer, M.R.: Measuring Colour, 4th edn. Wiley, Hoboken (2011)

$V(\lambda)$, $V'(\lambda)$

► [Spectral Luminous Efficiency](#)

Vantage Theory of Color

Adam Głaz

Institute of English Studies, Maria Curie-Skłodowska University, Lublin, Poland

Synonyms

[VT and color](#)

Definition

Vantage Theory (henceforth VT) is a cognition-based model of color categorization proposed by a one-time student of Brent Berlin and Paul Kay, Robert E. MacLaury [1–6]. The model's major tenet is that humans construct color categories as one, two, or occasionally three vantages (i.e., points of view), a category being an assembly of its vantages.

Overview

Vantage Theory was proposed after MacLaury and his coworkers had conducted interviews

with about 900 speakers of 116 Mesoamerican languages (Mesoamerican Color Survey, part of World Color Survey), later enriched with data from a wide spectrum of world languages. The interviews consisted of three procedures – *naming*, *focus selection*, and *mapping* – and were performed with the use of the Munsell set of color chips, 320 chromatic and 10 achromatic (Fig. 1).

First, in the procedure of naming, the informant was shown the chips one by one in random order and asked to name each. The naming ranges of each color term were then marked on the derandomized array. Next, the informant was asked to choose the focus (best example) of each color term used. Finally, he/she was shown the arranged set without the naming ranges and asked to indicate all chips he/she would refer to with a given term. This process of mapping proceeded in incremental steps, until the informant refused to continue. Thus, the naming range, the focus/foci, and the mapping range of each term were elicited.

Parallelisms between the categorizing behavior of informants and spatiotemporal orientation were observed, the analogy being drawn in an instinctive and neurally expedited manner. While constructing a color category, a person anchors their cognition in a given dimension of color (hue, brightness, or saturation) and relates color stimuli to that fixed coordinate through similarity or its lack. The fixed coordinate is a categorical equivalence of spatial landmarks, whereas attention to similarity or difference arises by analogy to



Vantage Theory of Color, Fig. 1 The Munsell color array. The horizontal axis hue, from red to purple, severed in the middle of red at column 40. The vertical axis is for

brightness. All chips at maximum saturation. On the left: ten achromatic colors, from white to black (© Hale Color Consultants, Inc. Reproduced with permission)

Vantage Theory of Color, Fig. 2 Modeling of the blue category in VT. *Bu* blue focus, *S* attention to similarity, *D* attention to difference

Levels	Fixed Coordinates	Mobile Coordinates	Entailments
1	Bu	S	focus, range
2	S	D	breadth, margin

experiencing relative motion. Constructing a color category is here illustrated with a hypothetical example in Fig. 2.

On level 1, the category is endowed with the blue focus, the starting point for category construction, and a range of color stimuli similar (*S*) to the category focus. When the categorizer starts emphasizing difference (*D*) more than similarity, the category is curtailed at a margin. Through an analogy to motion, VT defines color similarity and difference as reciprocal and gradable, with the endpoints of the cline being total identity and complete disparity. Speakers can shift their attention by moving between the two extremes. Hue (most commonly), brightness (less so), or saturation (rarely) can function as *inherently fixed coordinates*, whereas *S* and *D* are *inherently mobile*: each type can become the other for immediate purposes of constructing a figure-ground arrangement of coordinates (where the fixed coordinate is the ground or “given” and the mobile coordinate is the figure or “new”). This, however, does not deprive either type of its fixed or mobile status, respectively. *S* and *D* are indispensable, whereas

the inherently fixed coordinates depend on the domain of categorization. VT emphasizes hue, brightness, and saturation as important color dimensions based on the implied salience of these perceptual features that is typically found in perceptual experience. If the salience of these dimensions were nonuniform, or if other dimensions took precedent over hue, brightness, and saturation experience, then VT analyses could still be applied to these alternative constructs in a similar fashion.

The ground-to-figure arrangement, in which a mobile coordinate is fixated to serve as ground for the introduction of a new value, is called a *vantage*. Crucially for the categorizing process, there may be two or sometimes three vantages (points of view) on a category, the category being the sum or assembly of the vantages that compose it. In space-time, the same event involving motion is perceived differently depending on one’s location (e.g., on a train vs. while standing by the track). As an example of point of view in color, consider the hypothetical cool category in Fig. 3, named with two terms, *x* and *y*. The two points of view are

DOMINANT VANTAGE				RECESSIVE VANTAGE		
	x			y		
Entailments	Fixed Coordinates	Mobile Coordinates		Fixed Coordinates	Mobile Coordinates	Entailments
focus, range	Bu	S	1	Gn	D	focus, margin
		↗			↗	
breadth	S	Gn	2	D	Bu	curtailment
		↗			↗	
margin	Gn	D	3	Bu	S	range

Vantage Theory of Color, Fig. 3 Modeling of the cool category in VT

the blue-focused similarity-based *dominant* vantage and the green-focused difference-based *recessive* vantage. They are characterized by a reversal of coordinates that swaps an emphasis on similarity to the focus with an emphasis on difference from the focus.

In the blue-focused dominant vantage, the emphasis on *S* is juxtaposed with *Gn* (green), the second focus of the category, which is then juxtaposed with *D*. The recessive vantage starts with the *Gn* focus. However, here the categorizer first establishes the boundary of the vantage by concentrating on difference vis-à-vis a blue category focus, *Bu*, which is introduced as a mobile coordinate on level 2 but returns to its inherently fixed capacity on level 3. There, it is treated as (weakly) similar to *Gn*, the process defining the range of the vantage.

Together, two vantages as points of view are the way VT models the complexity of categories that is seen in the empirical data. According to VT analyses, in such data one can explain the existence of cool or warm categories as the cognitive trading-off of emphases of similarity and difference.

The most important differences between the two vantage types are listed in Table 1.

Figure 4 illustrates these differences with the cool category in Zulu.

The dominant *hlaza* names 62 chips, as opposed 56 for the recessive *kosazana*, but it only spans 20 columns vs. 25 that *kosazana* does. *Hlaza* is focused in G28, which is very near elemental blue in F29, whereas *kosazana*’s first focus (the second focus being disregarded for the present purposes) falls on C17, as many as

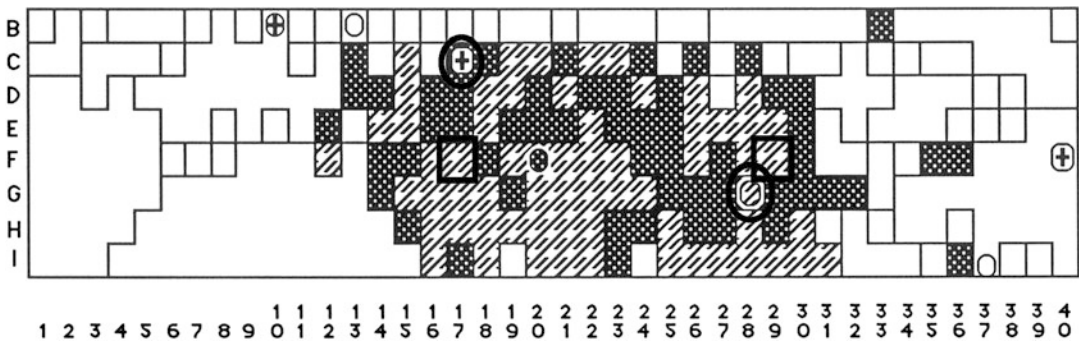
Vantage Theory of Color, Table 1 Characteristics of the dominant and recessive vantages. Elemental colors are “the purest, most intense perceptions” of red, yellow, green, blue, white, and black [1, p. 467], although black and white are traditionally not treated as colors. Elemental colors have their specific locations in the Munsell array, cf. Fig. 4

Dominant vantage	Recessive vantage
Greater number of chips covered	Smaller number of chips covered
Range more concentrated – over a more compact area	Range more dispersed – over a larger area
Focus more centralized relative to elemental colors	Focus less centralized relative to elemental colors

three rows above elemental green in F17 (interestingly, elemental green in F17 is named with the blue-focus *kosazana*). These visible effects (entailments) of the two points of view on the category are like a person’s spatiotemporal location being determined through reference to stationary and moving objects.

Semantic Relations Between Vantages

In addition to modeling color dimension relations as just described, there are also three major types of semantic relations between the dominant and recessive vantages (*near synonymy*, *coextension*, and *inclusion*) that influence the ways color is ultimately categorized. These are “segments of a continuum, not discrete kinds of relation” [1, p. 112].



Vantage Theory of Color, Fig. 4 Naming and focusing of the cool category in Zulu (Bantu group, Niger-Congo family). *Lighter hatching* blue-focused dominant *hlaza*, *darker pattern* green-focused recessive *kosazana*. *Dark*

squares elemental blue (F29) and elemental green (F17). *Dark ovals* foci for *hlaza* (G28) and *kosazana* (C17) (Figure received by the author from Robert E. MacLaury)

In near synonymy, the vantages are very much alike in terms of focus selection and range, and the differences are minimized. An example is the warm category, termed *he* and *lu*, constructed by a speaker of Jicaque (or Tol), an isolate spoken in Honduras [1, p. 123].

Coextension is a unique kind of relation and requires more attention. The fullest account can be found in [1]; see also [2–4]. Coextension was first observed in the warm category of Uspantec (Uspanteco), a Mayan language of Guatemala, and later in many interviews in Mesoamerica and elsewhere. The first four characteristics below are more common than the remaining ones, though all are subject to some degree of variation:

1. One category is named with two different root terms.
2. Each of the two terms is focused in reference to a different elemental hue.
3. The mapping of each term encompasses the focus of the other.
4. There is substantial overlap of the mapping of the two terms.

The more variable features are:

1. Mappings of the two ranges progress in opposite directions.
2. Naming ranges are intermixed, so that chips named with one term may be surrounded by those named with the other.

3. Foci of one or sometimes both terms are polarized, moderately when a term is focused between the category margin and the relevant elemental hue, in extreme cases when the focus falls outside the naming range of the term.

Figure 5 shows coextension in an early phase (closer to near synonymy than to inclusion).

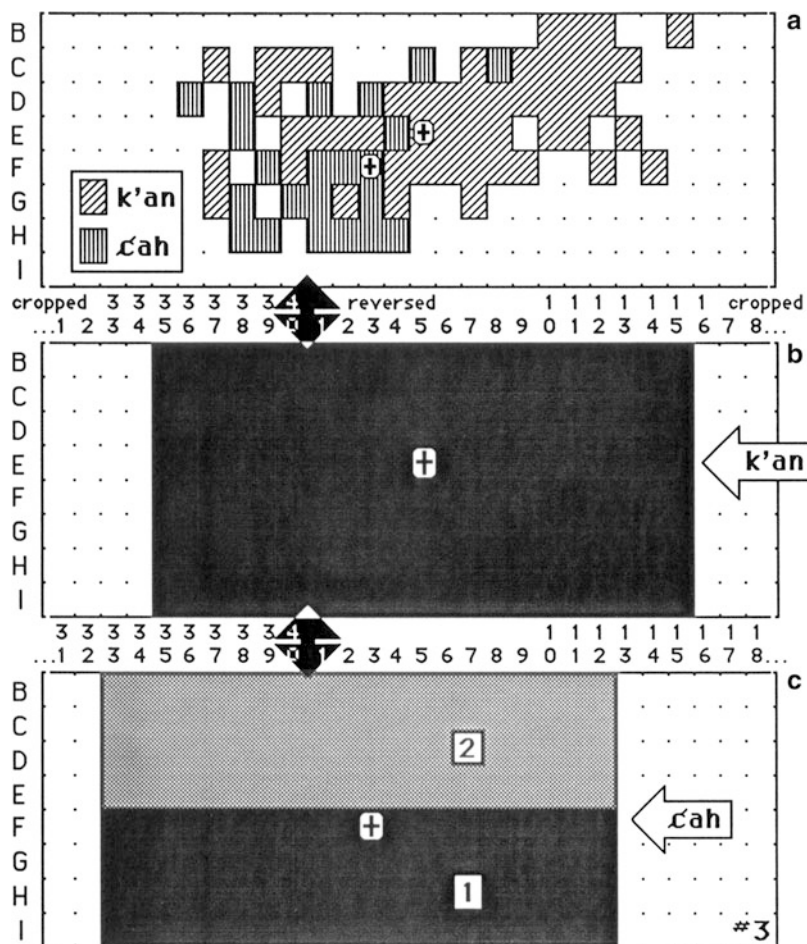
Another example is the Zulu cool category in Fig. 4.

Coextension cannot be explained solely in terms of perceptual dimensions; instead, it is the observer who “assumes opposite slants on the same sensations and names them differently from each angle” [1, p. 113]. Coextension can thus be thought of a variable semantic bias or strategy. This provides a strong argument in favor of subjectivity and speaker agency in categorization and meaning construction.

In inclusion, the naming and/or mapping ranges of the subordinate (recessive) term fall inside that of the superordinate (dominant) term. This happens when, as a result of strong attention to *D*, one of the ranges tends to “drift away” but both still share fixed cognitive coordinates. An example is the warm category in a speaker of Aguacatec (Awakateco), a Mayan language of Guatemala [1, pp. 195–196, e.g., 4].

When the strength of *D* rises even more, the cognitive link between the vantages is broken and the two ranges separate: inclusion becomes *complementation*, a relation between the dominant vantages of distinct categories. Extreme value of *D* causes category split.

Vantage Theory of Color,
Fig. 5 Coextension in an
early phase. Warm in
Tzeltal (Mayan, Tzeltalan),
Paraje Nabil, Tenejapa,
Chiapas, Mexico, male
65, 1980, **a** naming and
foci, **b–c** mappings
(Figure received from
Robert E. MacLaury)



Finer Distinctions

VT models finer aspects of color categorization as *frames*, *stress*, and *viewpoints*.

A frame is a closed system of interdependent parameters, such as emphases on *S* vs. *D*. Occasionally, three vantages on a category are constructed, e.g., in the warm category in Aguacatec [1, pp. 115–116]. In a non-framed analysis, the three vantages are called *dominant*, *recessive*, and *ultra-recessive*, but a framed analysis links the dominant and the recessive terms (the red-focused *k'aq* and the yellow-focused *q'an*) into frame I, while the recessive and the ultra-recessive terms (*q'an* and the brown-focused *sq'inko'x*) constitute frame II. In frame II, the relationship between the recessive and the ultra-recessive term is analogous to that between the

dominant and the recessive term in frame I. Thus, the terms *dominant*, *recessive*, and *ultra-recessive* are relative rather than absolute.

Stress is mental proximity to either the fixed or the mobile coordinates in a vantage. In the color domain, stress may be put on hues or on relations between them. In the former case, one of the hues may “draw” the category to itself – this is called *skewing* [7]. The category may thus divide, as happened in the Mayan language of Kekchí (Guatemala, Belize and El Salvador) [5, p. 57]. In the latter case, focus placement may be random, foci may substitute for each other, and mismatches may be observed between naming ranges, foci and mappings both for a single speaker and between informants. Informants may maintain the dominant-recessive pattern without preference for which hue is dominant or

recessive: the pattern itself seems to be more important than the hues on which it is based. Stress on mobile coordinates has been noted in Mazatec (or Huautla, Oto-Manguen, Mexico) [1, pp. 312–315, 5, pp. 58–59] or Mam (Mayan, Guatemala) [1, pp. 294–306]. The behavior is consistent and appears chaotic only for speakers used to stressing fixed coordinates, as those of English.

Viewpoints are gradable degrees of subjectivity/objectivity with which a person constructs a category or conceptualizes an object or scene [1, pp. 280–283, 4, pp. 528–529, 5, pp. 44–49, 54]. The notion has been found useful in modeling several Mesoamerican languages (Nahuatl, Cakchiquel, Northern Tepehuán, Quiché, Chinantec, or Lacandón). It is a good example of how VT can be extended beyond color: the conception has been applied in several accounts of linguistic behavior [1, p. 284, 8, 9].

Other Notions

Although originating in the color domain, VT is claimed to have universal application and pertain to categorization at large. Aspects of VT not discussed here include, among others, brightness-based categories; full vs. partial inversion of coordinates; submerged vs. reflective vantages; the role of *S* and *D* in categorical evolution; the flip-flop, or oscillation, of color term meanings; dual and triple foci; non-discriminatory vs. analytic vs. synthetic thinking; the spotlight effect; and more [1, 2, 4]. On a more general level, VT relates to the question of linguistic relativity, representing a non-Whorfian stance but one which endorses variety in linguistic behavior. The variety, however, occurs within a cognitively universal but plastic mechanism of categorization [3].

Vantage Theory's Unique Contributions

With VT, MacLaury has contributed to research on color categorization in three major ways. First,

he has redefined the notion of a category, traditionally identified with a color term. In VT a term names a vantage on a category, and there may be one, two, or occasionally three such vantages. Second, the scholar has identified, defined, and modeled coextension, a relation between vantages that at face value appears chaotic or indicative of errors in data elicitation. Third, VT stresses individual variation and personal preference in color categorization. It is an attempt to model the variation and reconcile it with both the classical perceptual constructs of hue, brightness, and saturation and cognitive emphases on similarity and difference. Vantages on categories help to systematically explain how the same stimuli may be differently interpreted by different individuals but also how the same individual might interpret the same stimuli as different on different occasions.

Future Directions

So far, with few exceptions, VT has remained largely dormant though full of potential. In order to utilize this potential, it is proposed that the theory should be made more accessible to the color research community by producing textbook-type publications (a noteworthy precedent is Sect. “9.3” of [10], though it remains isolated and rather limited in scope). Also, because of its author's premature passing, VT is still open to extension and elaboration.

Cross-References

- [Berlin and Kay Theory](#)
- [World Color Survey](#)

References

1. MacLaury, R.E.: *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*, 1st edn. University of Texas Press, Austin (1997/2011)

2. MacLaury, R.E.: Vantage theory. In: Taylor, J.R., MacLaury, R.E. (eds.) *Language and the Cognitive Construal of the World*, pp. 231–276. Mouton de Gruyter, Berlin (1995)
3. MacLaury, R.E.: Linguistic relativity and the plasticity of categorization. In: Pütz, M., Verspoor, M. (eds.) *Explorations in Linguistic Relativity*, pp. 251–293. John Benjamins, Amsterdam (2000)
4. MacLaury, R.E.: Introducing vantage theory. *Lang. Sci.* **24**(5–6), 493–536 (2002)
5. MacLaury, R.E.: Vantage theory in outline. <http://serwisy.umcs.lublin.pl/adam.glaz/vt/VT-Outline.pdf> (1999). Accessed 7 Dec 2011
6. Allan, K.: Categorizing percepts: vantage theory. In: Brown, K. (ed.) *Encyclopedia of Language and Linguistics*, 2nd edn, pp. 252–253. Elsevier, Oxford (2005)
7. MacLaury, R.E.: Skewing and darkening: dynamics of the cool category. In: Clyde, L.H., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 261–282. Cambridge University Press, New York (1997)
8. Glaz, A.: Let me hear you talk and I'll tell you where you are. *Naukovi zapysky Nizhynskogo Derzhavnogo Universytetu imeni Mykoly Gogola. Filologichni nauky* (Research Papers of Nizhyn University. Philology), pp. 90–97 (2007a)
9. Glaz, A.: Vantage theory: a newcomer to the cognitivist scene? In: Fabiszak, M. (ed.) *Language and Meaning*, pp. 91–112. Peter Lang, Frankfurt am Main (2007)
10. Allan, K.: *Natural Language Semantics*. Blackwell, Oxford (2001)

VEP

- [Visual Evoked Potentials](#)

Visual Appearance

- [Appearance](#)

Visual Contamination

- [Color Pollution](#)

Visual Evoked Potentials

Neil Parry

Vision Science Centre, Manchester Royal Eye Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Centre for Ophthalmology and Vision Sciences, Institute of Human Development, University of Manchester, UK

Synonyms

[Event Related Potential](#); [ERP](#); [VEP](#)

Definition

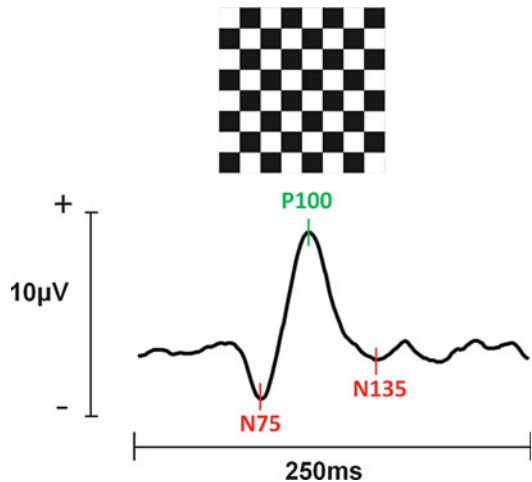
The visual evoked potential (VEP) is a means of extracting from the spontaneous electrical activity in the brain, electrical changes that are directly related to a specific brain action. It is a means of analyzing the Electroencephalogram (EEG). The event-related information (“signal”) is extracted from the overall electrical activity (“noise”) by some form of correlation. Usually, repeated samples of EEG are triggered by a particular event (e.g., an abrupt visual change) and averaged together. Because this event causes a particular type of brain response, at a predictable time, the process of averaging these time locked responses accentuates the event-related signal and smoothes out any uncorrelated activity. By using a stimulus that only contains time-locked changes in color, the VEP can provide an indirect measure of color information processing in the brain.

Electronic supplementary material: The online version of this chapter (doi:[10.1007/978-1-4419-8071-7_107](https://doi.org/10.1007/978-1-4419-8071-7_107)) contains supplementary material, which is available to authorized users.

Scalp-Recorded Electrical Signals from the Brain

The brain is constantly generating large amounts of noisy electrical signals, emanating from the vast number of neurones firing at any given time. This ever-changing activity can be recorded by attaching electrodes to the scalp. The spontaneous activity, known as the Electroencephalogram (EEG) (<http://www.springerreference.com/docs/html/chapterdbid/116485.html>), reflects all the current motor, sensory and cognitive processes. If we wish to study brain activity that is related to a particular function (e.g., hearing or seeing) we can isolate that part of the EEG which occurs in response to specific stimulation by sounds or visual stimuli. In order to elicit such stimulus-related visual response, known as the visual evoked potential (VEP), it is necessary to cause a *change* in the stimulus, since the evoked potential is actually a measure of a *change* in electrical activity.

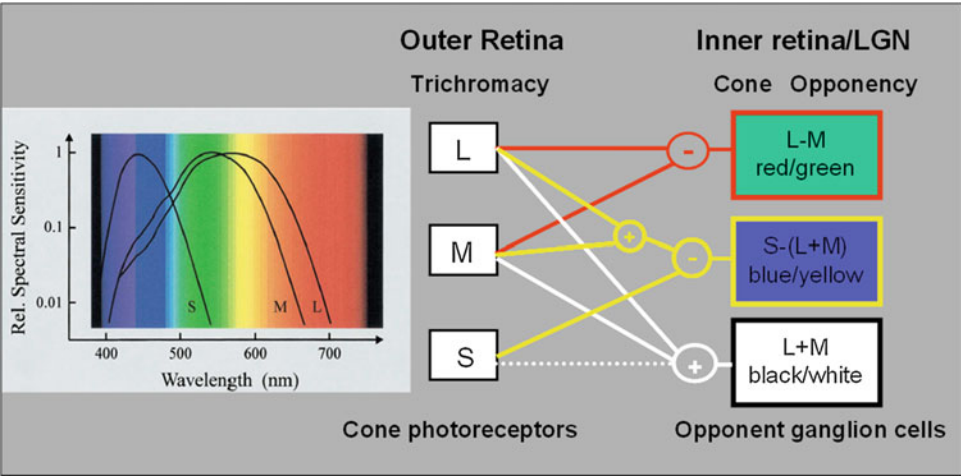
The crudest form of visual stimulus would be an instantaneous flash of light. This is rather like taking a sledgehammer to crack a nut, and is not very well suited to probing specific types of brain function. However it is often used clinically to get some idea of whether there is a gross connection between the eye and the brain, for instance in cases of cortical visual impairment. Patterned stimuli are more frequently employed; typically these are carefully specified repetitive patterns such as checkerboards or gratings. These have to be temporally modulated to elicit a response; usually this takes the form of pattern reversal or pattern onset-offset. Taking the example of a black-white checkerboard, in pattern reversal mode the black and white squares exchange places at fixed intervals, every few hundred milliseconds. In on-off presentation, the checks are replaced periodically by a blank gray screen so that there is no change in space-averaged color or luminance. It is crucial that artifacts are avoided: an overall luminance change between the two phases would itself elicit an evoked response, but this would have no relationship to the spatial content of the stimulus. Figure 1 shows a typical response to a reversing black and white



Visual Evoked Potentials, Fig. 1 VEP elicited to a black and white checkerboard with each square measuring 1 degree of visual angle. The checks reversed every 250 ms. The key wavelets are negativities at about 75 and 135 ms, and a positivity at about 100 ms. The waveform was produced by averaging together fifty 250ms samples of EEG, each of which was triggered by an identical event (the instantaneous reversal of the checkerboard). Timings are relative to this trigger event

checkerboard; this has a widespread use in clinical ophthalmology and neurology.

The electrogenesis of the evoked potential is not well understood, but the signals recorded from an electrode placed on the scalp above the occipital (visual) cortex are said to reflect mainly the activity of early visual areas V1 and possibly V2. The signals are recorded using a differential physiological amplifier, which takes the difference between the potential at the “active” site (e.g., the occiput) and a second electrode placed at a nonvisual or “reference” site (e.g., the vertex or an earlobe). In fact, the terms “reference” and “active” are misnomers as it is possible to record some level of visually related activity from anywhere on the scalp. An alternative name for this montage is unipolar. A bipolar electrode montage places both electrodes over the same cortical area, a few cm apart. This generates smaller signals but has the advantage that distant sources of electrical “noise” tend to be of the same magnitude at both electrodes and are therefore not as readily recorded using a differential amplifier.

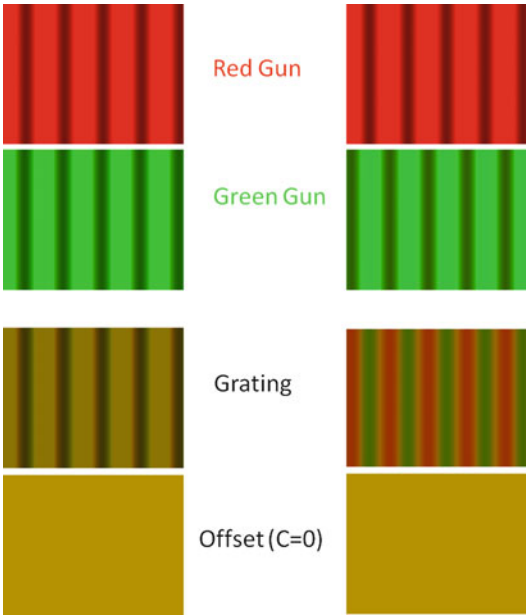


Visual Evoked Potentials, Fig. 2 Spectral characteristics of the cone photoreceptors; subtractive and additive mechanisms giving rise to opponency; with thanks to Prof D. McKeefry

Chromatic Evoked Potentials

A spatial stimulus is defined in terms of many parameters, including its size (how many degrees of visual angle it subtends on the retina), the pattern type (e.g., sinusoidal or square-wave gratings), its periodicity or spatial frequency (i.e., number of spatial cycles per degree of visual angle), the type of temporal modulation (e.g., sinusoidal or square-wave) and temporal frequency (number of temporal cycles per second). In order to elicit a chromatic response, the stimulus has to vary in color alone, in both spatial and temporal domains. It is important to achieve activation of color opponent mechanisms (see Fig. 2; ► [Color Vision, Opponent Theory](#)).

The simplest and most well-established method is to use a sinusoidal grating composed of two colors, say red and green (Fig. 3), which has been adjusted for each subject so that it is isoluminant. This means that, for an individual, there is no change in luminance across the stimulus, but only a change in color or chromaticity. This is usually achieved with a psychophysical procedure such as heterochromatic flicker photometry (HFP), in which the two colors are exchanged at a fairly high rate (about 16Hz), and their luminance reciprocally adjusted until the perception of flicker (which results from residual luminance contrast) is minimised. In on-off mode, the pattern would



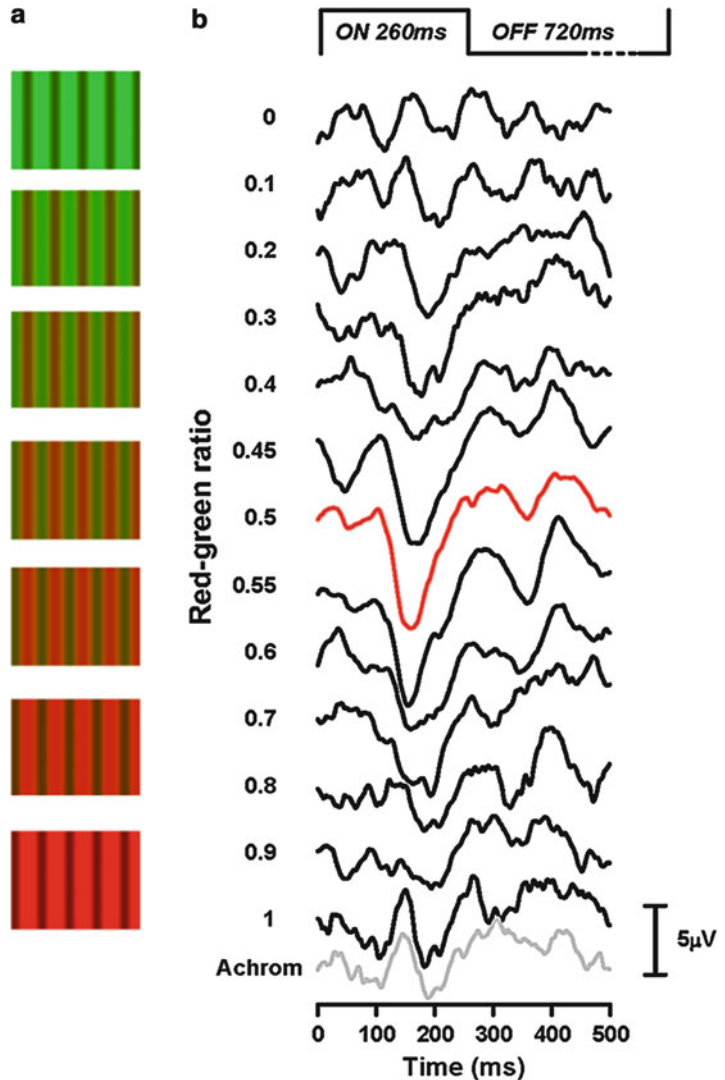
Visual Evoked Potentials, Fig. 3 Combining red and green gratings in phase to make an achromatic (yellow) stimulus and out of phase to make a chromatic (red-green) grating. Also shown is the offset phase when contrast is zero; note that this is the same for both stimuli

be replaced with a plain yellow field of the same space-averaged hue and luminance.

In Fig. 4, a series of VEPs are shown in which luminance contamination is present in varying quantities. At isoluminance (the middle trace,

Visual Evoked Potentials, Fig. 4

VEPs to a range of red-green ratios. The isoluminant response (when, for this subject, $RGR = 0.5$) is shown in red. The gray plot is the equivalent achromatic VEP, recorded to a yellow/black grating. Note that there is no significant difference between this response and those to the red/black or green/black gratings

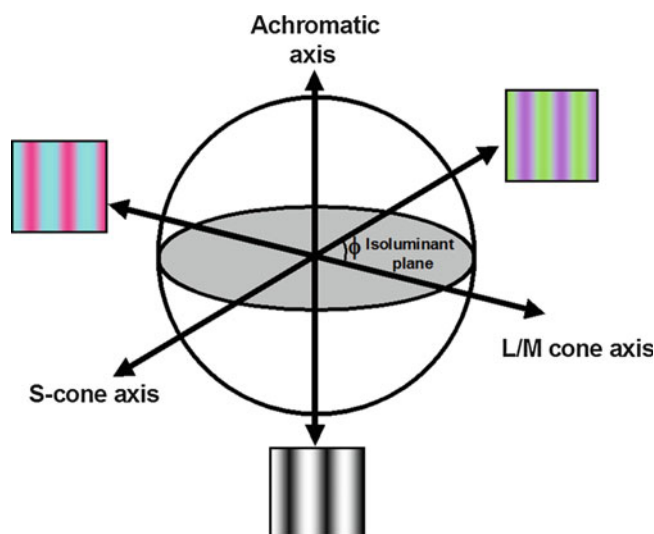


shown in red), the onset response takes on a different shape, the achromatic positivity (seen at top and bottom of the figure) being replaced by a chromatic-specific negativity [1–5]. This achromatic-chromatic difference is not seen when checkerboards or reversal presentation are employed, and this is possibly because these stimuli are not optimal for the chromatic visual system, which, at least for red-green stimuli, is handled by the parvocellular (P) visual pathway [6]. The P pathway is dominated by sustained nerve fibers, whilst reversal favors transient responses, and perhaps therefore activates preferentially the magnocellular system [7].

Although red-green stimuli are relatively easy to generate on a graphics display using the red and green primary phosphors, they usually contain significant contrast in both red-green and blue-yellow color opponent pathways. Since these properties are handled by different retino-cortical pathways (respectively the parvocellular and koniocellular streams), it is desirable to optimize the stimulus so that it activates a particular pathway. Physical colors are often represented in a 3-dimensional space, an example of which is given in Fig. 5. See [► [Psychological Color Space and Color Terms](#)] for more on color spaces. The two component colors of the stimulus are selected

Visual Evoked Potentials,

Fig. 5 MBDKL color space; with thanks to Prof D. McKeefry



from this space, forming a vector. This vector can be located at a particular orientation in order to restrict activation to one channel (say blue-yellow), whilst keeping the others (red-green and luminance) constant. In Fig. 6, a range of VEPs have been recorded to onset of chromatic stimuli which all lie on the isoluminant plane in MBDKL space [8, 9]; thus the luminance channel is held constant. The vector has been rotated for each successive recording so that at some point it passes through each color channel. When the chromatic axis (ϕ) is $0-180^\circ$, the stimulus is modulated along the L/M cone axis (“red-green”); when $\phi = 90-270^\circ$, it preferentially activates the S-cone axis (“yellow-blue”). 10 additional recordings have been made at intermediate angles where one would expect a mixture of red-green and blue-yellow signals. The blue-yellow VEP, which taps the sluggish koniocellular pathway, has a much longer latency (time from onset of the stimulus to peak of the component) than the M or L-cone mediated responses (see Fig. 6 inset).

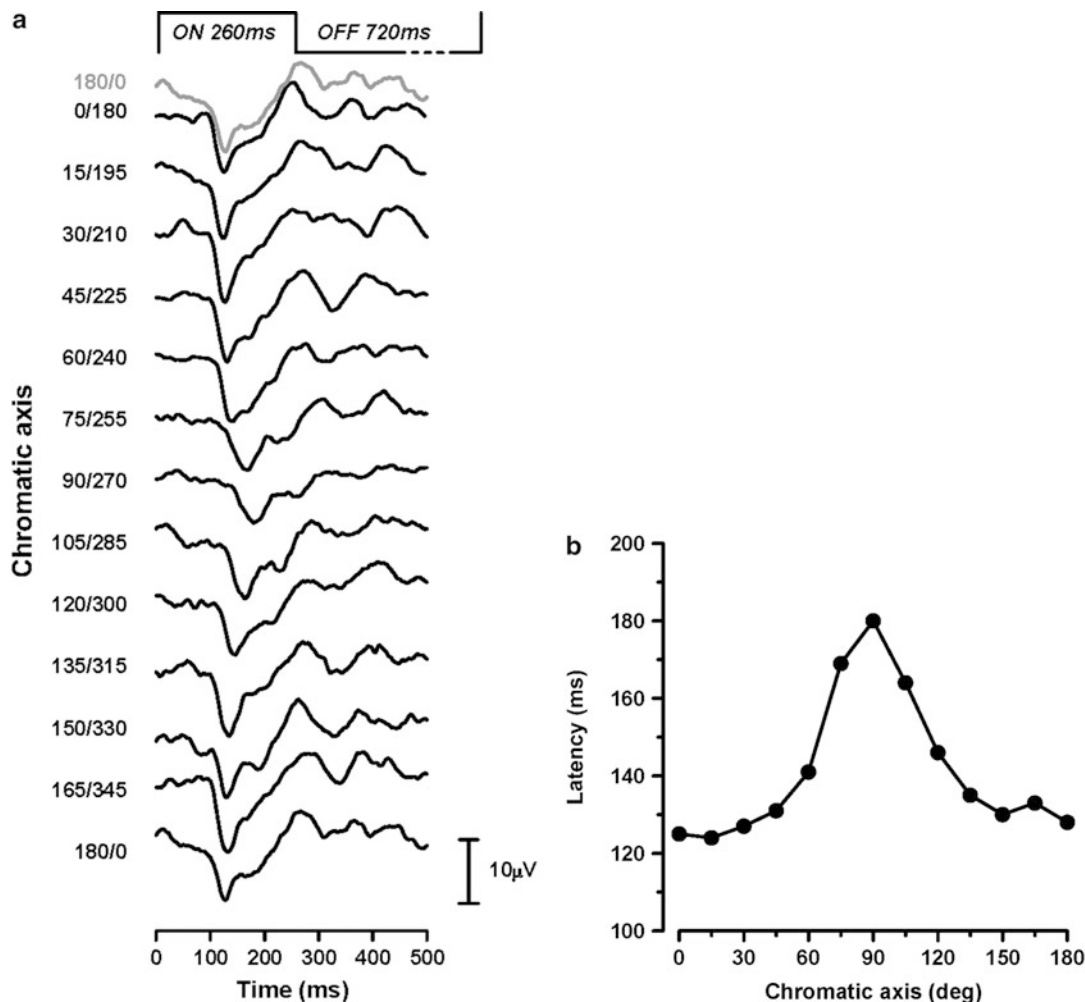
Limitations

There are a few factors which need to be taken into account when recording chromatic evoked potentials. Firstly, these stimuli may be contaminated by luminance information due to chromatic aberrations. The key to avoiding the many complications

arising from chromatic aberrations is to use low spatial frequency ($<5\text{c/deg}$) sinusoidal gratings and to restrict the number of cycles visible. There is another good reason to use low spatial frequencies: the chromatic visual system is tuned to these, showing a low-pass tuning curve with a half-height of about 3 c/deg and an upper limit of about 10 c/deg [10]. The issue of stimulus size becomes particularly important when blue-yellow stimuli are used. This is because the central retina contains a higher concentration of yellow macular pigment, meaning that it is not possible to set a single value of isoluminance across the whole stimulus [11, 12]. Isoluminance also varies from subject to subject, even in those with normal color vision, largely because of the high inter-subject variability in the proportion of long-wavelength to medium-wavelength photoreceptors [[► Cone Fundamentals](#)]. Thus it is always advisable to establish isoluminance for each subject tested. The presence of color vision deficiencies is also important of course. In extremis, dichromatic individuals lack red-green opponency [[► Color Vision, Opponent Theory](#)], and therefore do not generate red-green chromatic evoked potentials (see [Uses](#), below).

Uses

What benefits do chromatic VEPs bring, aside from giving us a tool to help understand



Visual Evoked Potentials, Fig. 6 (Left) Occipital VEPs recorded to a range of chromatic axes, where 0/180 is cardinal red-green and 90/270 is cardinal blue-yellow

(see Fig. 5). (Below) Latency of the chromatic negativity as a function of chromatic axis

the visual system? VEPs are useful in assessing the vision of preverbal individuals, and several studies have used VEPs to study the development of color vision in infants [13, 14]. There are many potential clinical uses, from the diagnosis of color vision deficiencies to the assessment of drug toxicity [15–18]. Studies have applied the technique to measure the effect of environmental toxins, which might be expected to preferentially affect the highly sensitive color visual system [► [Color Perception and](#)

[Environmentally Based Impairment](#)]. The chromatic VEP has proven to be of value in neuro-ophthalmology [19–24], again because diseases such as optic neuritis seem to produce defective color vision. This may not be because the disease specifically targets color vision (for instance via the parvocellular pathway), but because the chromatic system is simply more sensitive to insult or injury, showing their effects more readily than the more robust achromatic visual system.

Cross-References

- [Color Perception and Environmentally Based Impairments](#)
- [Color Vision, Opponent Theory](#)
- [Cone Fundamentals](#)
- [Psychological Color Space and Color Terms](#)

References

1. Carden, D., Kulikowski, J.J., Murray, I.J., Parry, N.R.A.: Human occipital potentials evoked by the onset of equiluminant chromatic gratings. *J. Physiol.* **369**, 44P (1985)
2. Murray, I.J., Parry, N.R.A., Carden, D., Kulikowski, J. J.: Human visual evoked potentials to chromatic and achromatic gratings. *Clin. Vis. Sci.* **1**, 231–244 (1987)
3. Berninger, T.A., Arden, G.B., Hogg, C.R., Frumkes, T.E.: Separable evoked retinal and cortical potentials from each major visual pathway: preliminary results. *Br. J. Ophthalmol.* **73**, 502–511 (1989)
4. Rabin, J., Switkes, E., Crognale, M., Schneck, M.E., Adams, A.J.: Visual-evoked potentials in 3-dimensional color space – correlates of spatiochromatic processing. *Vision Res.* **34**, 2657–2671 (1994). doi:10.1016/0042-6989(94)90222-4
5. Porciatti, V., Sartucci, F.: Normative data for onset VEPs to red-green and blue-yellow chromatic contrast. *Clin. Neurophysiol.* **110**, 772–781 (1999)
6. Lee, B.B.: Visual pathways and psychophysical channels in the primate. *J. Physiol.* **589**, 41–47 (2011). doi:10.1113/jphysiol.2010.192658[pil] jphysiol.2010.192658
7. Kulikowski, J.J., Parry, N.R.A.: Human occipital potentials evoked by achromatic or chromatic checkerboards and gratings. *J. Physiol.* **388**, 45P (1987)
8. Derrington, A.M., Krauskopf, J., Lennie, P.: Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 241–265 (1984)
9. MacLeod, D.I.A., Boynton, R.M.: Chromaticity diagram showing cone excitation by stimuli of equal luminance. *J. Opt. Soc. Am.* **69**, 1183 (1979)
10. Mullen, K.T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J. Physiol.* **359**, 381–400 (1985)
11. Robson, A.G., Parry, N.R.A.: Measurement of macular pigment optical density and distribution using the steady-state visual evoked potential. *Vis. Neurosci.* **25**, 575–583 (2008)
12. Parry, N.R.A., Robson, A.G.: Optimization of large field tritan stimuli using concentric isoluminant annuli. *J. Vis.* **12**(12): 11 (2012). doi:10.1167/12.12.11
13. Parry, N.R.A., Murray, I.J.: Electrophysiological investigation of adult and infant colour vision deficiencies. In: Dickinson, C.M., Murray, I.J., Carden, D (eds.) *John Dalton's Colour Vision Legacy*. Taylor Francis (1997)
14. Crognale, M.A.: Development, maturation, and aging of chromatic visual pathways: VEP results. *J. Vis.* **2** (6): 2 (2002). doi:10.1167/2.6.2
15. Kelly, J.P., Crognale, M.A., Weiss, A.H.: ERGs, cone-isolating VEPs and analytical techniques in children with cone dysfunction syndromes. *Doc. Ophthalmol.* **106**, 289–304 (2003). doi:10.1023/a:1022909328103
16. Tobimatsu, S., Celesia, G.G.: Studies of human visual pathophysiology with visual evoked potentials. *Clin. Neurophysiol.* **117**, 1414–1433 (2006). doi:10.1016/j.clinph.2006.01.004
17. Gomes, B.D., et al.: Normal and dichromatic color discrimination measured with transient visual evoked potential. *Vis. Neurosci.* **23**, 617–627 (2006). doi:10.1017/s0952523806233194
18. Pompe, M.T., Kranjc, B.S., Brecelj, J.: Chromatic VEP in children with congenital colour vision deficiency. *Ophthalmic Physiol. Opt.* **30**, 693–698 (2010). doi:10.1111/j.1475-1313.2010.00739.x
19. Russell, M.H.A., Murray, I.J., Metcalfe, R.A., Kulikowski, J.J.: The visual deficit in multiple sclerosis: a combined psychophysical and electrophysiological investigation. *Brain* **114**, 2419–2435 (1991)
20. Porciatti, V., Sartucci, F.: Retinal and cortical evoked responses to chromatic contrast stimuli – specific losses in both eyes of patients with multiple sclerosis and unilateral optic neuritis. *Brain* **119**, 723–740 (1996). doi:10.1093/brain/119.3.723
21. Tobimatsu, S., Kato, M.: Multimodality visual evoked potentials in evaluating visual dysfunction in optic neuritis. *Neurology* **50**, 715–718 (1998)
22. Horn, F.K., Bergua, A., Junemann, A., Korth, M.: Visual evoked potentials under luminance contrast and color contrast stimulation in glaucoma diagnosis. *J. Glaucoma* **9**, 428–437 (2000)
23. Sartucci, F., Murri, L., Orsini, C., Porciatti, V.: Equiluminant red-green and blue-yellow VEPs in multiple sclerosis. *J. Clin. Neurophysiol.* **18**, 583–591 (2001). doi:10.1097/00004691-200111000-00010
24. Wu, F., Yang, Y., Li, H., Odom, J.V.: Relationship of chromatic visual-evoked potentials and the changes of foveal photoreceptor layer in central serous chorioretinopathy patients. *Ophthalmic Physiol. Opt.* **31**, 381–388 (2011). doi:10.1111/j.1475-1313.2011.00839.x

Visual Illusions

- [Color Spreading, Neon Color Spreading, and Watercolor Illusion](#)

Visual Pigments

- [Photoreceptors, Color Vision](#)

Visual Pollution

► Color Pollution

von Bezold Spreading Effect

► Assimilation

Vora Measure, ν

► Vora Value

Vora Value

Poorvi L. Vora¹ and H. Joel Trussell²

¹Department of Computer Science, The George Washington University, Washington, DC, USA

²Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA

Synonyms

Vora measure, ν

Definition

The *Vora Value*, ν , measures the recording accuracy of a set of color filters [1]. The definition of ν takes into account the viewing and recording illuminants and the target color space. The original definition is independent of the data set to be recorded, does not take into account recording noise, and is based on the Euclidean distance measure in the target color space.

Let $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_s]$ be an $N \times s$ matrix whose columns are the vectors (or functions) that define the target color space for a specified viewing illuminant. If we assume that the vectors of \mathbf{V} are

independent, then they define a basis for the span of \mathbf{V} , $\text{Span}(\{\mathbf{v}_i\}_{i=1}^s)$. To simplify notation, we denote the span of \mathbf{V} by $\mathbf{R}_s(\mathbf{V})$, where the subscript denotes the dimension of the vector space. For example, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ could represent the three *CIE XYZ* matching functions for viewing illuminant D65. Let the column vector \mathbf{f} of length N represent the reflectance spectrum of the sample to be measured. Then the target color space is $\mathbf{R}_s(\mathbf{V})$, and the desired s -stimulus vector is

$$\mathbf{t} = \mathbf{V}^T \mathbf{f}$$

In our example, \mathbf{t} would represent the *CIE XYZ* values of \mathbf{f} for viewing illuminant D-65. The value of s need not be restricted to $s = 3$. For example, one may use $s = 6$ to measure the *CIE XYZ* values of \mathbf{f} for viewing illuminants D-50 and D-65. For another example, $s = 4$ may be used to aid in color correction. For hyperspectral cameras and other multiband image recording systems, s may take on other, larger, values.

Let $\mathbf{M} = [\mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_r]$ be an $N \times r$ matrix whose columns represent the effective recording filters (combination of the scanning filters and scanning illuminant). If recording noise can be ignored, the vector of recorded values is

$$\mathbf{g} = \mathbf{M}^T \mathbf{f}$$

It is typically physically impossible – or, at least, impractical – to construct vectors $\mathbf{m}_i = \mathbf{v}_i$ that match the ideal color matching functions exactly. Further, recording and viewing illuminants generally differ. In general, $r \neq s$, $\mathbf{M} \neq \mathbf{V}$, and $\mathbf{t} \neq \mathbf{g}$. The vector \mathbf{g} is typically color corrected to obtain the best linear estimate of \mathbf{t} , $\mathbf{t}_{est} = \mathbf{A}\mathbf{g}$, where color correction matrix \mathbf{A} is chosen so as to minimize the mean square estimation error, $E\{\|\mathbf{t} - \mathbf{t}_{est}\|^2\}$, where $E\{\cdot\}$ is the expected value operator. If the components of \mathbf{f} are assumed to be independent, identically distributed with variance σ^2 , it can be shown that

$$\mathbf{A} = \mathbf{V}^T \mathbf{M} (\mathbf{M}^T \mathbf{M})^{-1}$$

and

$$E\left\{\|\mathbf{t} - \mathbf{t}_{est}\|^2\right\} = \sigma^2 \text{Trace}\left(\mathbf{V}^T \mathbf{V} - \mathbf{V}^T \mathbf{M} (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{V}\right)$$

The maximum value of $E\left\{\|\mathbf{t} - \mathbf{t}_{est}\|^2\right\}$ is $\sigma^2 \text{Trace}(\mathbf{V}^T \mathbf{V})$.

The Vora Value, v , is a measure of the effectiveness of the set of recording filters $\{\mathbf{m}_i\}_{i=1}^r$, in recording the color of a sample in the target color space, $\mathbf{R}_s(\mathbf{V})$. It measures how accurately \mathbf{t} can be estimated from \mathbf{g} after linear color correction and is defined as

$$v = \left(\sigma^2 \text{Trace}(\mathbf{V}^T \mathbf{V}) - E\left\{\|\mathbf{t} - \mathbf{t}_{est}\|^2\right\} \right) / \sigma^2 \text{Trace}(\mathbf{V}^T \mathbf{V})$$

Note that v increases linearly with a decrease in mean square estimation error and ranges in value from zero to one, where the maximum value of one corresponds to perfect recording.

The definition of v does not take into account recording noise, real data sets, or perceptual error measures, each of which would influence the error expression. For measures that take these into account, please see the section on *Other Measures* below. The original form of v is simple and provides an effective rough estimate of the accuracy of the filter set.

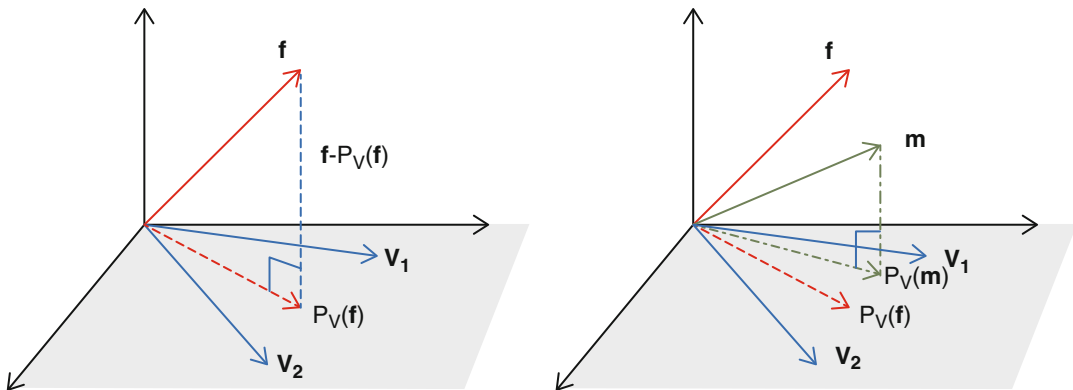
Motivation: The Measure v and the Q-Factor

To understand the properties of a good filter set, note that \mathbf{t} consists of the inner products (dot products) of the vectors $\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_s$ with the vector \mathbf{f} [2]. Thus \mathbf{t} depends only on the *fundamental* of \mathbf{f} , which is defined as $\mathbf{P}_V(\mathbf{f})$, its orthogonal projection onto the target color space. On the other hand, any components of \mathbf{f} that are orthogonal to all of $\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_s$ will not contribute to \mathbf{t} . Thus, in particular, $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$ does not contribute to \mathbf{t} .

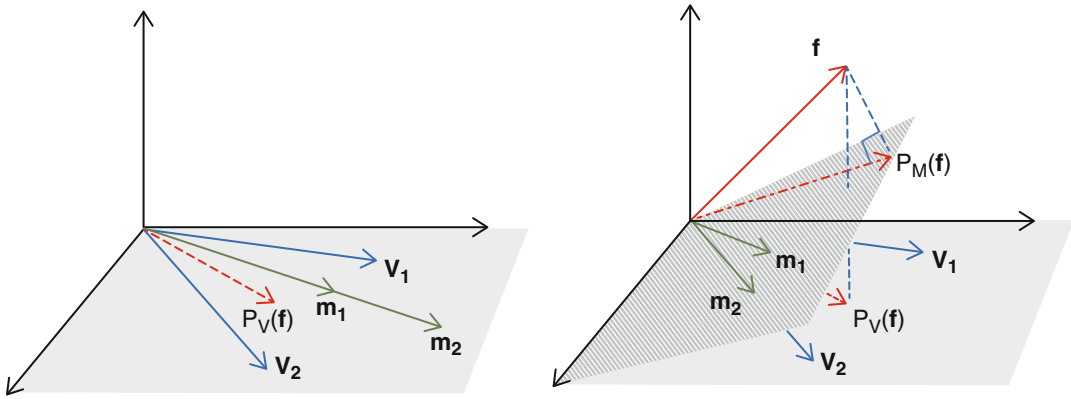
$$\mathbf{t} = \mathbf{V}^T \mathbf{f} = \mathbf{V}^T \mathbf{P}_V(\mathbf{f}) + \mathbf{V}^T (\mathbf{f} - \mathbf{P}_V(\mathbf{f})) = \mathbf{V}^T \mathbf{P}_V(\mathbf{f}) + \mathbf{0}$$

Figure 1a illustrates this for $s = 2$. Vectors \mathbf{v}_1 and \mathbf{v}_2 define the target color space, the horizontal plane. The reflectance spectrum is represented by three-dimensional vector \mathbf{f} . The vector \mathbf{t} is completely determined by $\mathbf{P}_V(\mathbf{f})$, the orthogonal projection of \mathbf{f} onto the horizontal plane. The inner product of $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$ with either of \mathbf{v}_1 or \mathbf{v}_2 is zero and hence does not contribute to \mathbf{t} .

Let \mathbf{m} represent a single effective recording filter, obtained by combining the filter transmission function and the recording illuminant. Ignoring recording noise, the value $\mathbf{m}^T \mathbf{f}$ or the inner product of vectors \mathbf{m} and \mathbf{f} is the recorded value. The component of \mathbf{m} that is orthogonal to the target color space, $\mathbf{m} - \mathbf{P}_V(\mathbf{m})$, contributes to inaccuracy in the



Vora Value, Fig. 1 (a) $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$ does not contribute to \mathbf{t} . (b) $\mathbf{m} - \mathbf{P}_V(\mathbf{m})$ measures parts of $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$



Vora Value, Fig. 2 (a) *Left*: Two or more scanning filters should not record energy in a single direction. (b) *Right*: Differences between $\mathbf{R}_r(\mathbf{M})$ and $\mathbf{R}_s(\mathbf{V})$ lead to inaccurate recording

recording of \mathbf{t} because it measures energy in $\mathbf{f} - \mathbf{P}_V(\mathbf{f})$. This is illustrated in Fig. 1b.

$$\begin{aligned} \mathbf{g} &= \mathbf{m}^T \mathbf{f} = \left(\mathbf{P}_V(\mathbf{m})^T + (\mathbf{m} - \mathbf{P}_V(\mathbf{m}))^T \right) \\ &\quad (\mathbf{P}_V(\mathbf{f}) + (\mathbf{f} - \mathbf{P}_V(\mathbf{f}))) \\ &= \mathbf{P}_V(\mathbf{m})^T \mathbf{P}_V(\mathbf{f}) \\ &\quad + (\mathbf{m} - \mathbf{P}_V(\mathbf{m}))^T (\mathbf{f} - \mathbf{P}_V(\mathbf{f})) \end{aligned}$$

Thus we have seen that only the fundamental of \mathbf{m} , $\mathbf{P}_V(\mathbf{m})$ is useful for recording accuracy. The Colorimetric Quality Factor (CQF or q-factor, defined by Neugebauer in 1956) of \mathbf{m} measures the accuracy of \mathbf{m} from this perspective, and is defined as the energy in the vector $\mathbf{P}_V(\mathbf{m})$, as a fraction of the total energy in \mathbf{m} .

$$q(\mathbf{m}) = \|\mathbf{P}_V(\mathbf{m})\|^2 / \|\mathbf{m}\|^2$$

The value of $q(\mathbf{m})$ is larger when $\mathbf{P}_V(\mathbf{m})$ is a larger fraction of \mathbf{m} . If $\mathbf{m} \in \mathbf{R}_s(\mathbf{V})$, $\mathbf{P}_V(\mathbf{m}) = \mathbf{m}$, and \mathbf{m} has a maximum q-factor of one.

While the CQF captures an important requirement of color scanning filters, recording filters are not used by themselves and it is necessary to evaluate filter sets as opposed to single filters. As in the case of a single filter, it is important for the filter set to measure the energy in $\mathbf{P}_V(\mathbf{f})$. Thus the approach underlying the q-factor is useful. However, while a single filter can record energy only in one “direction” of the target color space (e.g., a single filter can measure one of the CIE

tristimulus values) the purpose of a filter set is to measure energy in all the directions of the target color space. Thus, for example, if we wish to record CIEXYZ tristimulus values, the filter set should accurately measure X, Y, and Z values. An evaluation criterion for a set of recording filters should hence evaluate whether it measures all the energy in $\mathbf{P}_V(\mathbf{f})$ – that is, whether it measures the energy in $\mathbf{P}_V(\mathbf{f})$ in all directions of the target color space.

The generalization of the q-factor to the evaluation of a filter set is not trivial. To see this, consider, for example, a set of three perfect red filters, each with a q-factor of one. While these are clearly inadequate to measure tristimulus values, any function of individual q-factors will indicate that the set is perfect. This limitation is illustrated in Fig. 2a for $s = 2$. The two filters \mathbf{m}_1 and \mathbf{m}_2 are both perfect, lie in $\mathbf{R}_r(\mathbf{M})$ and have q-factors of 1. However, one is a multiple of the other. Hence the filter set is not capable of measuring that part of $\mathbf{P}_V(\mathbf{f})$ that is orthogonal to the filters. It is hence not capable of measuring all the energy in $\mathbf{P}_V(\mathbf{f})$.

The general case is illustrated for $s = r = 2$ in Fig. 2a. The two recording filters \mathbf{m}_1 and \mathbf{m}_2 do not lie in the target color space, which is the horizontal plane, but in the shaded plane. The desired recorded values depend on $\mathbf{P}_V(\mathbf{f})$. The values that are recorded in practice depend on $\mathbf{P}_M(\mathbf{f})$.

$$\begin{aligned} \mathbf{g} &= \mathbf{M}^T \mathbf{f} = \mathbf{M}^T \mathbf{P}_M(\mathbf{f}) + \mathbf{M}^T (\mathbf{f} - \mathbf{P}_M(\mathbf{f})) \\ &= \mathbf{M}^T \mathbf{P}_M(\mathbf{f}) + \mathbf{0} \end{aligned}$$

Hence the set of filters is perfect when $\mathbf{R}_r(\mathbf{M}) = \mathbf{R}_s(\mathbf{V})$, and the error is not dependent on individual filters but on the spaces. Thus $E\{\|\mathbf{t} - \mathbf{t}_{est}\|^2\}$ depends on how much of the energy of $\mathbf{R}_r(\mathbf{M})$ is contained in $\mathbf{R}_s(\mathbf{V})$, or, equivalently, how much of the energy of orthogonal filters spanning $\mathbf{R}_r(\mathbf{M})$ is contained in $\mathbf{R}_r(\mathbf{V})$. It may be shown that

$$v(\mathbf{V}, \mathbf{M}) = \sum_{i=1}^r \frac{q(\mathbf{o}_i)}{s},$$

where $\{\mathbf{o}_i\}_{i=1}^r$ is an orthogonal basis for $\mathbf{R}_r(\mathbf{M})$ and $q(\mathbf{o}_i)$ is the q-factor of \mathbf{o}_i . The use of orthogonal filters prevents correlation among filters from artificially increasing the value of the measure (as in Fig. 2a).

Additional insight into the power of the Vora Value is obtained by considering the case where the target space is the horizontal plane in 3-space. A set of three independent vectors, none of which lie in the plane, form a basis of the 3-space. None of the vectors will have a q-factor of one. It is clear that the span of the three vectors includes the target space. For this case, $v = 1$, as it should. Thus, we can create a filter set that produces perfect tristimulus values from imperfect filters. The cost of this is that it requires more filters than the dimension of the target color space, i.e., $r > s$.

The Measure v and Principal Angles

It may be shown that

$$v(\mathbf{V}, \mathbf{M}) = \sum_{i=1}^s \frac{\cos^2(\theta_i)}{s}$$

where θ_i is the i^{th} principal angle [3] between the target color space, $\mathbf{R}_s(\mathbf{V})$, and $\mathbf{R}_r(\mathbf{M})$.

Other Measures

The measure v is based on a mean-square error in the target color space assuming independent, identically distributed data. It does not take into account real data sets, perceptual error, or recording noise.

Finlayson and Drew [4, 5] address the issue of real data sets. They propose a similar measure based on the *White Point Preserving Vora Error With*

Maximum Ignorance With Positivity (WPPVEMIP), which assumes that one is maximally ignorant of the distribution of \mathbf{f} except for knowing that all values in the vector \mathbf{f} are positive and that the color correction procedure preserves white points. They demonstrate that the *WPPVEMIP* is a more accurate estimate of the real error, and hence that the corresponding measure is more accurate than v .

Wolski et al. [6] address the issue of perceptual error by using a linear approximation of CIE Lab error in the neighborhood of the points in the (known) data set. The filter sets designed by this method minimize an error measure that is a better estimate of perceptual error than the mean square error used to define v .

Sharma and Trussell [7] address the issue of recording noise and propose a *Figure of Merit (FOM)* based on the approach of Wolski et al. The FOM is derived from an error that takes into consideration recording noise.

References

1. Vora, P.L., Joel Trussell, H.: Measure of goodness of a set of color scanning filters. *J. Opt. Soc. Am. A* **10**(7), 1499–1508 (1993)
2. Neugebauer, H.E.J.: Quality factor for filters whose spectral transmittances are different from color mixture curves, and its application to color photography. *J. Opt. Soc. Am.* **46**(10), 821–884 (1956)
3. Golub, Gene H., Van Loan, C.F.: *Matrix Computations*. Johns Hopkins University Press, Baltimore (1996)
4. Finlayson, G.D, Drew, M.S.: The maximum ignorance assumption with positivity. 4th color imaging conference: color, science, systems and applications, IS&T/SID, pp. 202–205 (1996)
5. Finlayson, G.D, Drew, M.S.: White-point preservation enforces positivity. 6th color imaging conference: color, science, systems and applications, pp. 47–52 (1998)
6. Mark, W., Bouman, C.A., Allebach, J.P., Eric, W.: Optimization of sensor response functions for colorimetry of reflective and emissive objects. *IEEE Trans. Image Proc.* **5**(3), 507–517 (1996)
7. Sharma, G., Trussell, H.J.: Figures of merit for color scanners. *IEEE Trans. Image Proc.* **6**(7), 990–1001 (1997)

VT and Color

► [Vantage Theory of Color](#)

Wet or Chemical Finishes

- [Finish, Textile](#)

White Balancing

- [Color Constancy](#)

Wide-Dynamic-Range Imaging

- [High Dynamic Range Imaging](#)

Wood's Lamps

- [Ultraviolet Radiator](#)

World Color Survey

Paul Kay and Richard S. Cook
Department of Linguistics, University of
California, Berkeley, CA, USA

Definition

The World Color Survey collected comprehensive color-naming data from an average of 24 speakers

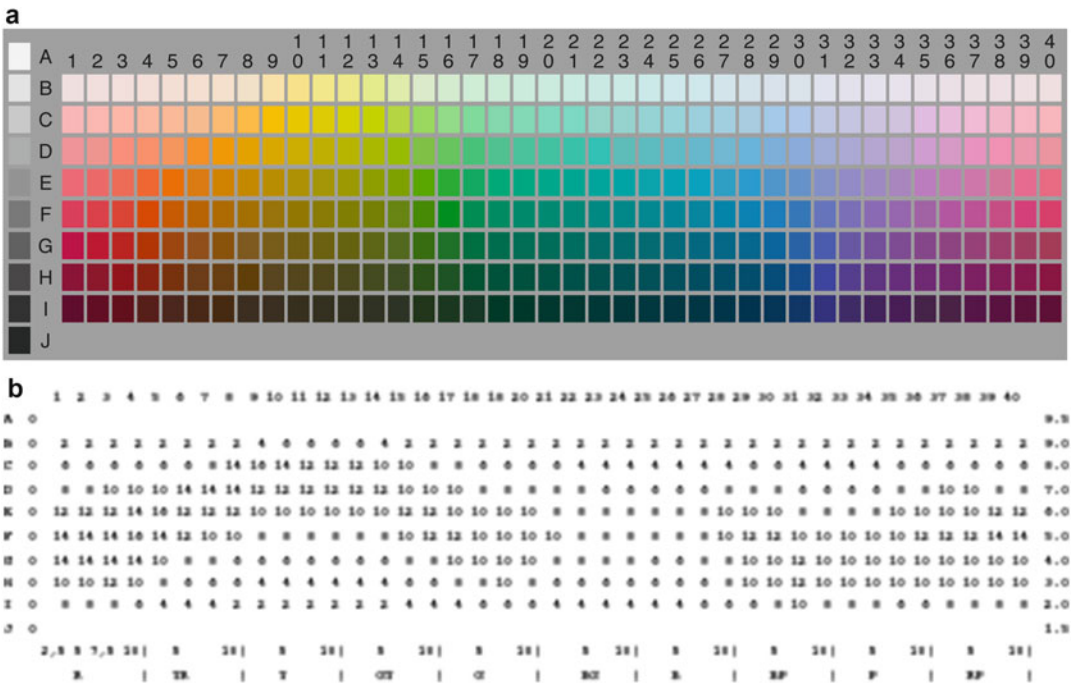
of each of 110 unwritten languages from around the world. Analysis of these data has resulted in a number of research publications. The data are available at <http://www1.icsi.berkeley.edu/wcs/data.html>.

Introduction

The World Color Survey (WCS) was undertaken to investigate the main findings of Berlin and Kay (B&K) [1]. These were (A) that there exist universal crosslinguistic constraints on color naming and (B) that basic color terminology systems tend to develop in a partially fixed order. To this end, the WCS collected color-naming data from speakers of 110 unwritten languages. The WCS data are available in the WCS Data Archive. This entry reviews the history of the WCS, including the creation of the online data archive, and describes some recent uses of the archive to investigate constraints on color naming across languages.

The WCS: History and Methodology

The WCS was begun in 1976 to evaluate the findings of B&K in a full-scale field study. B&K had investigated the color terminology systems of 20 languages in the following way. The stimulus palette used by Erik Lenneberg and John Roberts [2], consisting of 320 Munsell chips of 40 equally



World Color Survey, Fig. 1 (a) The WCS stimulus palette. (b) Munsell and WCS coordinates for stimulus palette of (a). The leftmost column and the top row give the WCS coordinates for lightness and hue, respectively. The rightmost column and the bottom two rows give the

Munsell coordinates for value and hue, respectively. Entries in the body of the table show the corresponding Munsell chroma numbers. (With regard to the A and J rows, there are no Munsell hues at the extremes of value (lightness): 9.5 (white) and 1.5 (black))

spaced hues and eight levels of lightness (value) at maximum saturation (chroma) for each (hue, value) pair, was supplemented by nine Munsell achromatic chips (black through gray to white) – an approximation of the resulting stimulus palette is shown in Fig. 1a and the corresponding Munsell coordinates in Fig. 1b.

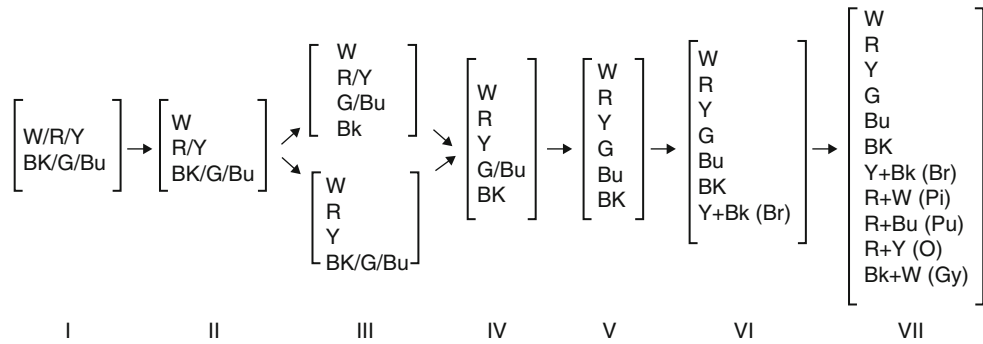
First, without the stimulus palette present, the major color terms of the collaborator’s native language were elicited by questioning that was designed to find the smallest number of simple words with which the speaker could name any color (*basic color terms*). Once this set of terms was established, the collaborator was asked to perform two tasks. In the *naming* task the stimulus palette was placed before the speaker, and for each color term *t*, a piece of clear acetate was placed over the stimulus board, and the collaborator was asked to indicate, with a grease pencil on the acetate sheet, all the chips that he or she could call *t*. In the *focus* task the stimulus palette was

shown as before, and the collaborator was asked to indicate the best example(s) of *t* for each basic color term *t*. B&K concluded that

- [1] The referents for the basic color terms of all languages appear to be drawn from a set of eleven universal perceptual categories, and [2] these categories become encoded in the history of a given language in a partially fixed order. [1]

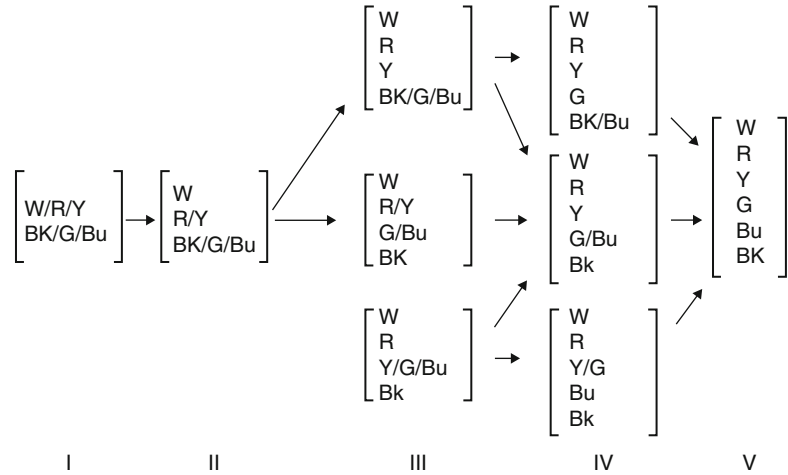
The original universal evolutionary sequence of color term development postulated by B&K is shown in Fig. 2. It has subsequently been revised in detail as more data has become available, but the main outlines of the original sequence have remained intact (see Fig. 3, for the most recent revision).

The B&K results were immediately challenged by anthropologists on the grounds that the sample of experimental languages was too small, too few collaborators per language were questioned (sometimes only one), all native collaborators also spoke English, the data were collected in the



World Color Survey, Fig. 2 Original B&K evolutionary sequence of color term development

World Color Survey, Fig. 3 Revised (2009) evolutionary sequence of color term development



San Francisco Bay area rather than in the homelands of the target languages, many regions of the world and language families were underrepresented or overrepresented in the sample of 20, and that the sample of 20 had too few unwritten languages of low technology cultures [3–6]. The results were nevertheless supported by various ethnographic and experimental studies conducted after 1969 and were from the start largely accepted by psychologists and vision researchers (e.g., [7–9]. See also [10], 498ff, [11], 133 ff).

Work on the WCS was begun in the late 1970s. Through the cooperation of SIL International (then the Summer Institute of Linguistics), which maintains a network of linguist-missionaries around the world, data on the basic color term systems of speakers of 110 unwritten

languages representing 45 different families and several major linguistic stocks were gathered in situ. Fieldworkers were provided with a kit containing the stimulus materials (330 individual chips in glass 35-mm slide sleeves for the naming task and the full stimulus palette for the focus task) as well as coding sheets on which to record collaborators’ responses. The included instructions requested that fieldworkers collect data from at least 25 speakers, both males and females, and urged them to seek out monolingual speakers insofar as possible. The modal number of speakers actually assessed per language was 25, and the average number was 24. (A facsimile of the WCS instructions to fieldworkers and of the original coding sheets is available on the WCS website.) The aim was to obtain names, category extent, and best examples of basic color terms in

each language – basic color terms being described in the instructions as “the smallest set of simple words with which the speaker can name any color.”

The WCS methodology coincided with that of the B&K study in the use of essentially the same set of Munsell color chips. One white chip was added in the WCS study that was whiter than any chip available at the time of the B&K study, making for a total of ten achromatic chips and an overall total of 330 chips, as shown in Fig. 1.

The WCS differed from B&K in the technique for eliciting naming responses. In the WCS procedure, no preliminary interview was administered to establish a set of basic color terms, and in the naming task the 330 individual color stimuli were shown to each cooperating speaker, one by one, according to a fixed, pseudorandom order, and a name elicited for each (in contrast with the B&K procedure of presenting the entire stimulus palette at once in eliciting naming responses). Fieldworkers were instructed to urge observers to respond with short names (although, depending on the morphology of the language, particular field circumstances, and local culture, there was considerable variation in the degree to which the field investigators were able to satisfy these desiderata). Identification of basic color terms, therefore, was done by the fieldworker as a result of the naming task itself, rather than through prior elicitation. The best example (focus) responses were elicited in the same way in both studies: once a set of basic color terms was isolated, the native observer was presented with the full palette and asked to indicate the chip or chips that represented the best example of each term, one by one.

Initial Analysis

Originally, the naming and best example data of the WCS were entered into separate files for each language; they were not compiled into a unified database until the early 2000s. In 2009, a monograph [13] appeared, based on analysis of data in this form. It contains a separate chapter for the color-naming system of each language, identifying the basic terms of the language through a

variety of ways of summarizing and displaying the data. An updated version of the original B&K evolutionary sequence was postulated and is depicted in Fig. 3.

Uses of the WCS Archive

The WCS data archive has been used in investigating two broad questions, one concerning *universals* and other concerning *variation*, of color naming, corresponding to the two major conclusions of B&K [1]. Numerous statistical studies utilizing the WCS online have been conducted.

Universals of Color Naming

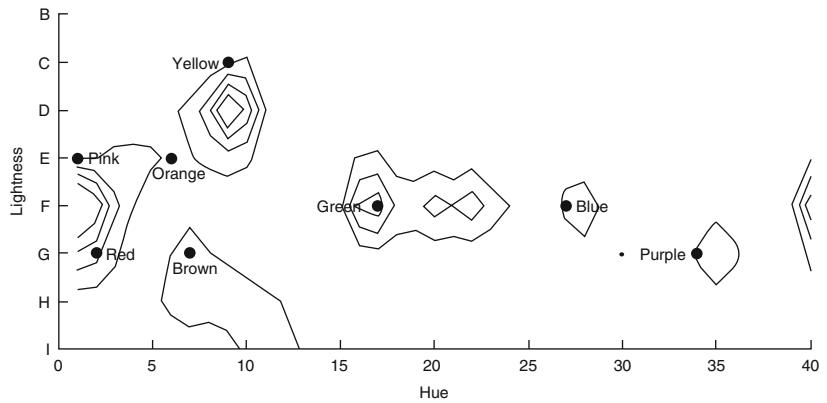
Since B&K found evidence for universals in color naming across languages, the existence of such constraints has generally been accepted in the scientific community. However, there have always been dissenters from this consensus (e.g., [2, 3]), and this dissenting view has recently gained prominence (e.g., [14–19]). Criticisms of the universalist position have come in two major varieties. The first points out that B&K’s findings were never objectively tested, as they relied on visual inspection of color-naming data. Lucy [15] challenges such a methodology as hopelessly subjective:

[Work in the B&K tradition] not only seeks universals, but sets up a procedure which guarantees both their discovery and their form. . . . when a category is identified . . . it is really the investigator who decides which ‘color’ it will count as . . . What appears to be objective - in this case, a statement of statistical odds - is [not]. ([15], p. 334)

On this view, B&K’s subjective methodology allowed them to impose their own universalistic assumptions on their data – so the universals are actually in the minds of the investigators, not in the languages of the world. The second strand of criticism points out that B&K’s data were drawn primarily from written languages and thus may not be representative. This point is coupled with analyses of particular unwritten languages, which are claimed to counterexemplify universal

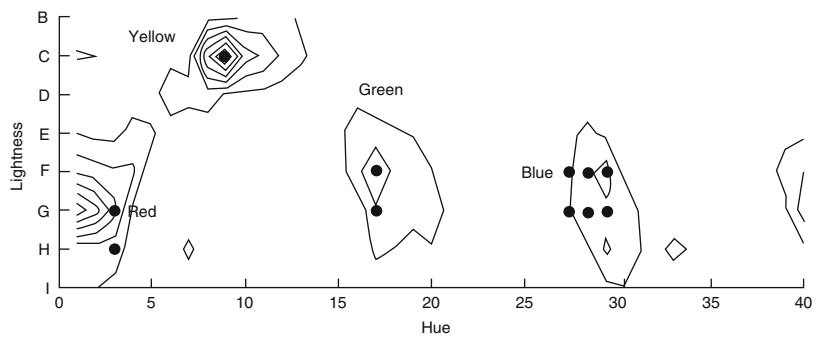
World Color Survey,

Fig. 4 Contour plot of WCS speakers' naming centroids, compared with English naming centroids (black dots); source for English naming centroids: [22]. The outermost contour represents a height of 100 centroids, and each subsequent contour represents an increment in height of 100 centroids (Source of figure: [21])



World Color Survey,

Fig. 5 Contour plot of focal color responses from WCS languages, with superimposed focal responses from English (B&K data, shown as dots), plotted against the stimulus palette (Source: [23])



constraints (e.g., Berinmo: [18, 19]; Hanunóo and Zuni: [15]). Subsequent, more detailed analyses of each of these languages have found that each fits the universal pattern [20]. Disputes of this sort over conflicting interpretations of individual color-naming systems could continue indefinitely. Objective statistical studies were needed to resolve the issue.

The WCS database has been used in a number of independent statistical studies to test the hypothesis that there are statistical constraints on the basic color-naming systems of languages. The weight of the evidence supports the conclusion that such universal statistical constraints exist. The centroids of the naming responses to all color categories documented in the survey were found to cluster in color space more closely than chance would dictate; in a Monte Carlo simulation of the WCS, in which on each of 1,000 trials the modal color-naming pattern of every WCS language was rotated a random hue angle, the actual WCS naming centroids were found to cluster more tightly than the naming centroids in any of

the 1,000 hypothetical (rotated) versions of the WCSs [21]. Figure 4 presents a contour plot of the WCS naming centroids compared to those of English.

The best example choices (foci) of all WCS color terms were found to also cluster more tightly than the centroids of naming categories, suggesting an intimate relation between “focal” colors and universal tendencies of color naming [23]; further, the WCS focal choices were found to cluster closely to those of English or other familiar written languages, as shown in Fig. 5.

Whereas the studies just discussed relied on color naming and focus choice data grouped by language, a clustering and concordance study based on the naming patterns of individual participants independently supports the conclusion that the WCS languages largely partition the color space in ways that, although often having fewer basic terms than English and hence fewer boundaries in their lexical “map” of color space, tend strongly to place boundaries in the same locations as do English and other familiar written

languages [24]. This study also detected a hierarchical order in the lexical partitions of color space compatible with that depicted in Fig. 3. The question naturally arises regarding the degree to which WCS “focus” judgments, that is, participants’ judgments of the most typical examples of named categories, will agree with judgments of unique hues. For example, to what extent do people’s judgments of the best example of green agree with their (actually, in this case, with other people’s) judgments of a green that contains neither blue nor yellow. In a study that pruned the WCS data to consider 38 languages that yield unequivocal results for the Hering fundamental hues, red, yellow, green, and blue, it was found that the focal judgments of the WCS participant speakers of unwritten languages agreed well with unique hue judgments of 300 speakers of several written languages [25]. Systematic, statistical comparison with the WCS database of languages claimed to violate universal color-naming tendencies has refuted that claim; however, a small number of other languages, not the object of such claims, have in fact been found to violate universal tendencies [26]. In a different study, languages with terms roughly equivalent to red, yellow, green, and blue nevertheless were found to differ slightly in their average focus placements, although this variation was greatly exceeded by that found among speakers of the same language [27]. A tentative explanation of the universal tendencies in color naming was found in a study that modeled hypothetical color-naming systems minimizing within-category distance in color space; good fit between the data generated by this model and the WCS data was achieved [28]. One explanation for variation across speakers within a given language is that there appear to be a small number of patterns of naming that occur among some speakers of many languages, with the speakers of few languages all following the same naming pattern [29]. Analysis of the WCS database has revealed that the categories named by basic color terms in the world’s languages tend to be convex sets in color space [30]. An iterated learning study in which a stable color-naming system is achieved by interacting hypothetical agent has shown that equipping such hypothetical agents with the

human just noticeable difference function at the start is sufficient to produce a final output that matches the WCS data well [31]. An iterated learning study using actual human learners showed that limiting a given simulation to just the number of color terms also produced systems of color naming that were statistically close to corresponding systems in the WCS [32]. Extensive narrative descriptions of the color-naming systems of each WCS language, supported by charts and tables and keyed to the overall universal classification scheme, are available in monograph form [13].

Cross-References

- [Berlin and Kay Theory](#)
- [Color Categorical Perception](#)
- [Color Vision, Opponent Theory](#)
- [Comparative Color Categories](#)
- [Dynamics of Color Category Formation and Boundaries](#)
- [Effect of Color Terms on Color Perception](#)
- [Infant Color Categories](#)
- [Multilingual/Bilingual Color Naming/ Categories](#)
- [Unique Hues](#)
- [Vantage Theory of Color](#)

References

1. Berlin, B., Kay, P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley/Los Angeles (1969)
2. Lenneberg, E.H., Roberts, J.M. *The Language of Experience: A study in Methodology*. Memoir 13 of International Journal of American Linguistics. (1956).
3. Durbin, M.: Review (1). *Semiotica* **6**, 257–278 (1972)
4. Hickerson, N.: Review of (1). *Int. J. Am. Linguistics* **37**, 257–270 (1971)
5. Collier, G.A.: Review of (1). *Language* **49**, 245–248 (1973)
6. Conklin, H.C.: Color categorization: Review of (1). *Language* **75**, 931–942 (1973). 1955 Hanunóo color categories. *Southwestern J. Anthropol.* **11**, 339–344
7. Brown, R.W.: Reference. *Cognition* **4**, 125–153 (1976)
8. Miller, G.A., Johnson-Laird, P.: *Language and Perception*. Harvard University Press, Cambridge, MA (1976)

9. Ratliff, F.: On the psychophysiological bases of universal color terms. *Proc. Am. Philos. Soc.* **120**, 311–330 (1976)
10. Kaiser, P. Boynton, R.M.B. *Human Color Vision*. Washington, DC: Optical Society of America (1996)
11. Boynton, R.M.: Insights gained from naming the OSA colors. In: Hardin, C.L., Maffi, L. (eds.) *Color Categories in Thought and Language*, pp. 135–150. Cambridge University Press, Cambridge (1997)
12. Collier, G.A., Dorfinger, G.K., Gulick, T.A., Johnson, D.L., McCorkle, C., Meyer, M.A., Wood, D.D., Yip, L.: Further evidence for universal color categories. *Language* **52**, 884–890 (1976)
13. Kay, P., Berlin, B., Maffi, L., Merrifield, W.R., Cook, R.: *The World Color Survey*. CSLI Publications, Stanford (2009)
14. Lucy, J.A.: *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. Cambridge University Press, Cambridge, UK (1992)
15. Lucy, J.A.: The scope of linguistic relativity: an analysis and review of empirical research. In: Gumperz, J.J., Levinson, S.C. (eds.) *Rethinking Linguistic Relativity*. Cambridge University Press, Cambridge, UK (1996)
16. Lucy, J.A.: The linguistics of color. In: Hardin, C.L., Luisa, M. (eds.) *Color Categories in Thought and Language*. Cambridge University Press, Cambridge, UK (1997)
17. Saunders, B.A.C., van Brakel, J.: Are there nontrivial constraints on colour categorization? *Brain Behav. Sci.* **20**, 167–228 (1997)
18. Davidoff, J., Davies, I.R.L., Roberson, D.: Colour categories in a stone-age tribe. *Nature* **398**, 203–204 (1999)
19. Roberson, D., Davies, I.R.L., Davidoff, J.: Colour categories are not universal: replications and new evidence from a stone age culture. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000)
20. Kay, P.: Methodological issues in cross-language color naming. In: Jourdan, C., Tuite, K. (eds.) *Language, Culture and Society*, pp. 115–134. Cambridge University Press, Cambridge (2006)
21. Kay, P., Regier, T.: Resolving the question of color naming universals. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9085–9089 (2003)
22. Sturges, J., Whitfield, T.W.A.: Locating basic colours in the Munsell space. *Color Res. Appl.* **20**, 364–376 (1995)
23. Regier, T., Kay, P., Cook, R.: Focal colors are universal after all. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8386–8391 (2005)
24. Lindsey, D.T., Brown, A.M.: Universality of color names. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16609–16613 (2006)
25. Kuehni, R.G.: Nature and culture: an analysis of individual focal color choices in World Color Survey languages. *J. Cogn. Cult.* **7**, 151–172 (2007)
26. Kay, P., Regier, T.: Color naming universals: the case of Berinmo. *Cognition* **102**, 289–298 (2007)
27. Webster, M., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R., Paramei, G., Dedrick, D. (eds.) *Anthropology of Color*, pp. 29–53. John Benjamins, Amsterdam (2007)
28. Regier, T., Kay, P., Khetarpal, N.: Color naming and the shape of color space. *Language* **85**, 884–892 (2009)
29. Lindsey, D.T., Brown, A.M.: World Color Survey color naming reveals universal motifs and their within-language diversity. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19785–19790 (2009)
30. Jäger, G.: Natural color categories are convex sets. In: Aloni, M., Bastiaanse, H., de Jager, T., Schulz, K. (eds.) *Logic, Language and Meaning. 17th Amsterdam Colloquium*, pp. 11–20. Springer, Amsterdam (2009)
31. Baronchelli, A., Gong, T., Puglisi, A., Loreto, V.: Modeling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2403–2407 (2010)
32. Xu, J., Griffiths, T. L., Dowman, M. Replicating color term universals through human iterated learning. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Portland, Oregon: Cognitive Science Society (2010)

Wright, David

Robert W. G. Hunt

Department of Colour Science, University of
Leeds, Leeds, UK



W. David Wright was a British physicist and color scientist who made very important contributions to colorimetry and visual science. In fact he is

generally regarded as one of the fathers of colorimetry as it is practiced today.

Wright graduated at Imperial College, London University, in 1926 and received a Ph.D. in 1929 and a D.Sc. in 1937. He was a Medical Research Council student at Imperial College from 1926 to 1929, and in this period, the spectral color-matching properties of ten observers were measured.

He was a Research Engineer at Westinghouse Electric and Manufacturing Co., Pittsburgh, USA, in 1929–30, where he undertook early research on color television.

On his return to England, at Imperial College, he became a Lecturer and Reader in Technical Optics in 1930–1951 and Professor of Applied Optics in 1951–1973. He was a Research and Consultant Physicist to Electrical and Musical Industries in 1930–1939. He became Kern Professor of Communications at the Rochester Institute of Technology in 1984–1985.

Main Interests

W. David Wright had a number of interests which included optics, vision, photometry, colorimetry, color perception, color applications, and color paintings. He supervised a succession of research students in color science, many of whom subsequently held senior positions in academia and industry.

Wright's most important research work was the measurement, for ten observers, of the way in which the colors of the spectrum are matched by beams of red, green, and blue light added together. This work, together with a similar study carried out by John Guild (at the National Physical Laboratory at Teddington, England, with seven additional observers), forms the basis of the international standard for measuring color established by the Commission Internationale de l'Éclairage (CIE). The quality of this experimental work was so high that the standard, although now more than eighty years old, is still in universal use.

Wright's researches also led, for the first time, to definitive descriptions of the main types of color deficiency (color blindness). His book

Researches on Normal and Defective Colour Vision, published in 1946, summarizes the results of his research on color vision. He also wrote the book *The Measurement of Colour*, the four editions of which provided a widely used practical guide to colorimetry from 1944 to the present time. His leisure interests included an appreciation of paintings and religious service and church activities.

Wright's books and key publications include:

The Perception of Light, Blackie, 1938.

Researches on Normal and Defective Colour Vision, Kimpton, 1946.

The Measurement of Colour, Hilger, 1944 (2nd ed., 1958; 3rd ed., 1964; 4th ed., 1969).

Photometry and the Eye, Hatton, 1950.

The Rays are not Coloured, Hilger, 1967.

He also authored about 80 original papers, mainly dealing with color and vision.

Wright received numerous awards and was a member of several professional bodies which included:

The Physical Society Thomas Young Oration, 1951.

Honorary D.Sc. City University, 1971.

The AIC (International Colour Association) Judd Award, 1977.

Honorary D.Sc. University of Waterloo, Canada, 1991.

He also provided public and professional service which included:

Founder of the Physical Society Colour Group, 1941.

Chairman of the Physical Society Colour Group, 1941–1943.

Vice-President of the Physical Society, 1948–1950.

Secretary of the International Commission for Optics, 1953–1966.

Chairman of the Physical Society Optical Group, 1956–1959.

President of the International Colour Association (AIC), 1967–1969.

Chairman of the Colour Group (Great Britain), 1973–1975.

Wyszecki, Gunter

Rolf G. Kuehni
Charlotte, NC, USA



Biography

Gunter Wyszecki was a German-Canadian mathematician/physicist who made important contributions to the fields of colorimetry, color discrimination, color order, and color vision [1]. He was born in Tilsit, East Prussia, Germany (today Sovetsk, Russia), in 1925. He attended the Technische Universität Berlin where he obtained a Dr.-Ing. degree, with a dissertation on normal and anomalous trichromacy [2]. In 1953 he was awarded a Fulbright Scholarship and for a year joined Deane B. Judd at the Colorimetry and Photometry Section of the US National Bureau of Standards in Washington, DC. In 1955 Wyszecki joined the National Research Council Canada in Ottawa where he became the leader of its Optics Section in 1960 and assistant director of the Division of Physics in 1982 and where he remained until his untimely death from leukemia on June 22, 1985.

Major Accomplishments/Contributions

Wyszecki is best known for his scientific contributions to and leadership in the International

Commission on Illumination (CIE). He was chairman of its Colorimetry Committee from 1963 to 1975, vice president of the organization from 1979 to 1983, and its president from 1983 until his death. During this period the CIE made many important recommendations in colorimetry, remaining valid today, such as 1 nm tables of the color-matching functions of the two CIE standard observers and the standard illuminants A and D65, addition of integrating sphere reflectance factor measurement as a recommended measuring geometry, the 1964 ($U^*V^*W^*$) and the 1976 CIELAB and CIELUV uniform color space and color-difference formulas, and others.

Metamerism Wyszecki introduced the important mathematical concept of “metameric blacks,” psychophysical definitions of blacks with tristimulus values 0, 0, 0 that, within limits, can be added to a spectral reflectance to form the various possible metamers having an identical set of tristimulus values under a given light [3]. With W.S. Stiles he also developed mathematical methods to calculate by various methodologies the number of possible metamers for given chromaticities, peaking at the achromatic colors [4].

Wyszecki seven-field colorimeter In 1965 Wyszecki developed the seven-field colorimeter with which an observer can view with both eyes one or more of seven hexagonal fields, each with separately controllable RGB sources, achieved with filtered light and mixed in an integrating sphere. It was successfully used in many different research projects.

Color matching and color-difference matching The MacAdam color-matching error ellipses of 1942 (1 observer) were extended in 1957 to 12 observers by Brown. In 1971 Wyszecki and Fielder determined color-matching error ellipsoids for three observers [5]. The latter two investigations demonstrated the considerable variability by observer. The seven-field colorimeter was also used for a novel color-difference matching experiment in which three fields were displayed and the observer had to adjust the third field so that its brightness matched the brightness of preselected colors with equal luminance in two fields and its chromaticity resulted in identical perceived differences between the colors in the triangular arrangement [6].

Heterochromatic brightness matching

Wyszecki and coworkers added important experimental data to the luminance of equally bright-appearing stimuli. Many chromatic stimuli, when compared to achromatic ones of the same luminance or luminous reflectance, appear to be lighter or brighter, to be “glowing,” an effect known as the Helmholtz-Kohlrausch effect [7].

Publications Wyszecki authored or coauthored 86 scientific papers and 3 books. The first book, *Farbsysteme*, was published in Germany in 1960 [8], describing various color order systems. He coauthored together with D. B. Judd the second and third editions of the latter's *Color in Business, Science, and Industry*, the third edition published after the passing of Judd [9]. He was the lead author, together with W. S. Stiles, of the monumental *Color Science: Concepts and Methods, Quantitative Data and Formulae*, with editions in 1967 and 1982 [10]. The second edition remains in print today as a highly important source of information in the field of color science.

References

1. Robertson, A.R.: Necrology of G. Wyszecki, AIC Newsletter No. 3, (Aug) 18–20 (1986)
2. Wyszecki, G.: Valenzmetrische Untersuchung des Zusammenhanges zwischen normaler und anomaler Trichromasie. *Farbe* **2**, 39–45 (1953)
3. Wyszecki, G.: Evaluation of metamer colors. *J. Opt. Soc. Am.* **48**, 451–454 (1958)
4. Stiles, W.S., Wyszecki, G.: Counting metameric object colors. *J. Opt. Soc. Am.* **52**, 313–319 (1962)
5. Wyszecki, G., Fielder, G.H.: New color-matching ellipses. *J. Opt. Soc. Am.* **61**, 1135–1152 (1971)
6. Wyszecki, G., Fielder, G.H.: Color-difference matches. *J. Opt. Soc. Am.* **61**, 1501–1513 (1971)
7. Wyszecki, G.: Correlate for brightness in terms of CIE chromaticity coordinates and luminous reflectance. *J. Opt. Soc. Am.* **57**, 254–257 (1967)
8. Wyszecki, G.: *Farbsysteme*. Musterschmidt, Göttingen (1960)
9. Judd, D.B., Wyszecki, G.: *Color in Business, Science, and Industry*. Wiley, New York (1975) (2nd ed. 1963, 3d ed. 1975)
10. Wyszecki, G., Stiles, W.S.: *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, New York (1982) (1st ed. 1967, 2nd ed. 1982)

Xenon Lamp

Wout van Bommel
Nuenen, The Netherlands

Definition

Lamps that produce light as a result of an electrical discharge, generated between two electrodes, in a high-pressure pure xenon gas, which is contained in a transparent bulb.

Note: So-called xenon gas discharge lamps for automobile lighting in fact are metal halide gas discharge lamps in which xenon gas does not contribute to the radiation but is added solely to facilitate ignition of the lamp. These lamps do not belong to the category of pure xenon lamps and are therefore not described here.

Xenon lamps are usually not used for general lighting purposes, and therefore, only the fundamentals of their working principle and of their properties are described in this encyclopedia.

Short-Arc Xenon Lamps

The light-emitting surface of short-arc xenon lamps is a near point source of very high luminance with excellent and constant color properties. They are developed as an energy efficient and safe substitute for the awkward carbon arc lamps used until the 1960s in cinema projection.

Today, they are used for cinema projection, film and studio lighting, follow spots in theatre lighting, searchlights and lighthouses. Smaller types are used in microscopy, spectroscopy, and for optical measurements. Because their spectrum resembles that of a black body radiator of 6300 K, and thus of that of sunlight, short-arc xenon lamps, with a luminance approaching that of the sun's surface, are also employed in sun simulators.

Working Principle and Construction

Short-arc xenon lamps exist of a single, double-ended, discharge tube of quartz with two electrodes of solid tungsten. The cathode is cone shaped with a sharp tip and doped with thorium oxide to facilitate electron emission. The anode is bombarded by the electrons and becomes very hot. It is flat shaped, large, and of heavy construction to facilitate heat dissipation and increase life. The high-power version lamps need either forced-air or water cooling. The very short electrode gap between cathode and anode varies, per type, between some 0.5 and 15 mm. The cone-shaped arc discharge stretches from cathode to anode and is consequently very small and of very high luminance. In vertical burning position, the arc exactly lines up with the electrodes. In horizontal burning position, convection forces may lift up the arc, and a magnet (a current through a coil) is in some types required for a stable, lined up arc position. The very high operating pressure of 10–50 atmospheres requires a cold pressure of some 3–15

atmospheres. During the manufacturing process, therefore, xenon is brought into the discharge tube in frozen form.

Short-arc xenon lamps must be operated on DC power because the very high arc currents require a different construction of cathode and of anode as described above. The high cold pressure requires high starting voltage peaks up to 50 kV; some types are therefore equipped with an auxiliary electrode to facilitate ignition.

Properties

System Luminous Efficacy

System luminous efficacies are from 15 to 50 lm/W. The lower values correspond to lower wattage and lower electrode gap versions.

Lumen Package and Luminance Range

A very wide range of lumen packages is available, varying from some 1000 to more than one million lumen (corresponding to some 75–20,000 W). The arc luminance varies, depending of wattage and electrode gap version, between roughly 20,000 and 500,000 cd/cm² (corresponding arc gap: 15 to 0.5 mm). The higher luminance value is the highest of all electric lamp types and is higher than the luminance of the surface of the sun of approximately 150,000 cd/cm².

Color Characteristics

Contrary to most gas discharge lamps, the spectrum of a short-arc xenon lamp approaches that of a near continuous spectrum and is practically independent of lamp type, wattage, dimming state and operating hours. In the visible range, it resembles a black body radiator of 6300 K. The color rendering index is better than 90.

Outside the visible area the lamps also produce ultraviolet radiation in the region down to 150 nm. The result is that ozone is produced during operation (ultraviolet radiation reacts with oxygen in the air). To avoid this effect some versions employ doped or coated quartz material that absorbs ultraviolet radiation with wavelengths smaller than 240 nm.

Lamp Life

Average lamp life varies with type from 500 to 2500 h.

Run-up and Reignition

Full light output is obtained almost instantaneously. Hot reignition is also immediate.

Handling

The cold pressure of 2–15 atmospheres requires precautions in lamp handling in order to prevent explosion. Protective coverings are supplied with the lamp for safe transportation and handling.

Long-Arc Xenon Lamps

Long-arc xenon lamps were introduced in the late 1950s as an alternative for incandescent lamps in stadium floodlighting. For lighting purposes these large lamps have since the 1960s been completely superseded by metal halide lamps.

Cross-References

- [Daylighting](#)
- [Metal Halide Lamp](#)

X-Linked Inherited Color Vision Deficiency

- [Deuteranopia](#)
- [Protanopia](#)

XYZ

- [CIE Tristimulus Values](#)

List of Entries

A

Achromatic Color
Achromatopsia
Acquired Color Vision Impairment
Acquired Color Vision Loss
Adaptation
Aftereffect
Afterimage
Al-Biruni
Al-Farisi, Kamal al-Din Hasan ibn Ali ibn Hasan
Al-Haytham (Alhazen), Abū ‘Alī al-Ḥasan ibn al-Ḥasan ibn
Al-Tusi, Nasir al-Din
Anchoring Theory of Lightness
Ancient Color Categories
Ancient Color Terminology
Apparent Magnitude – Visual Magnitude
Apparent Magnitude, Astronomy
Appearance
Arc Lamps
Architectural Lighting
Art and Fashion Color Design
Artistic Rendering
Assimilation
Aura
Automotive Lighting
Autosomal Dominant Inherited Color Vision Deficiency
Avicenna, Abū ‘Alī al-Ḥusayn ibn ‘Abd Allāh ibn Sīnā

B

Balance
Bartleson, C. James
Basic Colors
Bayesian Approaches to Color Category Learning

Behnam’s Disk
Berlin and Kay Theory
Bezold–Brücke Effect
Billmeyer, Fred Wallace, Jr.
Binocular Color Matching
Binocular Color Perception
Binocular Color Vision
Blackbody and Blackbody Radiation
Black-Light Lamps
Blending
Blind Spot
“Blue Blindness”
“Blue-Seven Phenomenon”
Blue-Yellow Deficiency

C

Candela Distribution
Canonical Color
Car Lighting
Carbon Arc Lamp
CAT
Categorical Perception
Chevreul, Michel-Eugène
Chiasma Opticum
Chromatic Adaptation
Chromatic Contrast
Chromatic Contrast Sensitivity
Chromatic Image Statistics
Chromatic Processing
Chromostereopsis
Chromotherapy
CIE 1931 and 1964 Standard Colorimetric Observers: History, Data, and Recent Assessments
CIE 1976 $L^*a^*b^*$
CIE 1976 $L^*u^*v^*$ Color Space

- CIE 1994 (ΔL^* ΔC^*_{ab} ΔH^*_{ab})
 CIE 2000 Color-Difference Equation
 CIE Chromatic Adaptation; Comparison of von Kries, CIELAB, CMCCAT97 and CAT02
 CIE Chromaticity Coordinates (xyY)
 CIE Chromaticity Diagrams, CIE Purity, CIE Dominant Wavelength
 CIE Color Appearance Model 2002
 CIE Color-Matching Functions
 CIE Color-Rendering Index
 CIE Cone Fundamentals
 CIE Fundamental Color Matching Functions
 CIE Guidelines for Evaluation of Gamut Mapping Algorithms: Summary and Related Work (Pub. 156)
 CIE Guidelines for Mixed Mode Illumination: Summary and Related Work
 CIE $L^*a^*b^*$
 CIE Method of Assessing Daylight Simulators
 CIE Physiologically Based Color Matching Functions and Chromaticity Diagrams
 CIE Special Metamerism Index: Change in Observer
 CIE Standard Deviate Observer
 CIE Standard Illuminant A
 CIE Standard Illuminant D65
 CIE Standard Illuminants and Sources
 CIE Standard Source A
 CIE Tristimulus Values
 CIE u' , v' Uniform Chromaticity Scale Diagram and CIELUV Color Space
 CIE Whiteness
 CIE94, History, Use, and Performance
 CIECAM02
 CIECAM02 (Standards: CIE)
 CIEDE2000, History, Use, and Performance
 CIELAB
 CIELAB (Standards: CIE)
 CIELAB for Color Image Encoding (CIELAB, 8-Bit; Domain and Range, Uses)
 Circadian Rhythms
 Circadian System
 Clapper-Yule Model
 Clinical Color Vision Tests
 Coarseness
 Coatings
 Coefficient of Utilization, Lumen Method
 Coherence
 Color Accord
 Color Adaptation
 Color Aesthetics
 Color and Culture
 Color and Light Effects
 Color and Light Interdependence
 Color and Lightness Constancy
 Color and Visual Search, Color Singletons
 Color Appearance
 Color Appearance Solid
 Color Atlases
 Color Bleeding
 Color Blindness
 Color Board
 Color Categorical Perception
 Color Categorization and Naming in Inherited Color Vision Deficiencies
 Color Category Learning in Naming-Game Simulations
 Color Centers
 Color Circle
 Color Combination
 Color Constancy
 Color Contrast
 Color Coordination
 Color Dictionaries and Corpora
 Color Directions
 Color Dynamics
 Color Effects to Humans and the Environment
 Color Environment
 Color Experience
 Color Field Painting
 Color from Motion
 Color Harmonies for Fashion Design
 Color Harmony
 Color Image Statistics
 Color Mixture
 Color Models
 Color Order Systems
 Color Palette
 Color Perception and Environmentally Based Impairments
 Color Phenomenology
 Color Pollution
 Color Pop-out
 Color Preference

Color Processing, Cortical
Color Psychology
Color Qualia
Color Range
Color Roles
Color Scene Statistics, Chromatic Scene
 Statistics
Color Scheme
Color Selection
Color Signals
Color Similarity Space
Color Solids
Color Spaces
Color Spreading
Color Spreading, Neon Color Spreading, and
 Watercolor Illusion
Color Statistics
Color Stereoscopic Effect
Color Stereoscopy
Color Synesthesia
Color Syntax
Color Trends
Color Union
Color Uses
Color Vision Abnormality
Color Vision Screening
Color Vision Test
Color Vision Testing
Color Vision, Opponent Theory
Color Wheel
Colorant
Colorant, Environmental Aspects
Colorant, Halochromic
Colorant, Natural
Colorant, Nonlinear Optical
Colorant, Photochromic
Colorant, Textile
Colorant, Thermochromatic
Colorant, Thermochromic
Coloration, Fastness
Coloration, Mordant Dyes
Coloration, Textile
Coloring
Color-Magnitude Diagrams
Color-Opponent Processing
Color-Separation Overlay
Color-Word Interference

Colouring
Colour-Opponent Processing
Combustion Lamp
Comparative (Cross-Cultural) Color Preference
 and Its Structure
Comparative Color Categories
Complementary Colors
Compositing and Chroma Keying
Computer Depiction
Computer Modeling
Cone Fundamentals
Cone Fundamentals, Stockman-Sharpe
Cone Pigments
Cone Spectral Sensitivities
Congenital Color Vision Abnormality
Contextual Color Design
Contrary Colors
Contrasting Colors
Correlated Color Temperatures
Cortical
Crawford, Brian Hewson
Cross-Cultural Color Preference
Culture and Color

D

Dalton, John
Daltonism
Daylight
Daylight Illuminants
Daylight Spectra
Daylighting
d-Block Elements
Deep Compositing
 ΔE^*_{94}
 DE^*_{ab}
 ΔE_{00}
Demichel Equations
Depth in Color
Deutanopia
Diagnosis of Defective Color Vision
Diagnostic Color
Dichromacy
Dichromat Color Naming
Diffuse Lighting
Dry or Mechanical Finishing
Dye
Dye Functional

Dye, Functional
 Dye, Ink-Jet
 Dye, Liquid Crystals
 Dye, Metal Complex
 Dye, Photodynamic Therapy
 Dyeing
 Dyes
 Dynamic Color Design
 Dynamic Color Spreading
 Dynamics of Color Category Formation and
 Boundaries
 Dyschromatopsia

E

Early Color Lexicons
 Early Expression of Color in Language
 Eco-/Genotoxicity of Synthetic Dyes
 Effect of Color Terms on Color Perception
 Efficient Coding
 Elastic Light Scattering
 Elastic Scattering
 Electrical Control Gear for Lamps
 Electrodeless Low-Pressure Mercury Lamps
 Electromagnetic Radiation
 Elementary Colors
 Emotional Color Effects
 Environmental Chemistry of Synthetic Dyes
 Environmental Color Design
 Environmental Influences on Color Vision
 ERP
 Event Related Potential
 Evolution
 Expressive Rendering
 Eye Sensitivity Functions

F

Favorite Colors
 Fechner, Gustav Theodor
 Fechner's Colors and Behnam's Top
 Fechner-Behnam Subjective Color
 Filling-In
 Finish, Textile
 Fitting (Anglo-Saxon English)
 Fixture (American English)
 Flame Light Lamps
 Focal Colors
 Forecast

Four-Dimensional Color Vision
 Functionality of Color
 Fundamental Colors

G

Galaxy Color Magnitude Diagram
 Gamut Volume
 Ganglion cells
 Ganzfeld
 Geniculate Pathways
 Gestalt Grouping Principles (or Laws)
 Ghost Image
 Glare
 Glints
 Global Illumination
 Gloss Meter
 Glossmeter
 Glow Discharge Lamps
 Goethe, Johann Wolfgang von
 Goniochromatism
 Goniometer
 Goniophotometer
 Graininess
 Grassmann, Hermann Günther
 Green-Blindness
 Guild, John

H

Halftone Layer Superposition Equations
 Halogen Lamp
 Hard-Edge Painting
 HDR
 HDRI
 Headlamps
 Helmholtz, Hermann Ludwig von
 Hering, Karl Ewald Konstantin
 Hertzprung-Russell Diagram
 High- and Low-Pressure Sodium Lamp
 High Dynamic Range Imaging
 High-Pressure Mercury Gas-Discharge Lamp
 High-Pressure Mercury Lamp
 High-Pressure Mercury-Vapor Lamp
 Highway Lighting
 History, Use and Performance
 Hue Difference, Delta H
 Hue Opponency
 Hue Sequence

I

Ibn Rushd (Averroes)
 Ibn Sahl, Abu Sa'd al-'Ala'
 ICC L*a*b*
 ICCLAB
 Ideal Observer Models of Color Category Learning
 Illuminance Meter
 Illusory Colors
 Image Burn-In
 Image Fidelity
 Image Quality
 Image Similarity
 Impressionism
 Inappropriate Use of color
 Incandescence
 Incandescent Bulb
 Incandescent Lamp
 Indirect Illumination
 Individual Differences
 Indoor Lighting
 Induction Lamp
 Infant Color Categories
 Influences Between Light and Color
 Instrument: Colorimeter
 Instrument: Photometer
 Instrument: Spectrophotometer
 Integrating Sphere
 Interactions of Color and Light
 Interior Lighting
 Interpolation of Spectral Data
 Interreflection
 Intrinsically Photosensitive Retinal Ganglion Cells
 Iodine Lamps
 ipRGCs
 Iridescence (Goniochromism)
 Irisation
 Isochromatic Color Confusions
 Itten, Johannes
 ITULAB
 Ives, Frederic Eugene

J

Judd, Deane Brewster

K

Katz, David
 Kohlrausch, Arnt

Kries, Johannes Adolph von
 Kruithof Curve
 Kubelka, Paul

L

Lab
 Labradorescence
 Lambert, Johann Heinrich
 Land Retinex Theory
 Lanthanide Ions
 Lanthanides
 Lanthanoid Ion Color
 Layering
 Le Blon, Jacob Christoph
 Light Bounces
 Light Depreciation
 Light Distribution
 Light Pollution
 Light Sensors
 Light Stimulus
 Light, Electromagnetic Spectrum
 Light-Emitting Diode, LED
 Light-Emitting Diode, OLED
 Lighting Control Systems
 Lighting Controls
 Lighting Design
 Lighting for Indoor Spaces
 Lighting for Interior Spaces
 Lighting Management
 Linguistic Influences on Color Perception
 Linguistic Relativism
 Lippmann, Jonas Ferdinand Gabriel
 Lorenz-Mie Theory
 Lorenz-Mie-Debye Theory
 Lovibond, Joseph Williams
 Low-Pressure Glow Discharge Lamps
 Low-Pressure Mercury Gas Discharge Lamps
 Lumen Depreciation
 Luminaires
 Luminance Measurement Device
 Luminance Meter
 Luminescent Materials
 Luminous Efficiency Function
 Luminous Intensity Distribution
 Luminous Intensity Meter
 Luxmeter

M

M, P, and K Pathways
 MacAdam, David L
 Mach Bands
 Magno-, Parvo-, Koniocellular Pathways
 Mass Pigmenting
 Matting
 Maxwell, James Clerk
 Mayer, Tobias
 Melanopsin Retinal Ganglion Cells
 Memory Color
 Mesoamerican Color Survey Digital Archive
 Metal Halide Lamp
 Metamerism
 Mie Solution
 Mie Theory
 Minimally Distinct Border
 Mixed Adaptation Condition
 Mixed Chromatic Adaptation
 Modern Art and Fashion Design
 Monochromator
 Mosaics
 Motion and Color Cognition
 mRGCs
 Multilingual/Bilingual Color
 Naming/Categories
 Munsell, Albert Henry

N

Natural Color Distribution
 Natural Colorants
 Natural Dyes
 Natural Scene Statistics
 Neo-impressionism
 Neon Color Spreading
 Neon Lamp
 Newton, (Sir) Isaac
 Nickerson, Dorothy
 Nil
 Non-Photorealistic Rendering
 Non-Visual Lighting Effects and Their Impact
 on Health and Well-Being

O

Op Art
 Opacity – Extinction
 Opalescence

Opponent Color Theory
 Opsin Genes
 Optic Chiasm, Chiasmal Syndrome
 Optical Art
 Optical Radiation
 Optical Spectrum
 Optoelectronic Light Source
 Ostwald, Friedrich Wilhelm

P

Paint
 Painted Dress
 Paints
 Pale Colors
 Palette
 Palmer, George
 Parafovea
 Parafoveal Vision
 Parallel Visual Pathways
 Pastel Colors
 Pastel Crayons
 Pastel Paintings
 Pastel Sticks
 Pastel Tones
 Pattern-Induced Flicker Colors
 Pearlescence
 Perceived Reflectance
 Perceived Value
 Perception by Night
 Perceptual Appearance
 Perceptual Color Effects
 Perceptual Grouping and Color
 Philosophy of Color
 Phosphors and Fluorescent Powders
 Photodetector
 Photoreceptor Cells
 Photoreceptors, Color Vision
 Photosensitive Retinal Ganglion Cells
 Photosensors
 Physiologic Scotoma
 Pigment, Ceramic
 Pigment, Inorganic
 Pigments
 Plasmonics
 Polychromatism
 Polychromy
 Preference for White in Asia

Prelinguistic Color Categorization/Categorical
 Perception
 Premetallized Acid Dyes
 pRGCs
 Primary Colors
 Primate Color Vision
 Primitive Colors
 Principal Colors
 Printing
 Protanopia
 Pseudoisochromatic Plates
 Psychological Color Effects
 Psychological Color Space and Color Terms
 Psychological Impact of color
 Psychological Primary Colors
 Psychologically Pure Colors
 Public Lighting
 Punctum Cecum
 Pure Colors
 Purkyně, Jan Evangelista

Q

Quartz Halogen Lamps

R

Rare Earths
 Rare-Earth Elements
 Rare-Earth Ions
 Rational Models of Color Category
 Learning
 Rayleigh and Mie Scattering
 Red Blindness
 Red-Green Deficiency
 Reference Standard Lamp
 Reflectance Standards
 Relational Color Space
 Relative Visibility Function
 Rendering Natural Phenomena
 Retinal Art
 Retinex Theory
 Retinogeniculate Pathways
 Retinogeniculocortical Pathways
 Richter, Manfred
 Road Lighting
 Roadway Lighting
 Runge, Philipp Otto
 Rushton, William A. H.

S

Saltzman, Max
 Sampling Problems in Computer Graphics
 Sapir-Whorf Hypothesis
 Schiller
 Schrödinger, Erwin
 Scintillating Scotoma
 Screening Plates
 Simple Colors
 Simulation
 Simulation and Image Synthesis of Natural
 Phenomena
 Simultaneous color contrast
 Sky Glow
 Sodium Gas Discharge Lamps
 Soft Colors
 Solid-State Light Source
 Sparkle
 Spectral Luminous Efficiency
 Spectral Power Distribution
 Spectral Power Distribution Curves
 Spectrometer
 Spectroradiometer
 Spectroreflectometer (for Reflectance
 Measurement)
 Sports Lighting
 Stain
 Standard Daylight Illuminants
 Standard Illuminating and Viewing Conditions
 Standard Lamp
 Standard Measurement Geometries
 Stanziola, Ralph A
 Stereo and Anaglyph Images
 Stereo Pairs
 Stereograms
 Stevens, Stanley Smith
 Stiles, Walter Stanley
 Street Lighting
 Stroop Effect
 Stroop Interference
 Structure of Color Preference
 Structure Property Relationships
 Subjective Colors
 Subsampling
 Surface Plasmon Polariton
 Surface Plasmon Resonance
 Surface Plasmons

T

Tactile Painting and Haute Couture
 Temporal Collective Color Preferences
 Tendencies
 Tessellations
 Tesserae
 Tetrachromatic Vision
 Textile Processing
 Textile Wet Processing
 Texture Measurement, Modeling, and Computer Graphics
 Tincture
 Tint
 Total Appearance
 Transition-Metal Ion Colors
 Transparency
 Trichromacy
 Tristimulus Colorimeter
 Tritanopia
 Tubular and Compact Fluorescent Lamp
 Tungsten Filament Lamp

U

UCS Diagram
 Ulbricht Sphere
 Ultraviolet Radiator
 Uniform Chromaticity Scale Diagram
 Unique Hues
 Unwanted Light Especially Referring to
 Unwanted Electric Lighting at Night
 USC Diagrams; Uniform Chromaticity Scales; $Y_u'v'_u'$

V

$V(\lambda)$, $V'(\lambda)$
 Vantage Theory of Color
 VEP
 Visual Appearance
 Visual Contamination
 Visual Evoked Potentials
 Visual Illusions
 Visual Pigments
 Visual Pollution
 von Bezold Spreading Effect
 Vora Measure, v
 Vora Value
 VT and Color

W

Wet or Chemical Finishes
 White Balancing
 Wide-Dynamic-Range Imaging
 Wood's Lamps
 World Color Survey
 Wright, David
 Wyszecki, Gunter

X

Xenon Lamp
 X-Linked Inherited Color
 Vision Deficiency
 XYZ