Human colour perception. A psychophysical study of human colour perception for real and computer-simulated two-dimensional and three-dimensional objects.

Item Type	Thesis
Authors	Hedrich, Monika
Rights	 The University of Bradford theses are licenced under a Creative Commons Licence.
Download date	2026-05-12 11:58:58
Link to Item	http://hdl.handle.net/10454/4304

Table of contents

1 Introduction

In this thesis a series of psychophysical experiments are reported aiming to investigate different aspects of human colour perception. Experiments have been carried out on colour memory, colour constancy and colour categorisation.

Although colour is a psychological phenomenon it is usually considered as an object property. Colour terms are often used to describe objects more precisely, e.g. blue car, red lipstick, black folder and so on, and some objects are even closely associated with a specific colour, e.g. green grass, red roses, etc. In order to be able to use colour terms consistently and persistently colours have to be remembered. However, solely remembering colours would not be enough to recognise, for example, our favourite pink shirt outside in bright sunlight and inside the office under artificial light. From a physical point of view the light that reaches our eyes from this shirt in these situations is very different but in both situations we perceive its colour as the same. Our visual system compensates for such illumination changes and enables us to perceive the environment as stable in the sense of colour, or in other words, to perceive the environment as colour constant.

Besides colour memory and colour constancy there is another crucial component in colour perception, colour categorisation. Humans are able to distinguish more than 2 million different colours but have only a very limited number of words to describe them. Consequently, a colour term describes not only a single colour experience but many. In everyday life this rough categorisation of colours has been proven to work efficiently because under most circumstances it is sufficient to know the colour category a colour belongs to. Imagine your friend tells you on the phone that he bought a new car and that its colour is red. Without having seen the actual car you have a clear idea what its colour looks like. Colour categorisation also plays an important role in colour memory and colour constancy. In most situations it is sufficient to remember the colour category of a previously seen object in order to recognise it, although the illumination under which it is viewed has changed.

In the present study emphasis is placed on the direct comparison of observers' colour memory, colour constancy and colour categorisation performances for (a) real and computer-simulated stimuli and (b) two-dimensional and three-dimensional objects. This is a novel methodological approach as most previous studies in this field have used solely a single stimulus configuration. It allows studying the effect of dimensionality on colour perception and investigating whether displayed stimuli are valid substitutions for real ones.

The overall aim of the present study is to investigate various aspects of human colour perception in colour normal participants. Therefore an observer screening is carried out in which observers' colour vision is assessed. As part of the screening process observers' colour memory is also tested to ensure that all participants have roughly equal colour memory.

The colour constancy experiments aim to investigate the effect of (a) dimensionality of the stimuli and (b) different illuminant changes on observers' performance. For (a), it is hypothesised that observers' performance is better when colours are learned as part of a three-dimensional instead as part of a two-dimensional environment, because the three-dimensional environment provides the visual system with more information/cues that are necessary to achieve colour constancy. Colour constancy is assessed for different illuminant change conditions - (b) - to ensure the generality of findings. To ensure also that differences in performance are not due to differences in memorability of colours, colour memory is tested for the same illuminants under which colour constancy is then test.

Another major objective of this study is to investigate whether computer-simulated and displayed stimuli are valid substitutions for real stimuli. Many experiments in vision research are conducted using displayed stimuli but so far, only few studies have investigated whether observers' performance is dependent on the actual kind of the presented stimuli. Colour memory and colour constancy experiments are therefore conducted using displayed as well as real stimuli. As indicated above, colour categories are an important feature of human colour vision and an extensive experiment is carried out in which colour naming for displayed and real stimuli is compared. Overall it is expected that these experiments lead to a general conclusion about the validity of substitution of displayed and real stimuli.

1.1 Overview of the following chapters

Chapter 2 gives a brief overview of the physiology of colour vision and refers to the theories of colour vision. The next chapter (3) describes the principles of colour spaces, the development of the CIE 1931 XYZ space and focuses in more detail on the Natural Color System (NCS), the CIE 1976 L*a*b* and RGB colour spaces. These spaces were selected because of their direct relevance for this study. Coloured paper samples from the Natural Color System were used to create experimental stimuli and all analyses regarding colour differences were conducted in the CIE 1976 L*a*b* colour space for object colours. A description of the RGB space was included as all real-world stimuli were also replicated and displayed on a monitor.

Chapter 4 provides a review of relevant literature and background information about colour memory and concentrates on the deterioration of colour memory over time and the effect of prior experiences on the memorability of colours. In chapter 5 an overview of the concept of colour constancy is given and several methodological and conceptual approaches are discussed in detail. Colour categories (chapter 6) are discussed with emphasis on their importance for colour vision research.

One of the primary objectives of the present study is to investigate observers' performance for real-world and computer-simulated stimuli. In this context the first part of chapter 7 reviews selected studies that are concerned with the quality of colour reproduction on monitors. The second part focuses on studies comparing directly observers' performances for real and displayed stimuli.

The methods are described in chapter 8. That chapter is divided into two major sections, the first one concentrates on the real-world setup and includes a detailed analysis of the experimental colours regarding their appearance. The second part provides details concerning the generation of the computer-simulated stimuli. Note that chapter 8 contains all information about the methods used in this study and that there are no further method sections in the following experimental chapters.

All experiments are reported in chapters 9 to 12, starting with the initial observer screening (9), which is then followed by the chapters about the colour memory (10), colour constancy (11) and colour categorisation experiments (12). General findings are subsequently discussed in chapter 13. This thesis concludes with a list of references and an appendix, which includes several tables with raw data and a list of publications resulting from the present study.

2 Physiology of colour vision

The human visual system is sensitive to electromagnetic radiation with wavelengths between approximately 380 to 780 nm. The process of vision starts when radiation of this wavelength range is absorbed by the photoreceptors in the retina.

The human retina contains two types of photoreceptors, called rods and cones. There are approximately 120 million rods and 6 million cones, which are unevenly distributed throughout the human retina (Curcio, Sloan, Packer, Hendrickson, & Kalina, 1987; Osterberg, 1935). Rods are completely absent from the centre of the fovea. Moving away from this area the number of rods increases and reaches its maximum at about 15 to 20 deg eccentricity before the density gradually declines again. Rods are responsive to low light levels i.e. in scotopic conditions. Cones have their highest concentration in the fovea. Outside the fovea their density decreases rapidly and reaches at approximately 10 to 15 deg eccentricity a constant level. Cones are responsive to high light levels i.e. in photopic conditions, and are fundamental for visual acuity and colour vision.

Within the cones three different photopigments have been identified. Each photopigment absorbs light over a broad range of wavelengths but has its maximum sensitivity at a different wavelength. The cones are named according to the wavelength range of their peak sensitivity: S-cones have their maximum sensitivity in the short wavelength range at approximately 420 nm, M-cones in the mid wavelength range at 534 nm and L-cones in the long wavelength range at 563 nm (Bowmaker & Dartnall, 1980). The sensitivity curve of a cone class is defined by the probability of a photon to be absorbed as a function of wavelength (figure 2-1). After a photon of a certain wavelength is absorbed this energy is transduced into an electrical signal by a complex photochemical reaction. This signal does no longer carry separate information about intensity and wavelength but only about the number of photons that have been absorbed by the cone (Principle of Univariance) (Naka & Rushton, 1966).

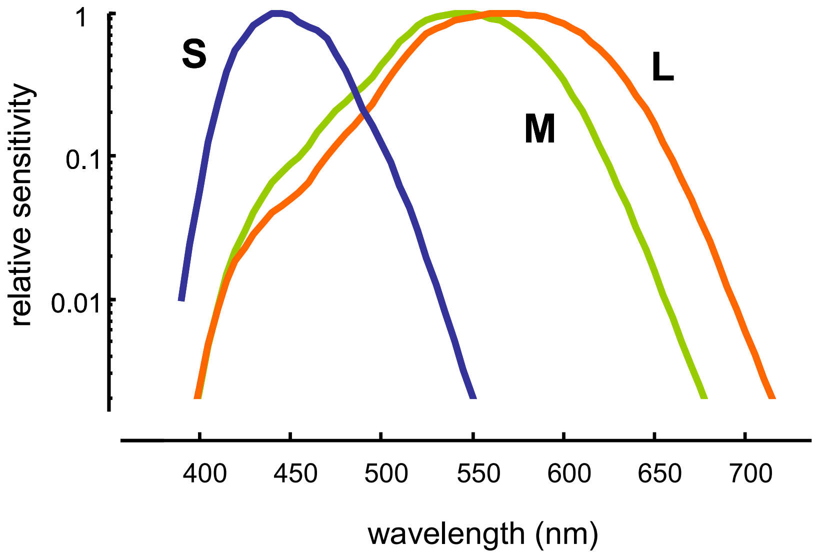


Figure 2-1. The relative sensitivities of the S-, M- and L-cones as a function of wavelength (derived from Stockman & Sharpe, 2000). Each curve is normalized to its maximum. Note that the peak sensitivities of the cones are shifted with respect to the ones mentioned above. The present data was obtained using psychophysical measurements in contrast to microspectrophotometry as used by Bowmaker & Dartnall (1980). The shifts are due to the transmission characteristics of the media of the eye.

The signals released by cones and rods are sent via bipolar cells to the ganglion cells. The information is not passed through directly as other cells such as horizontal and amacrine cells interact. Most ganglion cells receive their input, however, not from a single photoreceptor but from many. These neural networks allow condensing the signals of 126 million photoreceptors onto approximately 1 million nerve fibres through which the information leaves the eye.

The area in the visual field in which the presentation of a stimulus causes a response of a given cell is called receptive field. Receptive fields are circular and their size varies across the retina (increasing with eccentricity). In the periphery of the retina they can be up to 50 times larger than those in the fovea. The vast majority of ganglion cells have centre-surround receptive fields. That means that the quantity of light stimulating the receptors in the centre of the receptive field is compared with the quantity of light stimulating the receptors in the area surrounding the centre. There are two types of centre-surround ganglion cells, ON-centre and OFF-centre cells. ON-centre ganglion cells are excited when light strikes the centre of their receptive field and inhibited when light strikes the surround. OFF-centre ganglion cells, in turn, respond with a decrease of activity when light strikes the centre of the receptive field and with and increase of activity when the surround is stimulated (Schiller, Sandell, & Maunsell, 1986).

Ganglion cells differ not only with respect to centre type and field size but also in the actual cell type. The three most important and numerous types are midget, parasol and bistratified retinal ganglion cells. They are crucial for colour vision but carry also temporal and spatial information. Midget cells receive input from L- and M-cones. Information from central L- and M-cones is subtracted (chromatic opponency channel red/green) whereas information from peripheral L- and M-cones is added (contributing to the achromatic or luminance channel) (Derrington, Krauskopf, & Lennie, 1984). Midget cells project to the parvocellular layers of the lateral geniculate nucleus (LGN) and are therefore often called P cells. With approximately 80% they are the most numerous ganglion cell type on the retina. Parasol cells receive also input from M- and L-cones (L+M) (achromatic channel) and project to the magnocellular layers of the LGN (and are therefore often called M cells). They account for between 8 to 10 % of the ganglion cell population. The receptive fields of parasol cells are larger than those of midget cells. Bistratified or K cells account for less than 10 % of the retinal ganglion cells (Perry, Oehler, & Cowey, 1984); they receive input from S-cones as well as M- and L-cones (chromatic opponency channel blue/yellow). They project to the koniocellular layers of the LGN (Hendry & Reid, 2000).

Regarding colour vision the ganglion cells can be described as chromatically selective as their responses depend on differential inputs from the three cone classes. Figure 2-2 shows a schematic of the contribution of the three cone classes to colour opponency.

After processing visual information signals are transmitted from the ganglion cells to the LGN in the thalamus. The LGN consists of six major layers, numbered form 1 to 6.

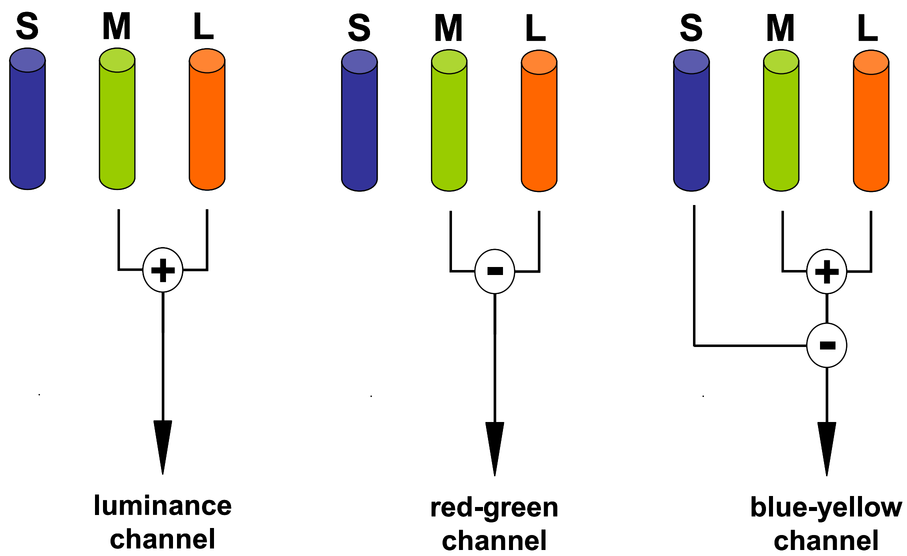


Figure 2-2. The colour opponent process. The luminance channel is formed from the sum of L- and M-cone signals (L+M); the red/green channel from the comparison of L- and M-cone signals (L-M); the blue/yellow channel from the comparison of S-, M- and L-cone signals (S-(L+M)). Adapted from Kaiser & Boynton (1996).

Layers 1 and 2 are magnocellular and receive input from the parasol ganglion cells, and all other layers (3 to 6) are parvocellular. Each of these six layers contains a koniocellular sublayer. The LGN receives input from both eyes; half of the layers (three main and three koniocellular layers) receive input form the contralateral eye, the other half from the ipsilateral eye. Figure 2-3 shows a schematic of the LGN.

Cells of the LGN have centre-surround receptive fields, which are very similar to those of ganglion cells. Characteristics of the retinal ganglion cells, such as colour opponency, are retained in the LGN as well as the strict segregation of the pathways. It appears that the LGN modulates the course of information from the retina to the visual cortex without processing the incoming signals considerably. However, it is important to note that the LGN is more than a simple relay station. It receives feedback from the visual cortex and other cortical areas (Montero, 1991) and that indicates a more complex role of the LGN (Derrington, 2001).

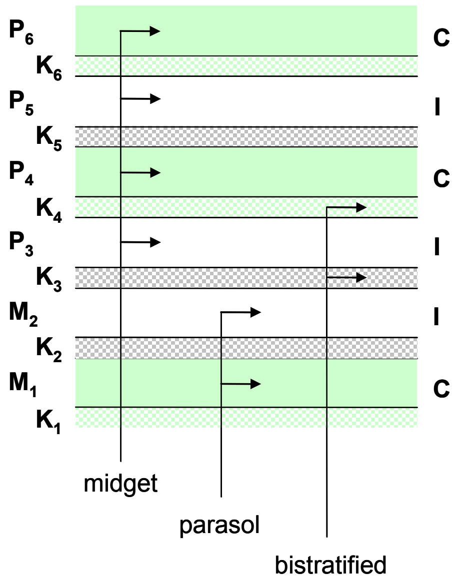


Figure 2-3. Schematic of the LGN. The major layers M_2, P_3 and P_5 (and the koniocellular layers K_2, K_3 and K_5) receive input from the ipsilateral (I) and M_1, P_4 and P_6 (and the koniocellular layers K_1, K_4 and K_6) from the contralateral (C) eye. Midget, parasol and bistratified ganglion cells of both eyes project onto different layers of the LGN, their pathways remain strictly segregated.

While neural processing of visual information is often described as being rather complex within the early stages of the visual system it becomes much more complex at the higher (cortical) stages. From the LGN, signals are transmitted to the primary visual cortex (V1), the first stage of cortical processing. V1 consists of six layers, which are numbered 1 to 6, starting from the layer nearest the surface. Layer 4 is divided into sub-layers named 4A, 4B, 4Cα and 4Cβ. Fibres from the LGN terminate predominantly in layer 4C (magno in 4Cα and parvo in 4Cβ) whereas koniocellular layers project to the superficial layers 2 and 3 (Hubel & Wiesel, 1977; Livingstone & Hubel, 1987). The incoming information is transmitted within as well as between the layers of V1 (Callaway, 1998).

Around 30 regions have been identified in the cortex to respond to colour (V2, V3, V4, MT, etc.) and it has been shown that complex interactions take place between them. This indicates that there is no single colour centre in the cortex, however, it remains unclear were ultimately the perception of colour is formed.

2.1 Theories of colour vision

In the nineteenth century two major colour theories had been introduced that account for colour vision, the trichromatic theory and the opponent-process theory. These two theories seemed to be contradictory to one another for a long time. It lasted until the middle of the twentieth century until it was discovered that both theories are compatible.

The trichromatic theory was first proposed by Thomas Young in 1802 and later revived by Hermann von Helmholtz. Their theory was solely based on observations from light mixing experiments rather than on anatomical studies of the eye. Young and Helmholtz had observed that any colour could be produced by mixing the light of three primary colours (red, green and blue). Although neither of these two scientists did know anything about the physiology of the retina they presumed that there were three different types of receptor with overlapping spectral sensitivities. This pioneering idea was later confirmed by the discovery of the three cone classes (e.g. Marks, Dobelle, & MacNichol, 1964; Nathans, Darcy, & Hogness, 1986) and the trichromatic theory became also known as the Young-Helmholtz theory.

The opponent-process theory was proposed firstly by the physiologist Ewald Hering and is based on the appearance of colour. He suggested, after observing the effect of afterimages, that colour is processed by the visual system in terms of opponent colour pairs. When we stare, for instance, at a red circle for a while and then look at a white piece of paper, we see green, the opponent colour of red. Hering specified originally two pairs of opponent colours, red and green, and blue and yellow. However, a third opponent pair, black and white, has been included as brightness is transmitted in a similar way as colour. Hering also noted that the perception of opponent colours is mutually exclusive; we never experience yellowish blue or greenish red. Physiological evidence that supports Hering's opponent-process theory was first found by Hurvich and Jameson (1957). Since then extensive research has been carried out to investigate the opponent colour processing in the visual system (e.g. De Valois, Abramov, & Jacobs, 1966;De Valois, De Valois, Switkes, & Mahon, 1997; Derrington et al., 1984).

Nowadays it is generally accepted that both theories of colour vision are partially correct and that colour processing occurs in two stages (De Valois & De Valois, 1993). In the first stage, on the receptor level, the Young-Helmholtz theory is correct with regard to the existence of the three cone classes and the subsequent transmission of the signal to the cortex, the second stage, can be explained by the opponent-process theory.

3 Colour spaces

3.1 Introduction

The idea to organise colours in logical arrangements has a long tradition (for review see Kuehni, 2003) and numerous colour spaces have been developed to satisfy very specific needs when dealing with colour. An internationally well-established standard for colour specification and communication is the CIE[1] 1931 XYZ colour space, which is based on the tristimulus values. Other colour spaces such as CIE 1976 L*a*b* (CIELAB) and CIE 1976 L*u*v* (CIELUB) have been introduced to facilitate the evaluation of colour differences; CIELAB was recommended to be used for the specification of surface colours and CIELUB for the specification of lights and self-luminous sources. The so far mentioned spaces are derived form colour matching but there are others that are derived from physiological data. Two well-established physiological colour spaces are the MacLeod-Boynton (MacLeod & Boynton, 1979) and the DKL (Derrington et al., 1984) colour spaces.
In this chapter the development of the CIE 1931 XYZ colour space is discussed as well as the CIE 1976 L*a*b* colour space, the Natural Color System and the RGB space. These colour spaces are explained in more detail because of their direct relevance to the present study.

[1] Commision Internationale de L'Eclairage – International Commission on Illumination

3.1.1 Arranging colour in a three-dimensional space

Each colour experience can be described by the three attributes hue, saturation and brightness. Hue refers to the actual colour such as blue, red, yellow. Saturation describes the purity of a colour, in other words, how much white it contains. Less saturated colours are known as pastel colours and a light blue, for instance, is nothing else than a poorly saturated blue. The third attribute is brightness and refers to the perceived intensity of the light. Hue, saturation and brightness are best illustrated in a three-dimensional space, as shown in figure 3.1.1-1, where each combination of these attributes corresponds to a single point in the space.

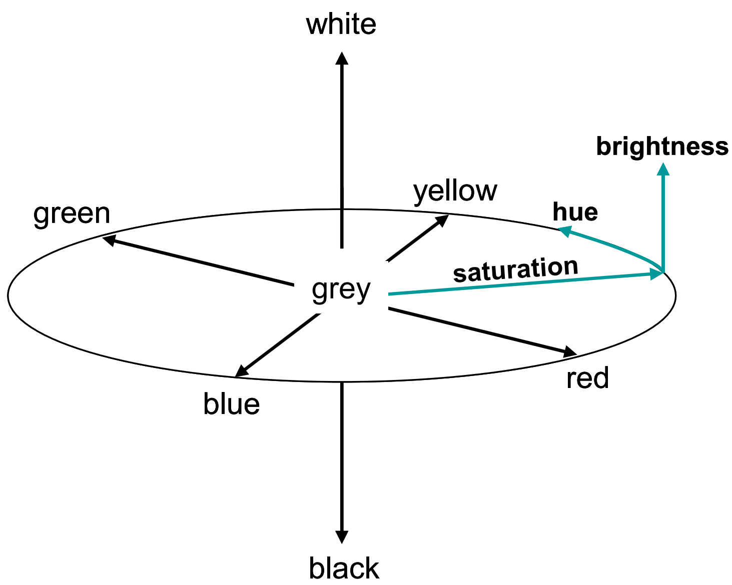


Figure 3.1.1-1. Schematic of a perceptual colour space. Brightness varies along the vertical axis and all hues lie in a horizontal plane. Saturation varies radially and decreases from the circumference to the centre of the circle.

3.2 Development of the CIE 1931 XYZ colour space

As mentioned earlier, a colour normal observer can match any colour by mixing no more than three primary colours. The primary colours, usually some kind of red, green and blue, are chosen arbitrarily. The only requirement is that a primary colour cannot be matched by a combination of the other two primaries. W. D. Wright (1946) chose, for example, primaries with the wavelengths of 650 nm (red), 530 nm (green) and 460 nm (blue) to determine colour matching functions. In his experiment the observer was presented with a bipartite field. In one half a spectral colour sample was presented and observers adjusted the intensities of each of the three primaries of the other half until the two halves appeared to be identical. Wright then determined colour matching functions with the amount of each primary that was needed to match the spectral light; the relative amount of each primary needed to match the wavelength of a colour sample is called tristimulus value.

Whenever a different set of primaries is chosen the colour matching functions change. The use of many different colour matching functions would be impractical because the tristimulus values alone would be insufficient but information about the primaries that were used must also be provided. In order to establish a unified colour system the CIE introduced in 1931 colour matching functions based on the primaries 700 nm (red), 546.1 nm (green) and 435.8 nm (blue). These functions are denoted by $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$, and the corresponding tristimulus values are called R, G and B. One particularity has to be mentioned. Most monochromatic colour samples can be matched by adding relative intensities of all three primaries. However, occasionally a negative amount of a primary is needed. That means that in order to match two colours, the observer has to add an amount of one of the primaries to the reference colour sample to achieve a match. This “negative” effect occurs regardless of which primary colours are chosen. It became evident that colour matching functions with only positive values would be more convenient for practical purposes. Therefore, the CIE transformed the RGB into the XYZ system, which is based on so-called imaginary primaries. The tristimulus values of this new system were called X, Y and Z and the colour matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$, respectively. The $\bar{y}(\lambda)$ colour matching function was chosen to match the photopic luminous efficiency function[1] (V_λ) as defined by the CIE. This means that the luminance of a colour stimulus can be determined by calculating its Y tristimulus value. Hence, two stimuli with the same Y tristimulus values have equal luminance, regardless of their X and Z values.

The tristimulus values XYZ are computed by multiplying a given spectral power distribution P_λ with the colour matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ (integrating over the visible spectrum); k is a normalization factor (Wyszecki & Stiles, 2000).

$$X = k\int P_\lambda \bar{x}(\lambda)d\lambda \qquad Y = k\int P_\lambda \bar{y}(\lambda)d\lambda \qquad Z = k\int P_\lambda \bar{z}(\lambda)d\lambda \tag{1}$$

The chromaticity coordinates are defined by:

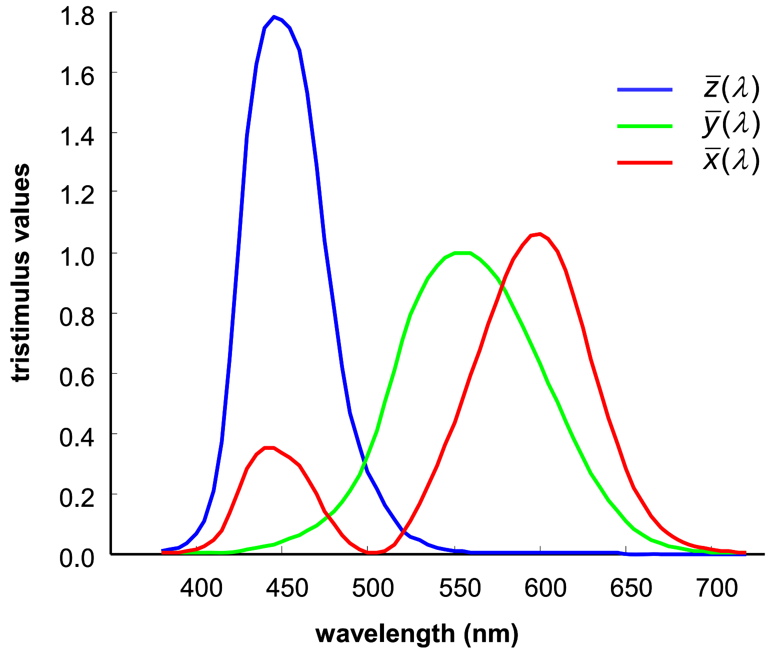
$$x = \frac{X}{X+Y+Z} \qquad y = \frac{Y}{X+Y+Z} \qquad z = \frac{Z}{X+Y+Z}$$

and thereby

$$x + y + z = 1 \tag{2}$$

The CIE 1931 colour matching function and the CIE 1931 xy chromaticity diagram are shown in figure 3.2-1.

[1] The photopic luminous efficiency function describes the mean spectral sensitivity of the human eye.



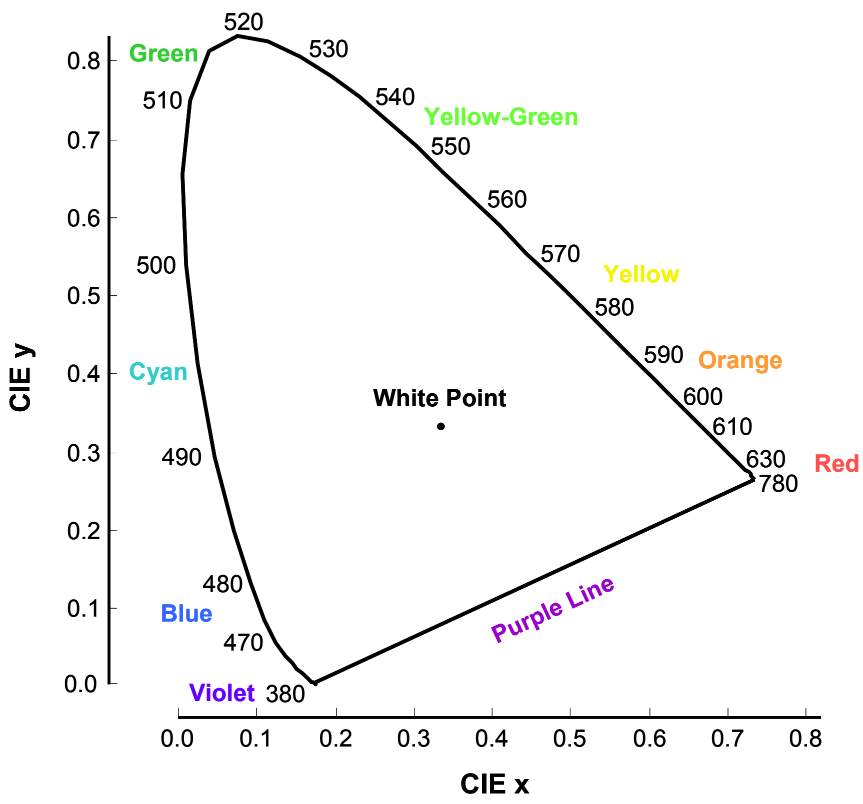


Figure 3.2-1. Top - CIE 1931 2-deg colour matching functions. Bottom - CIE 1931 xy chromaticity diagram. Spectral colours lie on the locus. The straight line that connects the limits of the spectral locus is called purple line. Colours along the purple line are mixtures of red and blue/violet light. Equal energy white is at xy (0.33, 0.33).

3.2.1 Metamers

In colour matching experiments generally two stimuli have to be adjusted until they are perceptually undistinguishable. Most commonly, one stimulus is monochromatic whereas the other one is a mixture of different wavelengths. In other words, the two stimuli appear identical despite different spectral composition. Such perceptually identical stimuli evoke the same tristimulus values and are referred to as metamers.

3.2.2 Blackbody locus

An often used reference point in the CIE xy chromaticity diagram is the white point (equal energy white). However, it is also common to relate/specify colours with respect to the Planckian or blackbody locus. A blackbody radiator emits light depending on its temperature, and for each temperature chromaticity and luminance coordinates can be computed. When the chromaticity coordinates are plotted into the CIE xy chromaticity diagram a continuous line is obtained, which is called blackbody locus (figure 3.2.2-1). Because of the correlation between colour and temperature points along this locus are specified by their correlated colour temperature.

The chromaticities of daylight and also the likes of some artificial light sources, for instance, candle light and tungsten light, are located on or near the blackbody locus. Changes between those are frequently experienced in our daily routine (e.g. change between daylight and tungsten light). Illuminants on the blackbody locus and changes between them will therefore be referred to as typical. Illuminants with chromaticities that do not lie on the blackbody locus will be referred to as atypical as well as changes between an illuminant on and one off the blackbody locus (e.g. change between daylight and a purple illuminant), because they are hardly ever experienced.

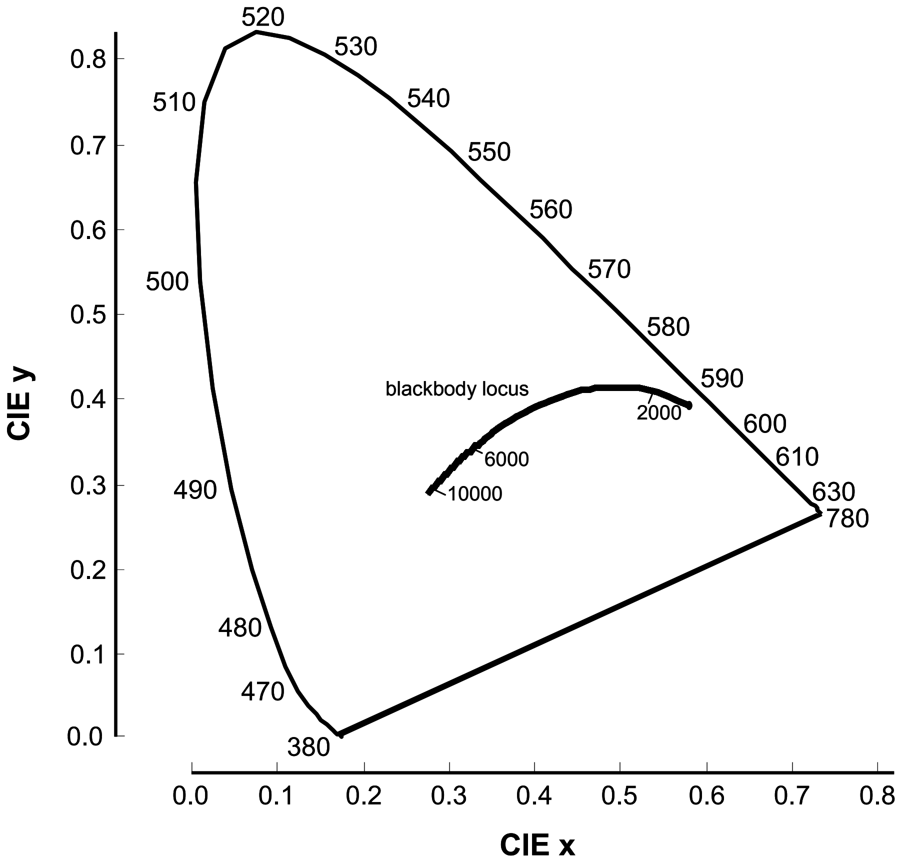


Figure 3.2.2-1. Location of the blackbody locus in the CIE xy chromaticity diagram. The blackbody radiator emits red light at low temperatures and at about 5600 K it emits almost perfectly white light; at higher temperatures the light becomes slightly bluish. The positions of the following colour temperatures are indicated: 2000, 6000 and 10,000 K.

3.3 CIE 1976 L*a*b*

The CIE xy chromaticity diagram was derived from colour matching behaviour and Euclidean distances do not reflect how easy it is to discriminate between two colours. In one region of the CIE xy chromaticity diagram two colours, which are separated by a certain distance, may be undistinguishable whereas in another region, two colours separated by the same distance are perceived as clearly different.

Fundamental studies addressing chromaticity discrimination were carried out by David L. MacAdam, of the Eastman Kodak Company. In order to specify the typical (trademark) yellow of the company MacAdam investigated by how much the tristimulus values of two colours had to deviate from each other to be perceived as different (MacAdam, 1942). Notice that he focused on the tristimulus values and not on physical characteristics of colours; colours may have different spectral reflectances and still evoke the same tristimulus values (i.e. metamers).

MacAdam studied colour discrimination not only for the “Kodak yellow” but for a wide range of different chromaticities, chosen from all over the CIE xy chromaticity diagram. Each reference colour was matched several times and afterwards an ellipse was fitted to the data. All chromaticity within such an ellipse (which is nowadays called a MacAdam ellipse) are undistinguishable from the reference colour which is located in the centre. These ellipses for all colours tested differ considerably in orientation and size; they are largest in the greenish and smallest in the bluish region of the chromaticity diagram. The MacAdam ellipses illustrate clearly that the CIE xy chromaticity diagram is not the appropriate colour space when it comes to perceptual colour differences.

In 1976 a new colour space was therefore introduced, the CIE 1976 L*a*b* or CIELAB colour space. This space is approximately perceptually uniform, where the spacing between two colours reflects the perceived difference between them. A non-linear transformation of the CIE 1931 XYZ space led to the new space with a lightness component L* and two chromaticity components a* (red/green) and b* (blue/yellow).

The quantities L*, a* and b* are defined by:

$$L^* = 116 \cdot \left(\frac{Y}{Y_n}\right)^{1/3} - 16$$

$$a^* = 500 \cdot \left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Y}{Y_n}\right)^{1/3}\right]$$

$$b^* = 200 \cdot \left[\left(\frac{Y}{Y_n}\right)^{1/3} - \left(\frac{Z}{Z_n}\right)^{1/3}\right] \qquad (3)$$

where X_n, Y_n and Z_n are the tristimulus values of a white point (Wyszecki & Stiles, 2000).

In the CIE L*a*b* space equally perceived colour differences are expressed by the same distance in the colour space. This distance is called ΔE_{ab} and is the Euclidean distance between two points. ΔE_{ab} is calculated from:

$$\Delta E_{ab} = \left[(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2\right]^{1/2} \qquad (4)$$

As mentioned above, the CIE 1976 L*a*b* space is only approximately perceptually uniform. More recently new colour difference formulae with improved perceptual uniformity have been introduced, for example CIEDE2000. The purpose of CIEDE2000 is to deliver a high degree of perceptual uniformity especially for small colour differences (around the just-noticeable-difference (jnd) between colour stimuli). However, for the evaluation of large colour differences ($\Delta E > 5$ units) the current CIE recommendation is still the CIE 1976 L*a*b* colour difference formula (Westland & Ripamonti, 2004). All colour differences analyses in this thesis are therefore conducted using this formula.

3.4 Natural Color System

The Natural Color System (NCS) was introduced by the Scandinavian Colour Institute (Skandinaviska Färginstitutet AB) in 1979. The system is a colour appearance system based on the unique colour perception of the six elementary colours red, green, blue, yellow, black and white. In his opponent colour theory Hering had already suggested that these six colours cause unambiguous colour sensations in observers. The Natural Color System was developed based on this principal idea. All colours of the system, apart from the elementary colours, can be characterised by judging their resemblance to at least two elementary colours. For example, purple can be specified as a mixture of red and blue, turquoise looks like a mixture of green and blue and brown like a mixture of red, yellow and black. The NCS was developed with the primary objective to obtain a description of colour perception and only secondarily to obtain a perceptually uniform space. For a detailed description of the origin and development of the Natural Color System see Hård, Sivik, & Tonnquist, 1996a, 1996b.

The Natural Color System is not a theoretical system but its colours are physically available in form of coloured papers or paints; at the moment 1950 different colour samples are available. The NCS is widely used in the areas of fashion and textiles, product and transport design, indoor design, architectural colour design and cosmetics among others (Scandinavian Colour Institute: Natural Colour System, 2004).

3.4.1 NCS colour space

The position of the six elementary colours constitute a three-dimensional space that has the shape of a double cone and in which red, green, blue and yellow lie in the same plane. Figure 3.4.1-1 shows a diagram of the NCS colour space.

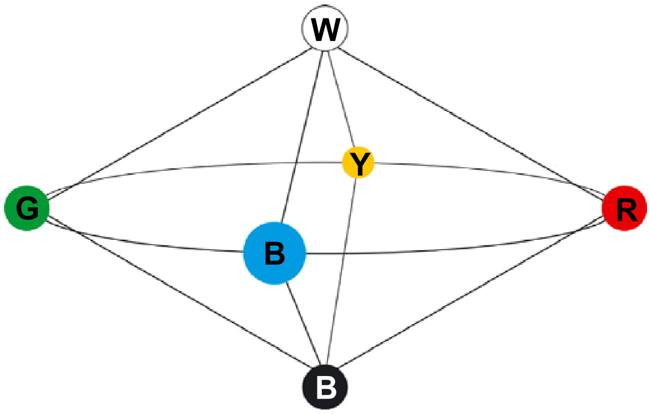


Figure 3.4.1-1. The NCS colour space is set up by six elementary colours. Modified from (Scandinavian Colour Institute: Natural Colour System, 2004).

Colour samples of this space are defined by three characteristics: hue, blackness and chromaticness. Hue refers to the actual colour. Furthermore, each colour sample is defined by the perceived quantity of black in the colour in comparison with pure black; this proportion is expressed by the blackness value. The term chromaticness is used to describe the saturation of a colour sample. All possible hues of the NCS lie on a colour circle, which results from a horizontal cut through the middle of the colour space (figure 3.4.1-2).

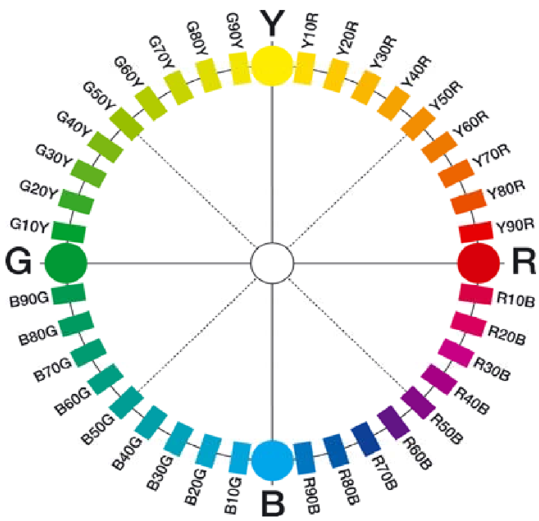


Figure 3.4.1-2. In the NCS colour circle colours change progressively from yellow to red, from red to blue and so on in steps of 10 perceptual units (Scandinavian Colour Institute: Natural Colour System, 2004).

Along the vertical axis of the NCS colour space lie the gray shades, reaching from white at the top to black at the bottom. A vertical cut through the colour space produces a so-called colour triangle (figure 3.4.1-3). There is a colour triangle for each hue. All colour samples of a triangle vary in blackness and chromaticness.

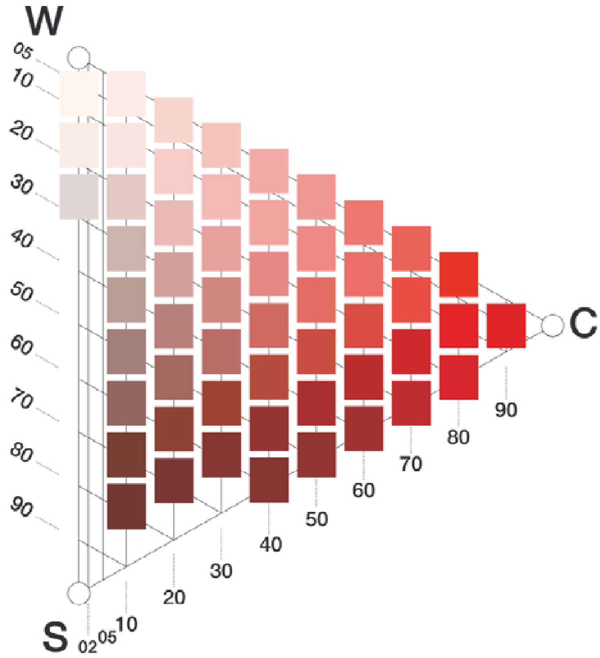


Figure 3.4.1-3. NCS colour triangle for the single hue Y90R. W stands for white, S for black and C for chromaticness. The chromaticness increases from the left to the right. The least saturated colour samples are next to the achromatic line. Blackness increases starting from the line between W and C and moving towards S (Scandinavian Colour Institute: Natural Colour System, 2004).

3.4.2 Colour notation

The NCS colour notation is based on the scheme that every given colour (apart from the elementary colours themselves) can be described as a mixture of two or more of the six elementary colours. The exact notation will be explained by the following example: S2030 – R70B. S indicates that this is a standardised NCS colour sample of the second edition. The number 2030 is called nuance and is a combination of blackness and chromaticness. Twenty indicates that the degree of resemblance to black is 20% and that the chromaticness is 30%. The hue is described by the last term of the notation. R70B means that it is a red (30%) with 70% blue. In other words this colour appears as a light bluish purple.

3.5 RGB colour space

The RGB colour space is a machine space, which arose together with modern display devices. In cathode ray tube (CRT) and colour television colours are generated by stimulating three different types of phosphors with an electron beam. The three phosphors are chosen to emit red, green and blue light (which leads to the abbreviation RGB) when stimulated and all other colours can then be generated by additive colour mixture. The phosphors used in different display devices are not identical in terms of their spectral output and because of this only a limited range of colours can be generated. The range of colours that can be generated is called the gamut of a display device. Colours are specified in terms of R, G and B units of activation and values reach generally from 0 to 255 (8 bit resolution) for each channel. Black can be generated by not stimulating any of the phosphors (RGB values 0 0 0) and white, in turn, is achieved when all phosphors are maximally stimulated (RGB values 255 255 255); grey shades have the same values of the three components and colour have varying values. By activating only the red phosphor, for example, a pure red colour is generated, whose brightness depends on the amount of activation. The brightest possible red is obtained with maximum output of the red channel (RGB values 255 0 0).

CIE XYZ values can be transformed into RGB values by multiplying them with a transformation matrix M, which can be obtained during the calibration process of the display device (equation 5). The calibration process consists of measuring the maximum output individually for each phosphor, determining the XYZ values which then form the entries of the transformation matrix. This transformation assumes a linear relationship between input and output (Berns, 1996; Lindbloom, 1989). However, that is not the case. The monitor’s output values are usually related to the RGB values which can be described by a power function of the form $output = RGB^{\gamma}$. Therefore it is necessary to once determine this power function and correct the RGB values such that input and output are linearly related. This procedure is called gamma correction (International Color Consortium, 1996).

Transformation from XYZ to RGB values (and vice versa):

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} & M^{-1} & \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} & M & \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

where $$M = \begin{bmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Z_R & Z_G & Z_B \end{bmatrix} \quad (5)$$

3.6 Colour appearance models

Another approach to colour specification are colour appearance models. Colour appearance models take viewing conditions, such as the state of adaptation of an observer, the spatial and temporal structure of the stimulus and the perceptual task being performed by the observer, into account as they affect colour perception/discrimination. The objective of colour appearance models is to provide a relation between a specification of a stimulus and the context in which it is viewed and its colour appearance. In other words, a colour appearance model adjusts images, so that an image that is viewed under one condition appears to have the same colours when it is viewed under a completely different condition (Fairchild, 2005).

A wide variety of colour appearance models such as von Kries, Hunt, RLAB, CIELAB, Nayatani, CIECAM97s and CIECAM02 have been developed; each of them having advantages but also limitations. The qualitative comparison of these models and their development is an active field of research, which is referred to in more detail in section 7.1.

4 Colour memory

4.1 Introduction to memory

The importance of memory becomes clear if we try to imagine being without it. Without memory we would be unable to remember our name and we would not be able to recollect any event of our past. We would also be unable to read, write or talk, to follow a conversation or to keep track of a program we see on television. This list could be extended by an almost infinite number of examples. It is evident that our lives would be extremely limited without any kind of memory.

The process of memory is divided into three stages: encoding, storing and retrieving. During encoding, sensory input is transformed into a form of mental representation. The process of retaining information in memory is referred to as storing and the active process of recovering information retained in memory and using it is called retrieval. The process of memory causes physiological changes in the brain; studying these changes is only one approach to understanding memory. Another approach is to think about memory as a process that takes place in several stores and that requires different systems and subsystems. This conceptualisation was introduced to organise different concepts of memory more efficiently and to facilitate the discussion of them. In the literature there are several models of memory. One model, which was proposed by Atkinson and Shiffrin (1968) is based on three types of memory (or stores): sensory memory/store, short-term memory/store and long-term memory/store. The sensory store is the first stage in the process of memory. Its characteristics are that information is hold only very briefly and that the capacity is very limited. Typically we are unaware of the sensory memory and it appears that its main function is to keep information long enough to transfer it to the next stage of memory, the short-term memory. The short-term memory has also a very limited capacity but information is stored for several seconds up to a couple of minutes and can be accessed immediately and effortlessly. However, already small distractions can interfere with the short-term memory and the stored information gets lost immediately. When we talk about memory in an everyday situation we usually refer to long-term memory. In the long-term memory we are able to retain information for very long periods; days, decades or even indefinitely. The question of how large the capacity of the long-term store is cannot be answered easily because there is no method to test its limits. It is evident that its capacity is enormous and some researchers even suggest that it is infinite (e.g. Bahrick, 2000).

Another model was proposed by Baddeley and Hitch (1974) and elaborated in (Baddeley, 1986), who introduced a concept called working memory, which replaced that of short-term memory. Baddeley and colleague suggested that working memory actually contains new information and information retrieved from the long-term memory. In other words, in the working memory information that has just bee obtained is related to information retained in the long-term store. Some very sophisticated and highly complex models of memory have been developed. Though, the diversity of things we can remember (e.g. events, smells, sounds, objects, people, emotions, etc.) forces researchers to concentrate on specific aspects of memory. This chapter focuses on the aspect of memory related to colour vision and the following paragraphs provide an overview of studies on this topic.

4.2 Introduction to colour memory

In colour memory research sensory memory, working memory and long-term memory have been addressed by using different delay-periods between seeing a stimulus for the first time and matching it (Boynton, Fargo, Olson, & Smallman, 1989; Collins, 1932; D'Ath, Thomson, & Wilkins, 2007; Epps & Kaya, 2004; Francis & Irwin, 1998; Hamwi &

Landis, 1955; Jin & Shevell, 1996; Newhall, Burnham, & Clark, 1957; Nilsson & Nelson, 1981; Perez-Carpinell, Baldovi, De Fez, & Castro, 1998a; Romero, Hita, & del Barco, 1986; Uchikawa, 1983). These fundamentally different approaches of testing colour memory impede the comparison of the results. Although the kind of memory involved plays an important role it is only one of many factors that affect the qualitative determination of colour memory. Another important factor is the kind of stimulus that is used to present a colour and whether an observer associates automatically a colour experience with it. Imagine a red circle on a grey background and, for instance, a polar bear. The shape of a circle is usually not associated with a specific colour whereas a polar bear is immediately associated with white. The effect of prior experience with stimulus objects and their familiarity to the observer were the main focus of numerous studies (Amano, Uchikawa, & Kuriki, 2002; Bartleson, 1960; Bodrogi & Tarczali, 2001; D'Ath et al., 2007; Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Heider, 1972; Hurlbert & Ling, 2006; Loftus, 1977; Olkkonen, Hansen, & Gegenfurtner, 2008; Perez-Carpinell, De Fez, Baldovi, & Soriano, 1998b; Seliger, 2001; Siple & Springer, 1983; Tarczali, Park, Bodrogi, & Kim, 2006; Yendrikhovskij, Blommaert, & de Ridder, 1999). A further factor that strongly influences the outcome of colour memory studies is the actual task observers are asked to perform. Sometimes observers have to pick a colour out of a larger selection (Bartleson, 1960; Epps & Kaya, 2004; Heider, 1972; Ling & Hurlbert, 2008; Perez-Carpinell et al., 1998a; Ratner & McCarthy, 1990) or they have to decide whether two colour stimuli are the same (Boynton et al., 1989; Francis & Irwin, 1998; Uchikawa, 1983). A frequently used task is also to adjust a colour sample until it matches another sample (Collins, 1932; D'Ath et al., 2007; Jin & Shevell, 1996; Nilsson & Nelson, 1981; Seliger, 2001) or until it appears achromatic (Olkkonen et al., 2008). Each of these methods has its advantages but unfortunately also some limitations.

Although individual research questions may differ considerably the overall motivation for researchers is to investigate (a) how stable is colour memory over time how and (b) accurate is colour memory and which factors affect its accuracy.

4.3 Colour memory over time

An early study on colour memory was reported by Collins (1932), in which colour memory was assessed for four monochromatic colours (blue, red, green, and yellow) for a delay of 15 sec. A spectrometer was used to present the test colours and for subsequent matching. The observers' task was to reproduce the original colour by changing the setting of the spectrometer. Collins reported that the four monochromatic test colours were not remembered with the same accuracy and that the accuracy in general varied considerably between observers; red and green were matched less exact than yellow and blue. However, no common direction of the shift between the matched and the original colour was found. For example, the matched red, i.e. the remembered red, shifted towards longer and green towards shorter wavelengths. Some of the six observers completed more than 100 trials per test colour over a period of several weeks in order to investigate whether a test colour was learned after completing several settings, i.e. whether accuracy improved with the number of performed settings. The results revealed such an expected learning effect but only for some observers and only for certain test colours. Collins, therefore, was unable to draw generally valid conclusions from her study.

Monochromatic stimuli were also used by Nilsson and Nelson (1981). They investigated whether colour memory deteriorates over time and chose several rather short delays; reaching from 0.1 to 24.3 sec. The 16 test colours were generated by a monochromator, whose settings had to be adjusted after the delay until the new colour appeared to be the same as the original. Nilsson and Nelson found that not all test colours were remembered with the same accuracy. They reported that blue, for example, was remembered more greenish and red more yellow and that the yellow-orange test colours were remembered most accurate. Despite small differences between individual test colours the overall accuracy with which the test colours were remembered was high and did not deteriorate over the different time-delays.

Furthermore, the authors found no difference within the four observers tested as opposed to the study by Collins.

Whether colour memory deteriorates over time and to which degree is controversial. For example, Francis and Irwin (1998) studied colour memory for similar delays (0.1, 1.0, and 10 sec) as Nilsson and Nelson. They, however, investigated not only the effect of delayed matching but also the influence of the context in which stimuli were presented. Francis and Irwin did not use a set of test colours but varied their (polychromatic) test colours from trial to trial to avoid learning effects. In order to study the influence of context, the test colours were presented in three conditions: (a) in isolation, (b) in a meaningful context and (c) in a meaningless context. The matching paradigm was also different from those mentioned above. Two stimuli were presented consecutively, with a delay of either 0.1, 1.0 or 10 sec, and observers judged whether these two stimuli were the same or different.

The results revealed that colours were remembered best when presented in a meaningless context and worst when presented in isolation. In contrast to the findings from Nilsson and Nelson, here the authors found deterioration for longer delays.

Other studies investigated the deterioration of colour memory over time comparing simultaneous with successive matching (Newhall et al., 1957; Perez-Carpinell et al., 1998a; Uchikawa, 1983). In all these studies it was found that memory deteriorated over time, in other words, that a remembered colour shifted with respect to the original. In the study by Newhall et al. (1957) all colour stimuli were generated by a colorimeter. Test colours (25 in total) were shown and observers (three) adjusted the settings of a second colorimeter to reproduce the test colour during the simultaneous and the successive matching. The delayed matching took place after 5 sec. For the successive matching Newhall et al. reported considerable variations in the accuracy with which colours were reproduced. Overall, the authors found that colours were remembered more saturated and slightly brighter than the original ones. Uchikawa (1983) came to the same conclusion with respect to the saturation. In his study two observers were presented with two stimuli (simultaneously or consecutively), which differed in saturation only. For the successive matching the second stimuli appeared after a 3 sec gap. In both cases observers judged whether the second stimulus was less or more saturated as the first one. In total colour memory was assessed for seven monochromatic test colours. The results showed that saturation was overestimated for most test colours, and that observers' performance was more accurate for the simultaneous than the successive comparison. Uchikawa reported, as many researchers before, that the shifts of the remembered colours were neither constant between test colours nor observers.

Newhall et al. as well as Uchikawa had used rather short delays for the successive matching. A study that included simultaneous matching and successive matching for considerably longer delays (15 sec, 15 min, and 24 h) was carried out by Perez-Carpinell et al. (1998a). In their study 100 observers were presented with an isolated Munsell colour chip for 10 sec (in total ten different test colours were used). Then the test colour was removed and after 15 sec a panel with 20 alternative Munsell colour chips was put in view. Observers' task was to select a chip that matched best the one seen before. The same task was repeated after 15 min and 24 h but without seeing the test colour again. The authors found that the selection was performed most accurately during the simultaneous matching and that performance deteriorated over time. Overall they found that orange was remembered most accurately and that dark colours tended to be remembered darker (or the same) and light colours lighter. Perez-Carpinell and colleagues reported substantial variations between observers' performances and between test colours.

D'Ath et al. (2007) and Hamwi and Landis (1955) studied the memory of hue and colour memory, respectively, over much longer periods than those reported above. However, these two studies approached this topic with different methods.

D'Ath and colleagues investigated the limits of memory for the hue of non-spectral colours for successive matching; matching was performed immediately after seeing the test colour, after a period of 1 h and 1 week. Their twelve test colours had equal saturation and lightness. After presenting a test colour for 10 sec on a monitor the displayed colour changed and observers (96) adjusted this new colour until it matched the previously seen one. The adjustment of an arbitrarily displayed colour was repeated after 1 h and 1 week without seeing the original test colour before. To avoid any confusion between test colours over such long periods each observer memorised only one test colour. D'Ath et al. found similar results as others before; colours (or strictly speaking hues) could not be reproduced accurately and accuracy varied considerably between observers. A comparison between the accuracy achieved for immediate and delayed matching revealed a significant decrease for the latter ones. However, the level of accuracy was almost the same for adjustments made after 1 h and 1 week.

Hamwi and Landis presented their eight test colours, which were chosen from the Color Harmony Manual[1], also in isolation. Each observer was asked to memorise several test colours and to pick a matching colour out of a selection of more than 100 colours after 15 min, 24 h and 65 h. Observers completed the same task several times and received feedback on their performance after each trial, i.e. the observers were trained. The authors were surprised by the overall good colour memory they had found but noted substantial differences between observers and test colours. The shifts in hue that had been revealed did no follow a common pattern but it was noted that colours were generally remembered darker than the original. Hamwi and Landis, in contrast to many other researchers, did not find a deterioration of colour memory over time. However, that might have been due to a ceiling effect as a result of the intensive training of the observers.

[1] Based on the Ostwald color system. For detailed information about the Color Harmony Manual see (Granville & Jacobson, 1944).

In a study by Jin and Shevell (1996) colour memory was assessed for a delay-period of 10 min. The authors employed three test colours (simulated Munsell colours shown on a monitor) and observers matched a memorised colour by adjusting the hue and saturation of a test patch, which was presented either on a grey background or as part of an abstract context (figure 4.3-1). The results revealed that the shifts of the remembered colours were similar for both conditions but often larger for the abstract context condition. Shifts were also reported (a) as being different in magnitude for the three test colours and (b) as not occurring in the saturation dimension only.

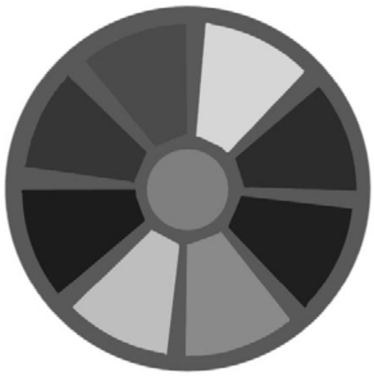


Figure 4.3-1. Test stimulus used by Jin and Shevell (1996), here in grey shades only. The test stimulus (circle in the centre) is embedded into a complex context.

4.3.1 Summary

The primary objective of the above mentioned studies was to investigate whether colour memory deteriorates over time and/or to establish the degree of colour memory after a certain time period. Intuitively one expects memory in general to deteriorate over time and there is no reason to believe that memory for colours would be an exception. This expectation was confirmed by most of the mentioned studies but not by either Hamwi and Landis (1955) or Nilsson and Nelson (1981). Regarding the study by Hamwi and Landis, this might be due to a ceiling effect that occurred because of the intensive observer training. Each observer repeated the task numerous times and received feedback after each matching. Hence, observers could improve their matching behaviour.

Nilsson and Nelson minimised the challenge for the observers in their study by asking to reproduce only one of the three attributes of a test colour, namely hue. In combination with the small number of observers that participated in this study (four; one of them was the first author) it is difficult to draw generally valid conclusions about the deterioration of colour memory, or the lack of it, over time.

Nilsson and Nelson were not the only ones who used a limited matching paradigm. Also in the study by D'Ath et al. (2007) observers adjusted the hue component only. Jin and Shevell (1996) asked their observers to adjust hue and saturation of a test patch until it matched the remembered colour and in the study by Uchikawa (1983) observers judged the saturation of two colour stimuli, which were otherwise identical. As different as the matching paradigms are, they all limit the range of possible errors to one or two dimensions, respectively, and it is disputable whether general conclusions about colour memory can be drawn from such studies. Instead of using an adjustment task several studies (Epps & Kaya, 2004; Hamwi & Landis, 1955; Ling & Hurlbert, 2008; Perez-Carpinell et al., 1998a) made their observers select a colour out of a larger array of colour samples. Although this is a more natural approach of testing colour memory, because in everyday life we cannot adjust colours either, presenting a fixed array yields particular difficulties. To choose an appropriate array size is crucial to the outcome of an experiment; too many distractors may confuse observers and the visual inspection would take very long whereas too few distractors, for example only one, would make the task extremely easy (assuming that they are easily distinguishable). However, the difficulty of a selection task depends not only on the array size put also on the perceived colour differences of the distractors. When they are perceptually too close together then it is almost impossible to select the matching colour correctly and, on the other hand, when they are perceptually too different the task becomes extremely easy. The distractors can also vary in all three attributes or only in one or two. Hamwi and

Landis (1955) showed rather large arrays for selection (672 and 168 colours) but did not provide detailed information about the perceptual colour differences between distractors and test colours. Therefore, the observed ceiling effect might have been also due to a, in perceptual terms, very easy array. In Ling and Hurlbert's (2008) study the array consisted of 16 colour samples that varied in one attribute only, either in saturation or hue. As mentioned already above, it is not possible to make a clear statement about colour memory from such a limited array. Conclusions can only be drawn regarding saturation and hue memory.

In the studies by Perez-Carpinel et al. (1998a) and Epps and Kaya (2004) selection was performed out of an array of 20 and ten Munsell colour chips, respectively, which varied in hue, chroma and value. That allowed over- and underestimating the chroma and value of the original colour as well as mismatching the hue. Despite a balanced methodology and large numbers of observers the authors of these two studies were unable to draw clear conclusion about colour memory.

There is a further difficulty in using a matching by selection paradigm. Because of the limited number of available colours none of these might perfectly match the colour an observer has in mind. Hence, he or she must come to a compromise between two colours. In other words, not a perfect match is achieved but the best possible.

Contrasting adjustment and selection matching paradigms it becomes clear that there is no perfect method to study colour memory. The adjustment of hue, saturation and lightness potentially allows the perfect reproduction of a remembered colour, nevertheless this method is highly artificial. Matching by selection is a much more natural task, however the array of colours from which the selection takes place must thoughtfully be chosen.

Table 4.3.1-1 provides an overview of studies regarding colour memory over time and their key features.

4.4 Memory colours for familiar objects

In the studies mentioned above researchers focused on the deterioration of colour memory over time. Abstract stimuli were used and the emphasis in these studies was placed on the physical characteristics of colour. Only Jin and Shevell (1996), and Francis and Irwin (1998) presented their stimuli both in isolation and in context. The context used by Jin and Shevell was abstract and meaningless and could not be related to prior experience. In contrast, Francis and Irwin used abstract as well as meaningful contexts and showed that colour memory was worse in the absence of context. In this section the studies focus mainly on the stability of colour memory for situations that require prior experience with coloured object and/or colours and words. Table 4.4-1 provides an overview of selected studies on memory colours for familiar objects.

The accuracy with which colours of familiar objects are remembered (memory colour) was determined in a study by Bartleson (1960). Observers were presented with a series of 931 Munsell colour chips and when the examiner named an object, observers selected a chip that best represented the colour of this object. Ten familiar objects were named in total: red brick, green grass, dry grass, blue sky, flesh, tanned flesh, broad-leafed summer foliage, evergreen trees, inland soil and beach sand. The group of 50 observers was divided into three groups, one consisting of completely inexperienced, the second of slightly experienced and the third of expert observers regarding colour (29, 17 and 4 observers, respectively). Bartleson found that the memory colours of all ten objects were recalled significantly different from the object's original colour. Most object colours were remembered as more saturated and lighter than the original one.

Summary table MEM1 here – Table 4.3.1-1

Although Bartleson found that the variations were not systematic she concluded that the remembered colours tended to be more characteristic of the dominant chromatic attribute of the object, i.e. grass was remembered more green, the sky more blue and so on.

The variety of familiar objects from which researchers can chose to study memory for object colours is almost infinite. Perez-Carpinell et al. (1998b) used eight familiar objects (aubergine, green watermelon, green lettuce, yellow lemon, orange, pink rose, brown chestnut and red tomato) to investigate whether there is a difference in memory for coloured objects between observers with and without intensive prior experience/knowledge in colour (each observer group consisted of 50 subjects). Furthermore, the authors studied whether the chromaticity of the illuminant influences memory performance. For a task the examiner named one of the eight objects and observers picked the colour sample that represented best the object's typical colour from an array they were presented with. An array consisted of ten NCS[1] paper samples, which surrounded the mean colour of the named object. The task was completed by each observer once under daylight and once under tungsten light. Perez-Carpinell and colleagues did not find significant differences in memory between either experienced versus inexperienced observers, or daylight versus tungsten light. However, they found that the accuracy with which object colours were remembered was object dependent. Most accurately remembered were the colours of aubergine and green lettuce whereas yellow lemon and orange were remembered worst. With respect to saturation the authors observed that relatively saturated and poorly saturated object colours tented to be overestimated and underestimated, respectively.

[1] Natural Colour System; this colour system is explained in detail in section 3.4.

Summary table MEM1 here – Table 4.4-1

The procedures used in the studies by Bartleson (1960) and Perez-Carpinell et al. (1998b) are very similar. Objects were named and observers picked a colour out of an array of colour samples, which they thought was the typical colour of the named object. The arrays that Perez-Carpinell and colleagues used consisted of ten colours however, no detailed information about these colours was provided. It remains unclear which criterion the authors applied creating the arrays. It is not known whether the ten colour samples were perceptually very similar or whether they varied in one attribute only, or in all three. Therefore, it is not possible to judge the difficulty of the arrays and how they might have influenced the results. It was reported that matching accuracy varied considerably between objects. However, this might have been due to perceptually inconsistent arrays rather than to actual differences in the memorability of object colours. Bartleson, in contrast, allowed the observers to pick from a not preselected array of 931 colour samples, which differed in hue and chroma from each other. The inspection of such a large array might have taken longer but the method was less restrictive than that applied by Perez-Carpinell and colleagues.

In both studies groups of colour experienced and inexperienced observers participated. While Perez-Carpinell et al. stated that there was no difference in performance between these two groups of observers this issue was not further elaborated by Bartleson.

The thought that the accuracy of memory colours might depend on prior education of the observer in colour was also pursued by Siple and Springer (1983). They had two groups of observers, one group was familiar with colour reproduction (six observers) and the other one (twelve observers) had no specific knowledge about colour. Siple and Springer assessed colour memory for six different fruits and vegetables and for three different conditions, which provided different levels of visual cues to the objects' identity. In the first condition a circular test patch was shown, which was surrounded by a selection of different fruits and vegetables. In the second condition the same circular test patch was shown but with texture, and in the third condition the circular test patch was replaced by a test patch in the shape of a fruit or a vegetable (without texture). In the first experiment a fruit or a vegetable was named and observers then adjusted the colour of the test patch (in all three conditions) until its appearance matched the colour the observer had in mind (memory instruction). The task of the second experiment was only slightly different. Now observers were asked to match the appearance of the test patch with the colour the fruit or vegetable should have according to observers' opinion (preference instruction).

Siple and Springer did not find a difference between the memory and the preferred object colour. The accuracy with which the colour of the fruits and vegetables was remembered varied considerably between the objects. The authors reported that hue and lightness were remembered generally quite accurately and that saturation tended to be higher than in the original object colour. However, the level of visual cues to objects' identity had no effect on matching accuracy. The hypothesis that colour experienced observers perform better, i.e. more accurately, in such experiments could not be confirmed; this finding is in agreement with that of Perez-Carpinell et al. (1998b).

Olkkonen et al. (2008) also studied the effect of shape and texture on colour appearance of eight familiar objects (fruits and vegetables). The 22 observers that participated in this study were presented with stimuli that had different levels of visual cues to the objects' identity; photographs, painted fruit, outline shapes, a disc with noise texture and a uniform disc were used as stimuli. Observers' task was to set the colour of a presented stimulus either to grey (perform an achromatic setting) or to the fruit's typical colour. Note, that the luminance distribution of the stimuli was kept fixed during adjustment. In total this procedure was repeated under five simulated illuminations. Figure 4.4-1 shows the eight test stimuli and two test field conditions.

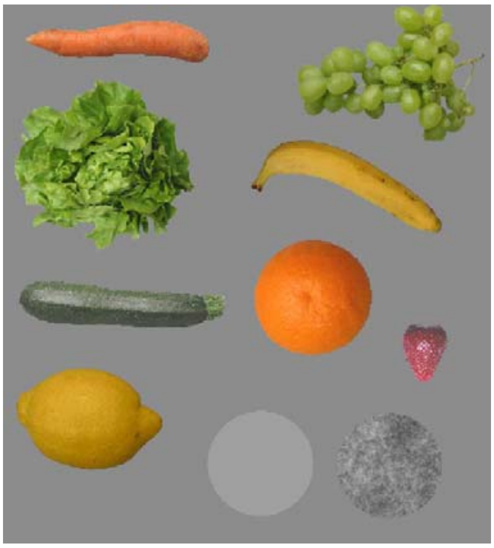


Figure 4.4-1. The eight fruit and vegetables used by Olkkonen et al. (2008). The figure includes also the uniform disc and disc with noise texture test field condition.

Olkkonen and colleagues found that the colour appearance of natural objects depends not only on their actual reflectance but also on the colour they are remembered with i.e. their typical colour. The more visual cues to an objects' identity are available the stronger is this effect. In other words, the adjustment of realistic photographs is more affected by the typical colour of a banana, for example, than the adjustment of an outline shape. This effect is independent of the illumination.

Despite similar experiments Siple and Springer and Olkkonen et al. came to opposite conclusions. Siple and Springer found no effect of integrated shape and texture cues on the variability of colour settings and suggested that memory colours are independent from form and texture information. However, Olkkonen and colleagues suggested the more form and texture information is available the more influenced are memory colours for familiar objects. One of the key differences of these studies was that Olkkonen et al. presented their stimuli in isolation whereas the test field in the study by Siple and Springer was surrounded by images of other fruits. It appears that the difference in stimuli-context led to fundamentally different outcomes.

4.5 Relation between colour terms and colour memory

The studies cited in the section above (4.4) assumed prior experience with the test stimuli in the sense that observers needed to know the objects and their usual colour appearance. Another approach investigating colour memory is to assume prior experience with combinations of specific words and colours, i.e. to relate a colour term to a colour sensation retained in memory.

It has been established that there are terms in every language to describe basic colour sensations. All colours that are described by the same basic term are grouped into a colour category and the colour that best represents a colour category is called focal colour (more about colour categories in chapter 6). That conceptualisation facilitates the efficient use of colour terms in everyday life, for example for object description, as everyone has an internal representation of these basic colours.

A study that used specifically focal colours was reported by Heider (1972). She carried out a series of experiments, in which the memorability of focal colours was contrasted with non-focal colours. Twenty-nine Munsell colour chips were used as test colours and presented to each of the 41 observers individually. Observers were shown a colour chip for 5 sec, after a delay of 30 sec an array of 160 colour chips was placed into view and observers selected the chip that looked like the one he or she had memorised. Heider found that focal colours were remembered significantly more accurately than non-focal colours. However, Heider's methodology was criticised by Lucy and Shweder (1979), who suggested that her findings were due to a biased stimuli selection. Lucy's and Shweder's criticism was based on an experiment they had carried out with, what they claimed to be an unbiased stimulus selection, and had failed to find such a superior memorability of focal colours.

In a study by Ratner and McCarthy (1990) the influence of specific colours and context on colour memory was compared. The colours of their stimuli were classified into focal and non-focal and context was divided into appropriate and inappropriate. For example, an appropriate or meaningful context is a red stop sign whereas a green one would be inappropriate. In total, colour memory was tested in four conditions: focal-appropriate, focal-inappropriate, nonfocal-appropriate, and nonfocal-inappropriate. First observers were presented with a picture and asked to memorise the colour of an indicated area. After a 30 sec gap three pictures were put into view where the critical area varied slightly in brightness, but which were otherwise identical. Observers were instructed to select the picture that looked most like the one they had seen before. The authors found significantly better memory for appropriate than for inappropriate colours. Though, the results did not reveal a difference in memory for focal and non-focal colours.

Here matching was performed by selecting one image out of three. Similar matching paradigms have been employed in numerous colour memory studies (Epps & Kaya, 2004; Hamwi & Landis, 1955; Y. Ling & Hurlbert, 2008; Perez-Carpinell et al., 1998a). However, the number of distractors in these latter studies was considerably higher than two, reaching from ten to 672. It is questionable whether selecting one image out of three, which varied in one area in brightness only, was an demanding enough task, considering also that luminance contributes less efficiently to colour memory than chromaticity (Sachtler & Zaidi, 1992).

The theory that focal colours are remembered more accurately was also investigated by D’Ath and colleagues (2007). Their hypothesis was that the more confident a colour can be named the more accurate this colour can be remembered. In other words, focal colours can be named confidently, and therefore they should be remembered more accurately than other colours. D’Ath et al. investigated this hypothesis within the study already reported above. Six observers completed the task mentioned earlier and in addition they were asked to name the test colours with basic colour terms and to indicate the level of how well this name represented the colour (level of confidence). The expectations D’Ath and colleagues had, were not confirmed. Although several colours that were named confidently were also remembered with great accuracy (e.g. purple), the results revealed also that some colours were remembered very accurately despite a lack of confidence in the moment of naming (red and orange). Consequently, the authors were not able to draw clear conclusions regarding the superior memorability of focal colours.

The controversy whether focal colours take an outstanding position regarding colour memory is well demonstrated by these three studies (see table 4.5-1 for summary). Heider (1972) found a superior memorability for focal colours but her findings were challenged because of the experimental procedure. Ratner and McCarthy (1990) did not find such superiority, this might have been due to a problematic methodology as they used a very limiting matching paradigm. Finally, the third study cited in this context (D'Ath et al., 2007) revealed superiority for some basic colour terms but not for others. To summarise, there is no clear evidence for superior colour memory for focal colours.

4.6 General summary

Intensive research has been carried out over the last decades to study colour memory and to find an answer to the question of how good humans are in remembering colour. The effect of time on colour memory has been studied as well as the memory colour for familiar objects and the influence of basic colour terms and focal colours on colour memory. The methods that have been used are diverse and equally diverse are the results that have been reported. There is a common agreement in only one point: humans' ability to remember colours is imperfect; provided that the performed memory task is not too easy.

Summary table MEM1 here – Table 4.5-1

4.6.1 Measuring colour memory

It is generally accepted that colours shift with respect to the original one when recalled from memory, but so far no measure has been established to quantify colour memory shifts. Without such a measure it is extremely difficult to compare levels of colour memory reported by different studies or to determine by how much colour memory deteriorated because of longer delay periods or because of removing context. For example, is a monochromatic colour that deviates by 8 nm from the original more accurately remembered than a colour that varies from the original by one chroma unit in the Munsell system?

4.6.2 Shifts in colour memory

Under most circumstances the remembered colour differs from the original one and despite great effort no consistent pattern has been identified that would allow predicting the shift between original and remembered colour. Shifts have been reported to occur for all three attributes of colour, hue (e.g. Bartleson, 1960; Collins, 1932; D'Ath et al., 2007; Epps & Kaya, 2004; Hamwi & Landis, 1955; Loftus, 1977; Perez-Carpinell et al., 1998a; Seliger, 2001), saturation (e.g. Bartleson, 1960; Jin & Shevell, 1996; Ling & Hurlbert, 2008; Perez-Carpinell et al., 1998b; Siple & Springer, 1983; Uchikawa, 1983) and lightness (e.g. Newhall et al., 1957; Tarczali et al., 2006).

Lightness shifts are reported less frequently because most studies focus on the recall ability of hue and saturation but they have been observed to occur for both possible directions. In other words, colours have been remembered as lighter as well as darker, depending on the experimental conditions. Hue shifts have been reported by almost all studies in colour memory. Despite the large amount of available data researchers have not been able to develop a generally valid rule that would describe hue shifts. The reason for this is that hue shifts are highly unsystematic. In almost every study it is stated that some hues are remembered with greater accuracy than others. However, it is impossible to rank hues in order of memorability because the overall findings are inconsistent. The range of distinguishable hues is enormous and therefore it is very unlikely that identical hues are used in different studies by coincidence. Consequently, the yellow stimulus used in one study might differ considerably from the yellow in another study, for example. A majority of studies revealed saturation shifts towards higher saturation, i.e. colours are remembered as more saturated than the original. Opposite findings have been reported less frequently but they should not be ignored. Therefore, shifts in saturation have to be classified as unsystematic just like lightness and hue shifts.

Even though shifts in colour memory do not follow a consistent pattern, several factors affecting the stability of colour memory, such as time and context in which colours are memorised (see Tate & Springer, 1971), have been identified.

5 Colour constancy

5.1 Introduction

In everyday life, we refer to colour as a constant property of an object and we are unaware that most objects only reflect light. A red balloon looks red because it reflects mainly the long-wavelengths of the spectrum when illuminated by white light. When the same balloon is illuminated by a specific green light it appears black because no light will be reflected. This is a rather extreme example but it demonstrates clearly that objects' colour appearance depends likewise on the reflectance properties of the object and the light it is illuminated with. When the same object we have seen before is illuminate by a different light source than the spectral composition of the reflected light changes and should theoretically lead to a changed colour appearance of the object. However, we know from experience that a banana looks yellow no matter which light it is under, in other words the change in the reflected light stays unnoticed. The fact that object colours appear unchanged despite a change in illumination is known as colour constancy (e.g. Jameson & Hurvich, 1989; Kaiser & Boynton, 1996), and in order to recognise a surface as unchanged the visual system must access information about the illuminant and the surface reflectance separately.

In a simplified world the light that reaches our eyes, commonly referred to as colour signal, is the product of the spectral power distribution of the illuminant and the spectral reflectance function of the surface (figure 5.1-1). Physically the reflectance spectra can be determined by measuring first the colour signal (with a spectroradiometer) wavelength by wavelength and dividing it then by the spectral power distribution of the illuminant. However, in the human eye the colour signal is detected by the three classes of cones on the retina, which are unable to transmit full spectral information.

They rather provide the visual system with information about their individual photon catch. This information changes when the colour signal changes but with only this information, the visual system cannot differentiate whether the alteration is due to an illumination or surface change. If there is no possibility to access separate information about the illumination and the surface reflectance from the colour signal how does the visual system achieve colour constancy?

In the paragraph above a simplified world is assumed though, the real world is more complex. Usually objects are surrounded by other objects and there is often more than just one light source. Therefore, the light reflected from an observed object not only depends on the reflectance properties of the object's surface but also on the illumination, other objects, their location and position with respect to the illuminant (or illuminants) and each other. Thus, to be colour constant the visual system has to compensate for these various changes.

5.1.1 Discriminating illuminant and surface changes

Colour constancy is usually defined as constant surface colour appearance despite an illuminant change. Craven and Foster (1992) however, studied whether observers are actually able to differentiate between an appearance change due to either an illuminant change or a change in surface reflectance. They presented two Mondrian patterns, which comprised 49 rectangular patches each, successively on a monitor. Observers were told that the second array of patches could differ from the first one because the overall illumination had changed or because the second array was made of slightly differently coloured patches.

Craven and Foster found that observers discriminated reliably between these two kinds of changes, in other words, observers had no difficulties to distinguish between an illuminant and a surface change.

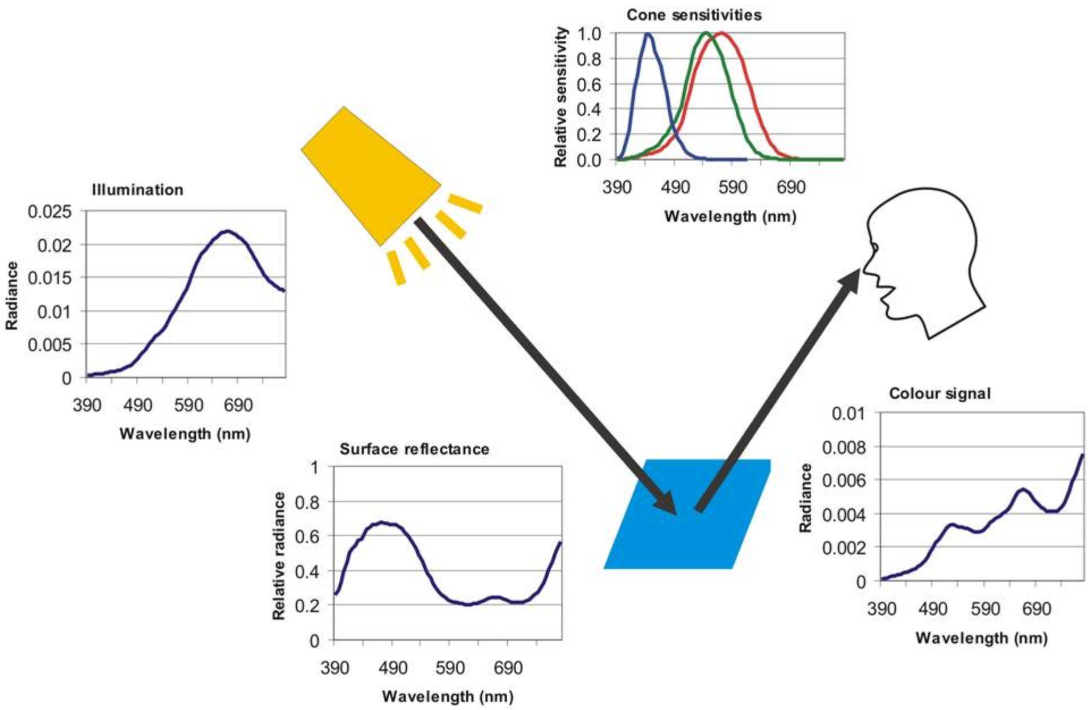


Figure 5.1-1. The emitted light hits a surface and is reflected; the reflected light reaches the eye of an observer who experiences colour. Each light source emits an illumination with a characteristic spectral power distribution. Together with the surface reflectance of an object, they create the colour signal, which reaches an observer's eye. The incident colour signal is absorbed by the cones, dependent on the cone sensitivities.

5.1.2 Why should the visual system be colour constant?

Natural daylight is not constant during a day but changes considerably from dawn until dusk. Much larger changes in illumination can even be experienced when coming from outside into a room with artificial illumination. However, the visual system compensates for these permanent changes and allows the stable perception of colour in our environment.

Colour is a reliable cue to object identification and it has also been shown to enhance scene recognition (Rinner & Gegenfurtner, 2000; Wichmann, Sharpe, & Gegenfurtner, 2002). As long as objects are under a constant illumination and context it is sufficient to simply remember the colour to recognise it as the same, though such situations are rare. It is much more common to encounter coloured objects in changed illumination and context conditions and to recognise colours they must firstly be remembered and secondly the illuminant and context change must be accounted for. Therefore, colour constancy can be considered as a sophisticated form of colour memory. In everyday life it is not always necessary to perceive colours as perfectly constant but to make confident judgements about colour categories. For example, the ability to decide whether a banana is yellow or green, and thus, is ripe or unripe.

5.1.3 Overview of the chapter

In the following section different methods of measuring colour constancy are discussed that are frequently used (section 5.2). Then mechanisms and cues contributing to colour constancy are reviewed in section 5.3. Section 5.4 starts with a brief overview of approaches towards colour constancy research and focuses then on experimental designs and stimuli used in this area of research. It follows a section (5.5), in which different methods are presented to quantify colour constancy. This chapter finishes with some final remarks regarding the topic (section 5.6).

5.2 Methods of measuring colour constancy

5.2.1 Matching to an internal standard

There are several methods to measure the level of observers' colour constancy. One approach is to test how well observers match a test surface within a scene in order to appear achromatic (e.g. Bäuml, 1999; Boyaci, Doerschner, & Maloney, 2004; Brainard, 1998; Doerschner, Boyaci, & Maloney, 2004; Fairchild & Lennie, 1992; Kraft & Brainard, 1999; Olkkonen et al., 2008; Rinner & Gegenfurtner, 2000; Yang & Maloney, 2001). In such a task observers adjust a specified surface to match their internal representation of white or any shade of grey. Speigle and Brainard (1999) showed that achromatic settings together with a gain-control model allow predicting colour constancy for chromatic stimuli rather well. However, it has been argued that achromatic settings only provide information about a single point in perceptual colour space, which cannot be generalised (Foster, 2003; Smithson, 2005).

In a modified version of this task observers match a test surface to their internal standard of a unique hue (red, green, blue and yellow); unique yellow appears neither reddish nor greenish, unique green appears neither bluish nor yellowish and so forth (e.g. Arend, 1993; Chichilnisky & Wandell, 1999; Smithson & Zaidi, 2004). Although this modified version allows measuring multiple colour loci it is limited to the four unique hues.

5.2.2 Asymmetric colour matching

Another approach to test colour constancy is how well observers match the colour of two surfaces seen under different illuminants (asymmetric colour matching). The two test surfaces, which can be real or computer-generated, are usually shown within an array of numerous surfaces. The observer can be presented with the two illuminant conditions at the same time, i.e. simultaneously (e.g. Arend & Reeves, 1986; Arend, Reeves, Schirillo, & Goldstein, 1991; Brainard, Brunt, & Speigle, 1997; de Fez, Capilla, Luque, & Perez-Carpinell, 2001; Foster, Amano, & Nascimento, 2001), or with one after the other, i.e. successively (e.g. Brainard & Wandell, 1992; de Fez et al., 2001; Foster et al., 2001). Using a successive asymmetric matching paradigm allows adapting to the changed illuminant conditions but relies on observers' ability to recall colours. On the other hand, simultaneous asymmetric matching does not require remembering colours but the level of adaptation depends on the time observers spend looking at one or the other surface. However, simultaneous asymmetric matching can also be performed under haploscopic viewing conditions (e.g. Lucassen & Walraven, 1996), where each eye is presented with a different illuminant condition. While this method allows controlling the state of adaptation it eliminates binocular cues to the scene geometry.

5.2.3 Colour naming

A third approach for measuring colour constancy discussed here is colour naming; observers assign colour terms to test patches under different illuminants (Granzier, Brenner, & Smeets, 2009b; Hansen, Walter, & Gegenfurtner, 2007; Kennard, Lawden, Morland, & Ruddock, 1995; Troost & de Weert, 1991; Uchikawa, Uchikawa, & Boynton, 1989). It is suggested that this method is the most direct approach to quantify colour appearance (Foster, 2003; Troost & de Weert, 1991) despite the weakness that only a very limited range of words are available to describe the approximately 2.28 million discernible colours (Pointer & Attridge, 1998). This limitation is often enforced by providing only a set of basic colour terms that have to be used to classify test patches. A colour naming procedure allows only for coarse classification, however, the precision of this method can be improved by rating (Schultz, Doerschner, & Maloney, 2006; Speigle & Brainard, 1996); observers indicate how well a colour term represents an actual test patch.

5.2.4 Instructions

The importance of instructions in the outcome of an experiment has been demonstrated by Arend and colleagues (Arend & Reeves, 1986; Arend et al., 1991). In an asymmetric matching task observers were presented with identical scenes but told to perform two different tasks. In one task observers were instructed to perform a hue and saturation match ('to match the hue and saturation of the test patch to those of the standard patch'), i.e. to perform a physical match between two surfaces. In the other task observers were told to match test and standard patch to 'look as if they were cut from the same piece of paper' (paper-match), i.e. observers were asked to perform an appearance match. While almost perfect colour constancy was revealed for the paper-match condition observers showed only little colour constancy for the hue-match conditions.

That observers are able to perform different tasks under identical experimental conditions was also demonstrated by Reeves et al. (Reeves, Amano, & Foster, 2008). The authors instructed their observers to perform a paper-match and a hue-match as described above but using a rating system instead of an adjustment task. Additionally observers performed a third task where they had to decide whether a material change had occurred. The authors showed that observers are capable of differentiating between physical and perceptual properties of the scene.

5.3 Mechanisms and cues contributing to colour constancy

It has been shown that colour constancy is not achieved by a single, but by the interaction of several, mechanisms and cues. It is the variety and combination of several cues that support colour constancy (e.g. Kraft & Brainard, 1999; Kraft, Maloney, & Brainard, 2002) and it has been suggested that the importance of particular cues varies depending on their availability and that the weighting of them is a dynamic process.

The following paragraphs provide a short overview of mechanisms and cues contributing to colour constancy.

5.3.1 Chromatic adaptation

Chromatic adaptation has been identified as a powerful mechanism contributing to colour constancy. Pioneering research in this area was conducted by von Kries, who postulated that the mechanisms of chromatic adaptation operate as gain controls on the cone signals. Since then an enormous number of studies have been dedicated to the investigation of chromatic adaptation (for review see Kaiser & Boynton, 1996; Webster, 1996). It was shown that von Kries' hypothesis can explain the phenomenon of colour constancy only partially (for review see Jameson & Hurvich, 1989) and that chromatic adaptation involves not a single but several mechanisms at low-level as well as high-level stages (e.g. Albright & Stoner, 2002; Zaidi, Spehar, & DeBonet, 1997).

Different mechanisms within chromatic adaptation have been identified to determine colour appearance. One of these is chromatic adaptation to the spatial average of a scene. This process occurs over a large spatial area and is rather slow, taking approximately 2 min to stabilise, i.e. approaching an asymptotic steady state at 90%-95% (Fairchild & Lennie, 1992; Fairchild & Reniff, 1995; Rinner & Gegenfurtner, 2000; Werner, Sharpe, & Zrenner, 2000). Another mechanism is chromatic adaptation to local colour contrast (Hurlbert & Wolf, 2004; Webster & Mollon, 1995). Colour contrast occurs through the interaction of adjacent surfaces. This process is almost instantaneous (Rinner & Gegenfurtner, 2000) and determines colour appearance radically (Webster, Webster, Malkoc, & Bilson, 2002).

The extent to which adaptation to the spatial mean of a scene and adaptation to local colour contrast mediate colour constancy was studied by Kraft and Brainard (1999) using an achromatic setting task. They also investigated the effect of adaptation to the most intense scene region, which has been suggested to be crucial for colour constancy (e.g. McCann, McKee, & Taylor, 1976). The setup used by Kraft and Brainard allowed isolating cues that activate these mechanisms and study of their effects on colour constancy separately. Under each condition observers showed a rather moderate level of colour constancy indicating that more than a single mechanism is necessary to achieve high levels of colour constancy.

5.3.2 Cone-excitation ratio

It has been suggested that not only the chromatic adaptation of cones are essential for colour constancy but also the ratios of cone-excitation. Foster and Nascimento (1994) showed that cone-excitation ratios stay almost invariant under illuminant changes and that this is a powerful cue to detect such changes. As mentioned above, Craven and Foster (1992) had found that observers can reliably discriminate between an illuminant change and a surface change, and Foster and colleagues argue that this is due to the stability of cone-excitation ratios.

Almost perfect invariance of cone-excitation ratios has been reported for a wide choice of typical illuminants[1] and scenes (artificial as well as natural scenes (Nascimento, Ferreira, & Foster, 2002)). The variance of cone-excitation ratios was subject of a study by Nascimento and Foster (1997), who investigated observers' sensitivity to these invariance violations. Observers had to discriminate between a real illumination change and one that was manipulated so that the cone-excitation ratios were preserved accurately. The manipulated illuminant changes were identified erroneously as real illuminant changes despite being highly unlikely to occur in the natural environment. The authors found that misidentification increased with increasing variance of the cone-excitation ratios. In other words, observers are sensitive to invariance violations and that a change in cone-excitation ratios implies a change in material.

5.3.3 Shadows

When light illuminates a scene, a series of interactions take place, which provide cues about the illuminant. Shadows are generally assumed to be such an illuminant cue (D'Zmura, 1992; D'Zmura & Iverson, 1993a, 1993b) as they provide information about the position and the number of light sources (Erens, Kappers, & Koenderink, 1993). The effect of cast shadows on observers' ability to estimate shapes (for example concave and convex spheres) has been investigated in several studies (e.g. Ramachandran, 1988). However, so far no study has been carried out investigating specifically the effect of shadows on colour constancy.

[1] Illuminants that lie along the blackbody locus (see section 3.2.2).

5.3.4 Specular highlights

Another cue to the illuminant are specular highlights. Specular highlights arise from shiny surfaces when illuminated and provide valuable information about the position of the light source or sources (figure 5.3.4-1). They also provide details about the chromaticity of the light source (D'Zmura & Lennie, 1986; Lee, 1986; Yang & Maloney, 2001; Yang & Shevell, 2003) as the spectral distribution of the specular highlights (arising from numerous surfaces such as plant leaves, painted surfaces, plastic surfaces, etc.) is very similar to that of the light source. However, exceptions to this are surfaces of homogeneous materials, for instance, polished metal.

Figure 5.3.4-1. There are two specular highlights on the black sphere, created by two white light sources which are in front and to the right of the sphere.

5.3.5 Mutual illumination

When light hits a scene it is reflected between surfaces and gives rise to mutual illumination, which provides information about the surface reflectance of the objects involved, chromaticity of the local incident light (Funt & Drew, 1993) and the scene geometry (Bloj, Kersten, & Hurlbert, 1999; Nayar, Ikeuchi, & Kanade, 1991) (figure 5.3.5-1). Kraft and Brainard (1999) demonstrated the contribution of mutual illumination to colour constancy by presenting observers with a scene that did not contain other illuminant cues than mutual illumination. Furthermore, all other mechanisms that are known to support colour constancy were silenced. Despite the reduced conditions observers' showed colour constancy, although very little.

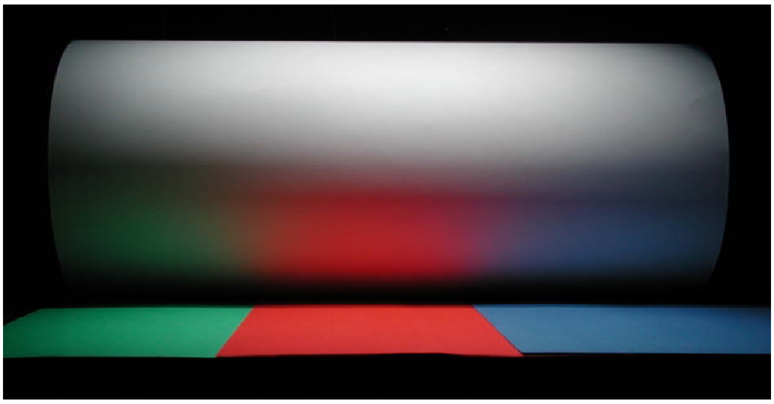


Figure 5.3.5-1. The scene is illuminated by a single light source directly above the setup. Light is reflected from the coloured papers onto the white cylinder introducing mutual illumination.

5.3.6 Cues to depth

Depth perception is a crucial part of human visual perception that allows judging distances and locating objects in the three-dimensional environment. When viewing a natural scene a number of cues provide detailed information about the spatial situation of the environment. Most of these cues such as size, interposition, perspective, shadows, shading and motion parallax are monocular depth cues. A binocular depth cue is retinal disparity, which is caused by the slightly different images projected into the eyes. However, there are also cues that are not directly related to scene/image properties such as convergence and accommodation (oculomotor cues to depth). Whenever an object is fixated the eyes are focused and converged by a certain amount that allows the visual system to estimate the distance between eyes and object, i.e. the state of convergence and accommodation provides useful information about absolute distances.

The influence of perceived depth on lightness and colour perception has been studied (e.g. Boyaci, Maloney, & Hersh, 2003; Gilchrist, 1977; Shevell & Miller, 1996) and it has also been discussed whether depth cues specifically affect colour constancy (Kraft et al., 2002; Werner, 2006; Yang & Shevell, 2002). For example, Yang and Shevell (2002) hypothesised that the absence of retinal disparity would lead to a misinterpretation of illuminant cues such as shadows and specular highlights and therefore affect colour constancy.

To investigate their hypothesis the authors presented rendered scenes, which were either created for stereo vision or identically for both eyes. The scenes comprised abstract stimuli with specular highlights and shadows that were consistent with a light source (figure 5.3.6-1, left). Observers performed colour constancy tasks in with and without retinal disparity conditions but note, that convergence and accommodation were identical for both viewing conditions. Yang and Shevell reported improved colour constancy for the with retinal disparity condition. They suggested that this improvement may have been due to a difference in segmentation of the three-dimensional scene depending on retinal disparity, and that such a difference may have influenced the illuminant estimation.

The effect of depth cues on colour constancy was also investigated in a study by Kraft et al. (Kraft et al., 2002). In their study observers viewed real surface stimuli (figure 5.3.6-1, right) either directly or through a telescopic viewing system (in both conditions the scene was viewed monocularly). The direct viewing condition provided the observers with depth cues such as motion parallax and accommodation (but not convergence), which were not available in the second viewing condition. The authors reported that observers’ colour constancy performance was similar for both viewing conditions. However, it remains arguable whether the comparison between these two viewing conditions was appropriate or whether binocular viewing would have led to a different outcome.

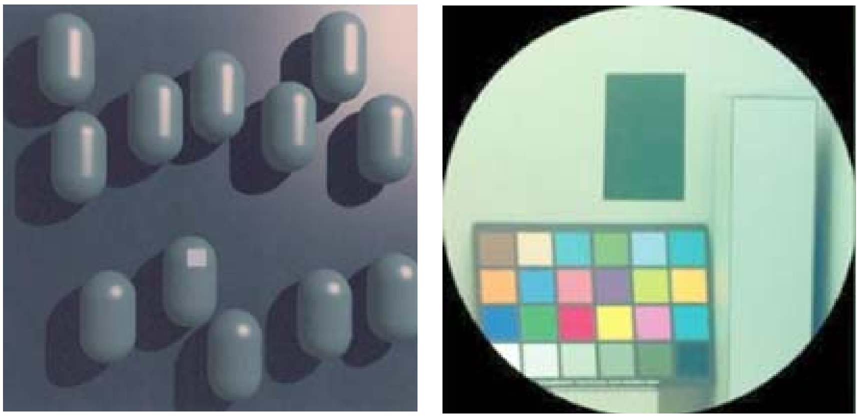


Figure 5.3.6-1. Experimental scenes used in the studies by Yang and Shevell (2002) (left) and Kraft and colleagues (2002) (right).

While Yang and Shevell found that presenting stimuli in 3D leads to improved colour constancy, Kraft et al. could not confirm this finding. This latter finding might have been due to the scene configuration, because the scene presented by Kraft and colleagues did not contain specular highlights and was poor in shadows. The removal of depth cues might not have affected observers' performance, if as Yang and Shevell suggest, the validity of specular highlights and shadows as illuminant cues decreases under monocular viewing conditions. The controversy surrounding whether cues to depth influence colour constancy remains. In a study by Werner (Werner, 2006) colour constancy was actually diminished when the test field and background were located in different depth planes. However, the only cue to depth in this study was retinal disparity.

5.4 Approaches towards colour constancy research

Colour constancy is an enormous field of research in which two major lines of investigation can be distinguished. One line focuses on studying human colour constancy performance and developing models that can explain the phenomenon. A second line of investigation concentrates on achieving colour constancy in computer vision systems. Numerous algorithms have been proposed that solve the problem of colour constancy to a certain extent (for review see Hurlbert, 1998; Maloney, 1999). Theoretically, these two lines can stand separately because models about human colour constancy can be formulated without drawing on computational solutions and vice versa. However, practically a complement of the two lines is desirable as, for example, a better understanding of the processes in the visual system may lead to improvements in algorithms.

Apart from the different methodological approaches, several ideas of how colour constancy is achieved by the visual system have been developed. One way to think about colour constancy is that the challenge for the visual system is to recover information about the illumination and the object reflectance in a scene from a single signal. If it is possible to estimate the illuminant correctly, then the visual system could accurately assess surface reflectance. In other words, it is assumed that if the visual system is able to extract the spectral reflectances of surfaces in a scene it achieves colour constancy. Therefore, it is essential to have a variety of visual cues in a scene regarding the illuminant. This approach is known as inverse-optics approach and has been considered in numerous models and algorithms (e.g. Brainard & Freeman, 1997; Maloney, 1999, 2002).

Another way to think about colour constancy is that cone-absorptions and subsequent neural signals from objects across differently coloured illuminants are compared, and that this comparison enables the visual system to be colour constant (e.g. Foster & Nascimento, 1994; Smithson & Zaidi, 2004; Zaidi, Spehar, & DeBonet, 1997). These approaches are fundamentally different but there is experimental evidence supporting both of them.

Experimental design

An important source of illuminant cues is three-dimensionality of the experimental scene. If illuminant cues are considered to be fundamental for the visual system to achieve colour constancy then experimental stimuli should be chosen accordingly. The variety of experimental designs and stimuli in colour constancy research is immense. The following paragraphs provide a short review of stimuli and setups used with an emphasis on (a) how dimensionality was incorporated into the experimental scenes and (b) studies using real 3D objects.

5.4.1 Two-dimensional abstract stimuli

Experiments have often been carried out with two-dimensional stimuli in two-dimensional environments, i.e. with computer-generated stimuli consisting of simple geometric forms, which were simulated as flat matte surfaces and presented under spatially uniform illumination on monitors (e.g. Arend & Reeves, 1986; Chichilnisky & Wandell, 1995; Bäuml, 1999; Jin & Shevell, 1996; Murray, Daugirdiene, Vaitkevicius, Kulikowski, & Stanikunas, 2006b). Many different layouts were created by varying the complexity of the surrounding area of a test patch. Nevertheless, as complex as such backgrounds may be, they can only provide a limited range of visual cues, significantly less than are usually found in a natural scene.

5.4.2 Two-dimensional images of natural (three-dimensional) scenes

Scenes that provide a much wider range of cues were used, for example, by Nascimento et al. (2002), Amano et al. (Amano, Foster, & Nascimento, 2006) and Foster et al. (Foster, Amano, & Nascimento, 2006), who presented images of natural scene such as rural and urban areas (figure 5.4.2-1). Although these images showed scenes with which observers were familiar and presented all characteristics of real objects they had one possible drawback, the images were solely 2D projections of 3D scenes. If cues to depth play a role in colour constancy then this images might have misled observers. Although monocular depth cues were preserved, the oculomotor depth cues arising from the images were deceptive and binocular depth cues were not available.

Figure 5.4.2-1. Examples of a rural (left) and an urban scene image (right) as used by Nascimento et al. (2002).

5.4.3 Rendered images viewed stereoscopically

A further step towards three-dimensionality is to present images stereoscopically. A number of studies have used this approach using computer-rendered complex scenes (e.g. Boyaci et al., 2004; Doerschner et al., 2004; Doerschner, Boyaci, & Maloney, 2007; Schultz et al., 2006; Yang & Maloney, 2001; Yang & Shevell, 2003). In the study by Boyaci et al. (2004), for example, the scene consisted of simple geometrical objects with different reflectance properties (matte, shiny and transparent) and was illuminated by a blue diffuse and a yellow point-like light source (figure 5.4.3-1). This scene provided a wide range of illuminant such as specular highlight, shadows and mutual illumination, and depth cues, e.g. interposition, perspective and retinal disparity.

However, there is a general dilemma regarding the presentation of 3D stimuli/scenes on monitors and the cues to depth because of cue conflict. While monocular and binocular (if integrated) cues suggest that the stimuli are three-dimensional the oculomotor cues (accommodation and convergence) indicate otherwise. The visual system is therefore provided with slightly ambiguous information about the viewed scene.

Figure 5.4.3-1. The rendered and stereoscopically viewed scene used in the study by Boyaci et al. (2004).

There is a considerable difference between the displayed images of natural 3D scenes mentioned above and the scenes here, which is the naturalness regarding the content of the scene. While observers can identify easily natural 3D scenes, a scene containing a number of geometrical volumes must be experienced as highly artificial (especially as some objects in the scene shown by Boyaci et al. seemed to float in space). However, the naturalness of the scene content might not be important for the task observers have to perform.

5.4.4 Real three-dimensional objects

Experiments using real 3D objects are rare in colour constancy research. However, Brainard (1998) introduced an experimental setup that consisted of real surfaces and objects. Observers could see the walls, floor and ceiling of the experimental room, two objects (a white table and a brown metal bookcase) and the test patch, which was a grey Munsell paper mounted on the back wall of the room.

Another real-world 3D setup was used by Kraft and Brainard (1999). Their setup included a chamber with several geometric volumes, a tin foil covered tube and an array of different coloured papers (figure 5.4.4-1 A). All objects could be removed from the scene. The back wall was replaceable and the scene illumination was computer-controlled.

Real objects and surfaces were also used by de Almeida and colleagues (de Almeida, Fiadeiro, & Nascimento, 2004) who presented two almost identical scenes concurrently to their observers. All objects and surfaces in the left and right scene were identical, except for a cube in the middle of each scene. In the left scene the 3D cube was made of real paper representing a test colour, whereas in the right scene it was a virtual image of a cube whose colour appearance could be manipulated via a computer.

Nascimento et al. (Nascimento, de Almeida, Fiadeiro, & Foster, 2005) used also a virtual cube (which was not identified as virtual by observers) as test object and presented it in scenes with varying complexity. The complex scene comprised numerous objects of different shapes that differed in colour, and in the poor-complexity-condition the cube was presented in isolation on a uniform background.

Ling and Hurlbert (2006) studied the effect of colour memory on colour constancy using a 3D dome and circular 2D colour patches. The 3D dome was either presented in isolation on a uniform background or surrounded by five 2D patches (figure 5.4.4-1 C).

A 3D setup was also used in a study by Zaidi and Bostic (2008) who investigated object identification across illumination changes. Four real objects (cylinders), of which three had the same reflectance properties, were shown on a variegated background (figure 5.4.4-1 D). Two cylinders were always viewed under one illuminant and the other two under a different one.

Granzier et al. (2009b) conducted an experiment in which coloured sheets of paper were presented in several natural indoor and outdoor environments (figure 5.4.4-1 B). Therefore, all objects and surfaces were real and viewed under most natural conditions.

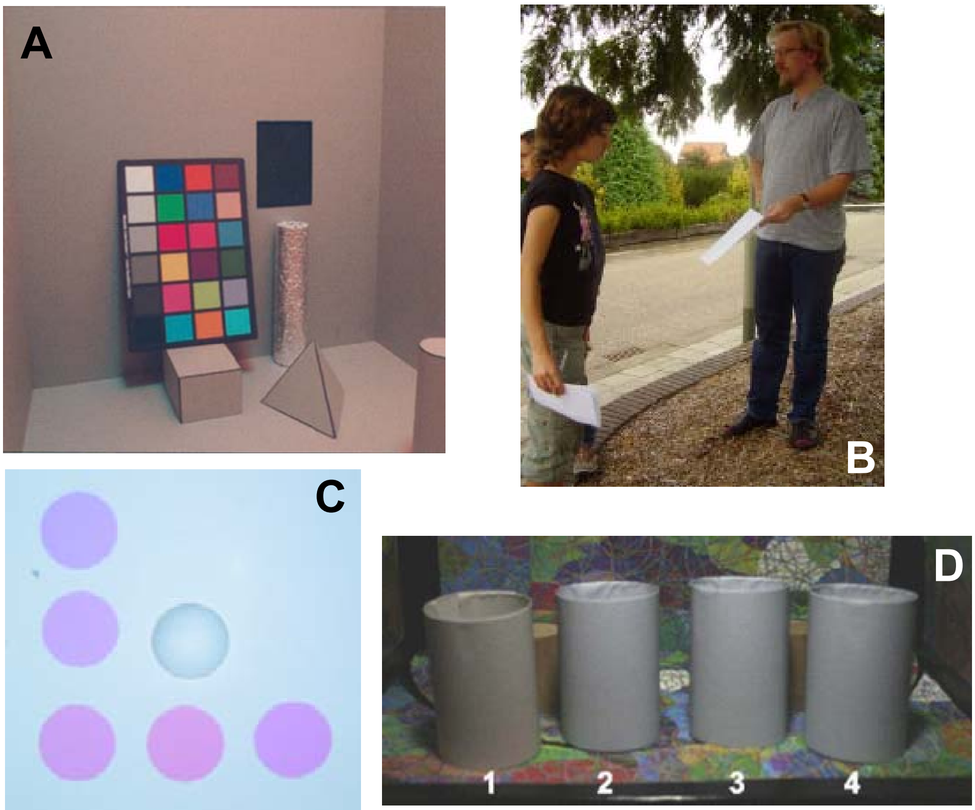


Figure 5.4.4-1. Experimental scenes used by (A) Kraft and Brainard (1999) (most complex configuration), (B) Granzier et al. (2009b) (outdoor environment), (C) Ling and Hurlbert (2006) (3D dome is in the centre of the scene) and (D) Zaidi and Bostic (2008) (all cylinders are under the same illuminant).

Although all the mentioned studies used real 3D stimuli the actual scenes varied considerably. As most studies had different objectives the experimental setups were designed to meet very specific standards. The studies by Brainard (1998) and Granzier et al. (2009b) aimed to investigate colour constancy under natural (or almost natural) conditions. Brainard presented observers with a scene that was not perfectly natural (in the sense that observers would not be able to encounter such a scene in everyday life) but included abstract stimuli such as 2D paper squares and well known objects such as a table. The scene was set up in experimental room in which the illumination condition was precisely controlled. On the other hand, Granzier and colleagues presented their stimuli in the most natural condition possible. By presenting coloured sheets of paper in everyday environments (six in total) they tested colour constancy in situations in which it usually occurs. Generally, both studies employed natural setups with which observers were familiar and which gave raise to a variety of illuminant cues.

Kraft and Brainard (1999) as well as Nascimento and colleagues (2005) used modifiable setups to study different mechanisms contributing to colour constancy and the influences of scene complexity, respectively. In complex viewing conditions these setups comprised several abstract stimuli such as squares, cubes and other geometrical volumes that provided numerous illuminant cues. However, in the study by Nascimento et al. observers were presented with the experimental scene for 1 sec under illuminant one and then for 1 sec under illuminant two, and judged whether the colour of the test object (virtual cube) changed accordingly with the illuminant change. The procedure did not incorporate adaptation which might be the reason why the authors did not find significant differences in observers' colour constancy performance between the high and low-complexity conditions. Kraft and Brainard allowed for adaptation but instead of a forced-choice procedure they employed an achromatic setting task. Variable scene complexity was used here to manipulate actively specific illuminant cues and the authors showed that colour constancy cannot be explained by a single mechanism or cue. (Not only the setups varied considerably between studies but also the actual task observers were asked to perform. The influence of different tasks is discussed below.)

Regarding scene complexity, the studies by de Almeida et al. (2004), Zaidi and Bostic (2008) and Ling and Hurlbert (2006) have in common that solely abstract stimuli were shown. The scenes used by de Almeida and colleagues, and Zaidi and Bostic provided several cues to the illuminant and observers were adapted to a mixed environment as scenes with different illuminations were viewed simultaneously. Ling and Hurlbert showed stimuli successively and allowed adaptation to a changed illumination condition however, because of the poor scene complexity only a very limited range of possible cues related to dimensionality arose from their scene.

5.4.5 Observers' tasks in experiments using real 3D objects

A variety of different matching tasks were employed in the studies using real 3D objects. It is unfeasible to determine the most appropriate matching paradigm to test colour constancy but they can be classified regarding their naturalness. In everyday life it must be decided whether a present object matches one in memory and objects' appearance cannot (can rarely) be adjusted until they match memory.

In Brainard (1998) and Kraft and Brainard (1999) the test patch was grey piece of paper, whose appearance was controlled by illumination, which was independent of the scene illumination. Observers were instructed to adjust the test patch until it appeared achromatic. Although the appearance of the test patch was entirely manipulated by illumination observers experienced its changed colour appearance as a result of a surface change. Instead of an achromatic matching task de Almeida et al. (2004) asked their observers to perform a chromatic setting in which a virtual cube had to be adjusted to match a reference cube so that they would appear to be made from the same piece of paper (the cubes were illuminated differently).

Matching by adjustment is a rather artificial procedure. Although it was ensured that appearance changes were experienced as being due to surface changes, it must have been rather unnatural to perform this task, as surface appearance can only rarely be adjusted continuously in the real world.

The primary objective of the study by Granzier et al. (2009b) was to assess colour constancy under natural conditions and that included also the matching paradigm. Coloured sheets of paper were presented in different environments and observers assigned them to colour terms. Observers were even allowed to move the sheets of paper and change their orientation.

A matching by selection paradigm was used in the studies by Nascimento et al. (2005) and Zaidi and Bostic (2008). Zaidi and Bostic presented four cylinders of which three had the same reflectance properties. Two cylinders were viewed under illuminant 1, two under illuminant 2 and observers had to identify the cylinder that had different reflectance properties. A similar task was performed in the study by Nascimento et al., where observers had to decide whether an appearance change of the test object (cube) was exclusively due to an illumination change or an actual surface change. Both these matching by selection tasks enforced a decision and mimic therefore an everyday situation quite well.

Ling and Hurlbert (2006) also used a matching by selection paradigm. A test colour was always presented on a 3D dome in isolation. After memorising a test colour, observers selected the matching colour patch from a selection of 2D patches. Throughout the experiment, the changes in appearance of the dome and the patches did not arise from an actual surface reflectance change but were generated by computer-controlled lighting illuminating white surfaces. The authors reported that the used paradigm was not ideal because the sudden apparent change of the surface colour of the dome and the patches was interpreted by the observers as artificial and not as a real surface change. Therefore, the observers might have made an appearance match instead of a surface match.

To summarise, the matching tasks described here differ considerably regarding their naturalness. The application of an adjustment task is appropriate to study exactly to what degree a visual system is colour constant. However, this might not be relevant for studying colour constancy how it is experienced in the real world and matching by selection or colour naming tasks imitate therefore everyday situations much better.

5.5 Quantifying colour constancy

It is common to express the level of colour constancy by a colour constancy index and different ones have been developed to satisfy specific needs. Numerous studies have applied the Brunswik ratio (BR) (Brunswik, 1928) or a modified version of it to compute a colour constancy index (e.g. Amano et al., 2006; Arend, 1993; Arend et al., 1991; Bäuml, 1999; Brainard et al., 1997; Cornelissen & Brenner, 1995; Daugirdiene, Murray, Vaitkevicius, & Kulikowski, 2006; Foster et al., 2001; Kulikowski & Vaitkevicius, 1997; Lucassen & Walraven, 1996; Murray, Daugirdiene, Stanikunas, Vaitkevicius, & Kulikowski, 2006a; Nascimento et al., 2005; Schultz et al., 2006; Troost & de Weert, 1991). The BR ratio is commonly defined as:

$$BR = 1 - \frac{PerceptualShift}{PhysicalShift} \qquad (6)$$

As the equation shows this ratio considers the perceptual shift of surface appearance as well as the physical shift that occurs due to an illuminant change. If a visual system is not colour constant at all, the perceptual shift is therefore identical to the physical shift (BR = 0). In this case, an observer would perform a chromaticity match. If a visual system is perfectly colour constant, it will compensate perfectly for the illuminant shift (BR = 1). There would be no perceptual shift and the observers would have performed a surface match.

This kind of index is usually two-dimensional as it considers the chromatic dimensions and ignores the luminance dimension, and colours are therefore commonly defined as

CIE u'v' chromaticity coordinates[1]. This allows comparing the Euclidean distances between observers' matches and (theoretically) perfect matches.

Instead of using CIE u'v' chromaticity coordinates Yang and Shevell (2002) proposed a colour constancy index where matches are specified in the MacLeod-Boynton colour space[2]. This index, and variations of it, differs from the previous one as it does not require measuring distances in an arbitrary colour space; such indices have been computed, for example, in studies by Smithson and Zaidi (2004) and Hansen et al. (2007).

Both kinds of indices just described have in common that they do not consider colour memory. However, if colour constancy is seen as a sophisticated form of colour memory then observers' ability to recall colours should be incorporated in a colour constancy index. Under laboratory conditions experiments can be designed to minimise the need of colour memory as, for example, in an achromatic matching task. When colour constancy is tested under real-world conditions, a present colour is compared to one in memory. To perform such a task successfully observers rely strongly on their colour memory. Ling and Hurlbert (2008) proposed a colour constancy index that includes the physical shift of a test stimulus as well as colour memory.

[1] The CIE 1976 uniform-chromaticity-scale (UCS) diagram is defined by the dimensions u' and v' and is approximately perceptually uniform. Note that all colours lie in a plane of constant luminance (Wyszecki & Stiles, 2000).

[2] The MacLeod-Boynton colour space is based on cone excitation (see MacLeod & Boynton, 1979).

Their index (CCI) is defined as:

$$CCI = 1 - \frac{\frac{\left(\overrightarrow{S_c} - \overrightarrow{S_m}\right) \cdot \left(-\overrightarrow{S_p}\right)}{\left\|\overrightarrow{S_p}\right\|}}{\left\|\overrightarrow{S_p}\right\|} \tag{7}$$

where $\overrightarrow{S_p}$ characterises the physical shift a stimulus undergoes by an illuminant change, $\overrightarrow{S_c}$ the constancy shift and $\overrightarrow{S_m}$ the pure memory shift. If the constancy shift equals the memory shift the index is one, indicating perfect colour constancy. The index drops to zero, indicating a complete lack of colour constancy, if an observer either shows perfect colour memory ($\overrightarrow{S_m}$ = 0) but performs a chromaticity match for the constancy task ($\overrightarrow{S_c}$ = -$\overrightarrow{S_p}$) or performs perfectly for the constancy task ($\overrightarrow{S_c}$ = 0) but the memory shift complies exactly the physical shift ($\overrightarrow{S_m}$ = $\overrightarrow{S_p}$).

A very different approach to quantify colour constancy was chosen by Granzier et al. (2009b). In their study observers assigned colour terms to paper samples and the level of colour constancy was simply measured by the percentage of correctly assigned terms.

To summarise, because of considerable methodological differences between studies a variety of different indices have been used to quantify colour constancy. Indices have been adapted to serve well under specific conditions but are not generally applicable. Quantifying colour constancy with a single number facilitates the comparison between studies, though the original data these indices are based on should be kept clearly in mind.

5.6 Final remark

Intensive research has been carried out over recent decades studying colour constancy. It has been demonstrated that there is no specific colour constancy mechanism but that colour constancy is achieved by the interaction of numerous mechanisms located in the retina and the cortex. While processes in the early stages of the visual system have been explored in-depth, those taking place in the cortical stages are less well understood (for review see Gegenfurtner, 2003). No evidence has been found of a specific colour constancy area in the cortex, although V4 has been identified to play a special role in colour constancy (e.g. Zeki, 1983a, 1983b). The absence of such a highly specified area seems rational as several studies have shown that colour constancy is influenced by other factors, e.g. memory and categorical colour judgements.

To conclude, immense knowledge about colour constancy and its underlying mechanisms has been acquired, and despite great efforts it remains unclear how and where exactly colour constancy is achieved in the visual system.

6 Colour categories

6.1 Introduction

Humans are able to perceive and distinguish approximately 2.28 million colours (Pointer & Attridge, 1998). In order to reason and communicate efficiently about colours we group them into categories. By definition, colour categories are described by observers with the same monolexemic terms and do not subsume within the range of any other colour word (Kay, Berlin, & Merrifield, 1991).

However, categorisation is not a colour specific phenomenon, in fact all our experience and knowledge is organised as concepts and categories in memory, because this is most economical for storing and retrieving data. By dividing the world into classes of things we decrease the amount of information we need to perceive, learn and recognise. Concepts and categories are used constantly during conversations as they allow conveying information about the world as we understand it and ourselves.

6.2 Origin of colour categorisation – nature versus nurture

There is a controversy taking place whether colour categories are universal or culture-specific, in other words, whether these categories are a product of nature or nurture (Davidoff, Goldstein, & Roberson, 2009; Franklin, Wright, & Davies, 2009; Kay & Kuehni, 2008; Kay & Regier, 2006; Roberson & Hanley, 2007). Universalists argue that colour categories are arranged around universally-shared prototypes in colour space, which are called focal points or colours (Berlin & Kay, 1969; Boynton & Olson, 1987; Heider, 1972; Kay & Regier, 2006; Regier, Kay, & Cook, 2005). The group that supports the idea of colour categories being culture-specific is often referred to as relativists. They argue that colour naming is the result of arbitrary linguistic conventions, and that differences in colour cognition are caused by cross-language differences in colour naming (Franklin & Davies, 2004; Roberson, Davidoff, Davies, & Shapiro, 2005; Roberson & Hanley, 2007; Roberson, Shapiro, Davidoff, & Davies, 2004). In order to find evidence that would support one or the other theory about the origin of colour categories, a cross-species study has been carried out (Fagot, Goldstein, Davidoff, & Pickering, 2006). Results from a matching task that was performed by baboons and humans revealed fundamental differences between the two test species. No evidence was found that colour categorisation occurs in baboon colour perception.

6.3 Basic colour terms

Cross-cultural studies have shown that the number of colour terms that are used to describe the whole range of perceptible colours varies considerably between languages. Berlin and Kay (1969) identified eleven basic colour terms in English (figure 6.3-1), which can be subdivided into three groups: (1) achromatic colour terms (black, white and grey), (2) primary colour terms (red, green, blue and yellow) and (3) secondary colour terms (orange, brown, pink and purple).

6.4 Colour categories in other areas of vision research

Although the origin of colour categories is a vivid field of research, colour categorisation is also of interest in other areas. For example, the contribution or involvement of different cortical regions to categorical perception of colour has been studied (e.g. Drivonikou, Kay, Regier, Ivry, Gilbert, Franklin, & Davies, 2007; Fonteneau & Davidoff, 2007; Holmes, Franklin, Clifford, & Davies, 2009; Ikeda & Osaka, 2007) as well as the effect of colour deficient vision on colour categorisation (e.g. Bonnardel, 2006; Cole,

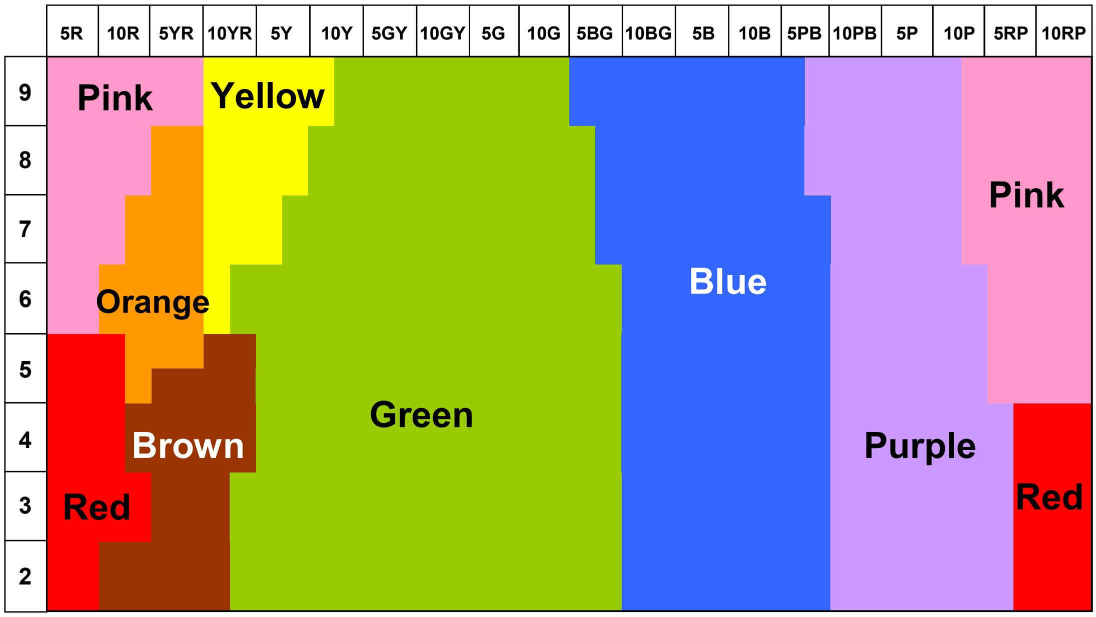


Figure 6.3-1. Colour categories in English. (Data redrawn from Roberson et al. (Roberson, Davies, & Davidoff, 2000). A set of 160 fully saturated Munsell colour chips was used to determine the colour categories. The array consisted of 20 evenly spread hues (levels 5 and 10) and eight levels of lightness (value 2 – 9).

Lakkis, Lian, & Sharpe, 2006; Kennard et al., 1995; Pokorny, Lutze, Cao, & Zele, 2006). Numerous studies have also been carried out to investigate the effect of colour categories on colour perception. For example, Boynton and Olson (1987) conducted a study to determine the exact locations of the eleven basic colours within the OSA colour system[1]. In a later study the same authors investigated the salience of chromatic basic colour terms (Boynton & Olson, 1990). Nine observers named 424 OSA colour samples, which were presented in isolation on a grey background, with monolexemic colour terms. The authors used response time, inter and intra-observer naming consistency to analyse observers' naming behaviour. Boynton and Olson (1990) found that colours that were assigned to a basic colour term were named with greater consistency than colours, which were assigned to non-basic colour terms. The

[1] This approximately uniform colour system was introduced by the Optical Society of America's Committee on Uniform Color Scales. For further information see David L MacAdam, 1974, 1978; Nickerson, 1981.

superiority of basic colours was also revealed by the response time analysis; basic colours were named considerably faster than non-basic ones.

The procedure of naming OSA colour samples was also used by Uchikawa et al. (Uchikawa, Uchikawa, & Boynton, 1989). They investigated the effect of achromatic surrounds on categorical perception of colours. Colour samples were presented within several surround conditions; no surround, small and large surround, and spatial separation between colour sample and surround (figure 6.4-1). The authors reported only small effects of different surrounds on observers colour perception. However, most findings are based on results obtained from the three authors themselves. It is arguable whether the reported outcome of the study was biased by the authors' extensive knowledge about colour categories and the purpose of the study.

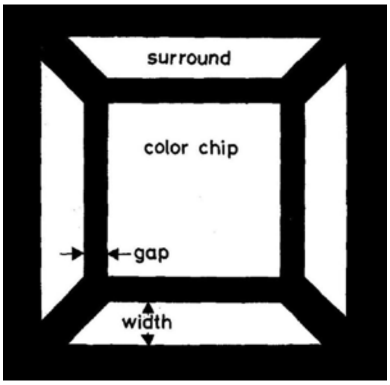


Figure 6.4-1. One of the surround conditions used by Uchikawa et al. (1989). Test sample and surround are not directly adjacent.

Kulikowski and Vaitkevicius (1997), Smithson and Zaidi (2004) and Hansen et al. (2007) studied the influence and the importance of colour categories on colour constancy. Colour categories were determined by colour naming in these three studies. Kulikowski and Vaitkevicius, who studied colour constancy under two illuminants found that observers' performance was best for the focal colours (the best example) of the red, green, yellow and blue categories. They suggest that this might be due to physiological properties of the visual system as red/green and blue/yellow are the basic opponent colour pairs.

Smithson and Zaidi determined the boundaries between the red and green, and the blue and yellow categories by classifying the appearance of a presented stimulus as either red or green or as either blue or yellow, respectively. The stimuli were shown in two illumination conditions and on different backgrounds, which were purposely biased. The authors reported that observers' colour constancy performance was largely unaffected by illumination conditions and different backgrounds.

In the study by Hansen et al. more than 400 different colours were assigned to eight colour categories; the German colour terms for red, orange, yellow, green, turquoise, blue, purple and grey were used (figure 6.4-2). This task was repeated under five illuminations and for various context conditions. Hansen and colleagues reported that boundaries between colour categories were rather constant apart from some minor but systematic deviations; these deviations were illumination and also context dependent.

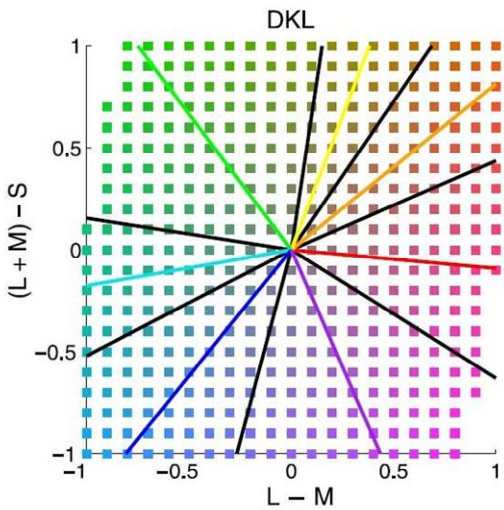


Figure 6.4-2. Location of colour categories in the DKL colour space[1] as determined by Hansen et al. (2007). Boundaries between categories are indicated by the black lines and the centres by the respectively coloured lines.

---

[1] Colour space proposed by Derrington, Krauskopf and Lennie (Derrington et al., 1984); based on physiological properties of the visual system.

A very different technique was used by Smithson et al. (Smithson, Khan, Sharpe, & Stockman, 2006) to study transitions between colour categories. A reverse Stroop task was employed to investigate the boundaries of several colour pairs. Observers were instructed to concentrate on a presented colour term and to ignore the colour in which it was printed as they had to respond to the meaning of the colour term only. Boundaries between colour categories were also determined by using a hue scaling task. Similar results were obtained for both techniques and the authors suggested that a reverse Stroop task is appropriate for the study of colour categories.

6.4.1 Colour memory within and across colour categories

The influence of colour categories on colour memory has been investigated in numerous studies (e.g. Bornstein, 1976; Boynton et al., 1989; Fonteneau & Davidoff, 2007; Ikeda & Osaka, 2007; Uchikawa & Shinoda, 1996). Despite very different methodological approaches the findings of these studies are in agreement, colours crossing a boundary of two categories are easier to distinguish or to remember than colours within a category. For example, Boynton et al. (1989) presented two stimuli consecutively with a delay of 10 sec and observers judged whether the stimuli were the same or different. A similar task was performed in a study by Uchikawa and Shinoda (1996) and in both studies it was found that observers gave the correct answer more frequently when the stimuli belonged to two different categories. Fonteneau and Davidoff (2007) studied neural correlates of colour categories. The observers' task was to detect specific colour features embedded within blocks of successively presented colour patches, and the authors reported shorter latencies for detecting colour features across categories than within a single category.

As an explanation of this effect it is suggested that the storage of colour in memory involves linguistic coding. Therefore, colours that cross the borders of colour categories might be easily verbally coded by assigning them different colour terms, whereas such coding is much more difficult (if not impossible) for colours within the same colour category.

6.5 Summary

Although it remains controversial whether the origin of colour categories is due to nature or nurture they are of general interest in colour perception. It seems that colour categorisation is rather robust. That might be an indication that it is more important for human observers to remember or recognise the category of a certain colour rather than recalling the exact chromaticity and lightness values.

7 Cross-media studies in vision research

Research in vision has historically been carried out with real objects while no other options were available. However, during the last decades technological innovations have fundamentally changed the way in which research is carried out. The availability of high-performance computers and high-quality monitors allows researchers to display computer-generated stimuli on monitors. This practice is widespread because it has numerous benefits over using real objects. For example, once a stimulus is created it can easily be manipulated and experimental conditions are precisely controllable. That allows experimenters for example to show sequences of stimuli for very short periods, in the range of milliseconds, which would be unfeasible with real objects. However, the use of displayed stimuli raises questions about (a) the appearance of these stimuli in comparison with real stimuli/objects and (b) the comparability of findings resulting from such experiments and real-object experiments.

In cross-media studies computer-generated stimuli are usually presented on cathode ray tube (CRT) monitors and are compared to real objects. In the literature displayed and real stimuli are often referred to as softcopy and hardcopy, respectively.

Two major lines of investigation can be identified in cross-media studies. One investigates the appearance and colorimetric aspects of colour reproduction, i.e. is a technological approach to colour reproduction. The other focuses on the direct comparison of observers' performances for identical (as far as possible) real object and computer-generated scenes. The latter is of great interest in vision research because it addresses the issue whether computer-generated and displayed stimuli are valid substitutions for real objects.

In the first section of this chapter several studies are reviewed concerning colour appearance models and the influence of viewing conditions under which these models are tested. The second section focuses on the comparability of findings resulting from real-world and computer-simulated experiments.

7.1 Technological approach towards colour reproduction

Consumers are able to experience the quality of cross-media colour reproduction easily. Imagine scanning a photo and displaying it on a monitor. It is expected that the softcopy looks exactly like the original hardcopy, but this is not always the case. For this reason, the colour reproduction industry has a great interest in developing models that allow converting images from one medium into another without altering their appearance. Displayed and print images are usually compared simultaneously by holding the print next to a monitor, or from memory.

The key intention in cross-media colour reproduction is therefore to produce precisely matching colours. In colorimetry two perfectly matching colours have identical tristimulus values, which can be due to (a) identical spectral power distributions or (b) metamerism (see section 3.2.1). Because of spectral differences of the primary colorants used to produce hardcopy and softcopy colours, all cross-media colour matches can only be of metameric nature. Metamerism is, to a certain extent, observer dependent as spectral sensitivities of the visual system vary between humans. The effect of observer variability in metameric colour matches across different media has been investigated in several studies (e.g. Alfvin & Fairchild, 1997; Oicherman, Luo, Rigg, & Robertson, 2008; Pobboravsky, 1988; Pointer, Attridge, & Jacobson, 2002; Rich & Jalijali, 1995) and significant inter-observer variabilities were reported in all studies except in Pobboravsky (1988). In all these studies, single colours were presented in isolation and observers adjusted the appearance of a test field until it matched the original colour. This type of study is fundamental although it approaches the issue of colour reproduction on a very basic level. The colour reproduction industry is generally concerned with far more realistic applications, for example the quality of displayed versus printed images.

7.1.1 Testing colour appearance models under varying viewing conditions

Numerous studies have used hard and softcopies of complex scenes to test different colour appearance models (e.g., Berns & Choh, 1995; Braun & Fairchild, 1996; Braun, Fairchild, & Alessi, 1996; Hseue, Shen, Chen, Hsu, & Liu, 1998; Katoh, 1994; Livens, Anthonis, Mahy, & Scheunders, 2003; Mandic, Delac, & Grgic, 2007; Shiraiwa, Hidaka, Mizuno, Sasaki, Ohta, & Usami, 1998; Song & Luo, 2002). The following paragraphs refer to some of them in more detail.

Braun et al. (1996) investigated the influence of viewing conditions and colour appearance models on observers' response behaviour, i.e. they wanted to determine the most appropriate viewing condition for cross-media image reproduction studies. In their study observers were presented with a printed image that was placed inside a booth with controlled illumination. Two images of the same scene, which had been created using different colour appearance models, were shown on a CRT monitor. The observers' task was to indicate, which of the reproductions looked similar to the original. Overall, the twelve observers completed the same task for all combinations of five colour appearance models (von Kries, CIELAB, RLAB, Hunt and Nayatani), five viewing (memory, successive binocular, simultaneous binocular, simultaneous haploscopic and successive haploscopic) and two illumination conditions (booth illumination D50 and CRT reproductions simulating D65, and booth illuminant A and CRT D65[1]). The five viewing conditions differed basically in two aspects, memory and adaptation levels. For two viewing conditions (memory and successive binocular)

[1] D50, D65 and A are CIE Standard Illuminants with correlated colour temperature of approximately 5000, 6500 and 2800 K, respectively (Wyszecki & Stiles, 2000).

observers had to remember the seen image for 60 sec during which they adapted to the changed illumination. For the other viewing conditions no memory was required and observers adapted to a mixed illumination environment (simultaneous binocular (figure 7.1.1-1)) or individually to different illumination conditions (both haploscopic viewing conditions). Braun and colleagues found that observers showed less variance in matching for the haploscopic than for the memory viewing conditions. They stated that under haploscopic viewing conditions good results were achieved but that this technique is highly artificial and that adaptation might not be as complete as expected. Despite slightly lower consistency the memory viewing condition was named by Braun et al. as the most appropriate condition for cross-media image matching experiments because this condition allows for full adaptation and is experienced as natural by observers. Regarding the tested colour appearance models, overall best performances were achieved for the RLAB model and the worst for the Nayatani model.

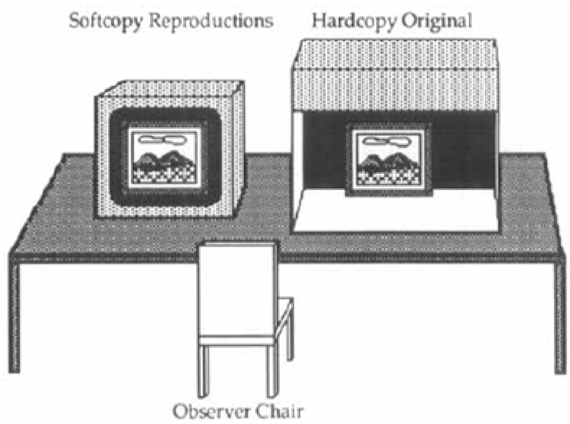


Figure 7.1.1-1. Schematic of the setup used by Braun et al. (1996) for the simultaneous binocular viewing condition. The computer-generated image was presented on a monitor (left) and the printed images in a light booth (right).

Hseue et al. (Hseue et al., 1998) studied cross-media matching performance for four colour appearance models (von Kries, Hunt, RLAB and CIELAB). They placed emphasis on the naturalness of the experimental conditions and therefore used the memory viewing condition recommended by Braun and colleagues (Braun et al., 1996), and unequal luminance levels (hardcopy 342 cd/m^2 and softcopy 60 cd/m^2). Instead of having the same luminance level for the monitor and the print images (Braun et al., 1996; Lo, Luo, & Rhodes, 1996) they presented the print images with a 5.7 times higher luminance level than the monitor images. The authors argue that this situation is more natural than a comparison between equal luminance levels as print images are normally viewed under much brighter conditions. During the procedure used by Braun and colleagues observers saw a print image once and then had to perform ten paired comparisons on the monitor. Hseue and colleagues, in contrast, allowed their observers to go back and look again at the print image when they wanted to refresh their memory of the original. Therefore, not all observers completed a series of paired comparisons following exactly the same procedure.

The results revealed that overall the most consistent performance was achieved with Hunt's colour appearance model. The RLAB model, which was named as the best one in the study by Braun et al. turned out to be the worst one. The authors reported a dependency of model performance on image characteristics, e.g. whether a landscape or a portrait was shown. Although these findings are based on a larger group of participants (30 in comparison to 12 in the study by Braun and colleagues) it remains unclear how the irregularities in the procedure affected the results.

Lo et al. (1996) carried out a similar experiment (nine observers) in which seven different colour appearance models (CIELAB, CIELUV, von Kries, BFD, Nayatani, Hunt and RLAB) were tested. Their findings were neither in accordance with those from Hseue et al. nor with those from Braun et al. Instead, Lo and colleagues found that performance was best for the Nayatani, Hunt and von Kries model and worst for CIELUV and CIELAB. However, they pointed out that this ranking was strongly image dependent (figure 7.1.1-2).

Figure 7.1.1-2. The six images shown in the study by Lo and colleagues (1996).

To summarise, the studies by Braun et al. (1996), Hseue et al. (1998) and Lo et al. (1996) were conducted under similar viewing conditions. The main difference was in the luminance levels under which images were shown. Braun et al. and Lo et al. presented the monitor and the print images under equated luminance level whereas Hseue et al. purposely chose them to be unequal. All these studies tested primarily the same colour appearance models but obtained very different results. This inconsistency does not allow us to draw general conclusions about the accuracy/quality of colour appearance models. Particularly, the ranking of the models reported by Lo et al. differs considerably from that in the other two studies and could be due to the small number of participants.

There are more colour appearance models available apart from those mentioned here and the development of new models is an ongoing process. The most recent one, for example, is the CIECAM02 (Moroney, Fairchild, Hunt, Li, Luo, & Newman, 2002), which was proposed by the CIE Technical Committee 8-01.

7.2 Performance comparability for cross-media studies

The results for real-world and computer-simulated stimuli have rarely been compared directly. As outlined above, nowadays many studies in vision research employ computer-generated and displayed stimuli. Although the results from studies employing such stimuli have been used to infer the processes of visual perception in natural environments, i.e. the real world, it is crucial to establish whether this substitution is valid.

In order to obtain two comparable sets of data, one collected from a real object experiment and the other one from an experiment using displayed stimuli, several difficulties regarding the experimental design have to be overcome. The main challenge lies in creating perceptually equal stimuli. The natural environment provides an extensive range of colours and brightness levels, which cannot be displayed on CRT monitors because of technical limitations. For rather simple stimuli, e.g. a red square on a grey background, it is sufficient to control colours and brightness levels but in order to create and display more complex scenes the interactions between surfaces and the appearance of different materials and textures have to be considered as well.

The following paragraphs provide details of studies that compared the outcomes of real-world and computer-simulated stimuli, some focusing specifically on colour appearance.

7.2.1 Comparing CRT and Munsell focal colours

A study by Berlin and Kay (1969) was ground-breaking in the field of colour categorisation and colour naming behaviour (see chapter 6), and motivated the work of many researchers (e.g. Boynton et al., 1989; Heider, 1972; Regier, Cook, & Kay, 2005; Roberson et al., 2004). In most of these studies either real surface colour samples or displayed colours were used. Kaufmann and O'Neil (1993) studied whether colour names and focal colours are identical for surface and displayed colours, i.e. whether there is a difference in perception for reflective and self-luminous surfaces. Their study was designed to identify the focal colours for the colour names white, grey, pink, red, orange, yellow, green, aqua, blue and purple when displayed on a CRT monitor. Subsequently these CRT focal colours were compared with those identified by Berlin and Kay for Munsell colour chips. Kaufmann and O'Neil created for each of the ten colour names a set of colours, from which observers (40 in total) then determined the best example of a colour name, in other words, they determined the focal colour. The comparison was conducted for only seven colour names; the colour names aqua, grey and white were excluded from the analysis.

Kaufmann and O'Neil found that the CRT focal colours for the colour names green, orange, purple and pink coincided almost perfectly with the Munsell focal colours. The CRT focal colours for red, yellow and blue were significantly different from the Munsell focal colours. The authors argue that the differences in CRT and Munsell focal colours were due to technical limitations of the CRT. They state, for example, that their yellow focal colour was created by maximum input of the red and green electron guns, resulting in the highest luminance value of the yellow colour set. Furthermore, this colour was also the most saturated one of this set. As already reported (see chapter 4), it is an often noticed phenomenon that observers tend to recall colours with higher levels of saturation. In the study by Perez-Carpinell et al. (1998a) yellow was one of the less accurately recalled colours and observers tended to pick the sample with the highest saturation. The same pattern of matching behaviour would also explain the difference between the blue CRT and the blue Munsell focal colour. The difference between the two red focal colours was due to technical issues rather than colour memory. The red Munsell focal colour was well outside the gamut of Kaufmann's and O'Neill's monitor and could therefore not be displayed. However, observers determined a red CRT focal colour which had approximately the same hue as the Munsell focal colour.

7.2.2 Matching surface and displayed colours

Kaufmann and O'Neill compared data resulting from a real surface experiment with data from an experiment using displayed colours. Hence, only the comparison was conducted across media but not the actual matching. A similar approach was chosen by Granzier et al. (Granzier, Brenner, & Smeets, 2009a) to study whether surface colours are matched fundamentally differently than displayed colours. In an earlier study (Granzier, Smeets, & Brenner, 2006) the same authors had reported that matches between surface colours were more accurate than between surface colours and displayed colours. They argue that this was due to the ambiguous interpretation of the displayed colour and suggest that colour matching within the same medium (surface colour to surface colour or displayed colour to displayed colour) would be more accurate and less variable than cross-media matching. Because if the visual system makes a fundamental distinction between reflective and self-luminous surfaces, then it does not make sense to match self-luminous and reflecting colours. On the other hand, if colours are always judged independently from their origins then one would expect to find no difference in cross-media matching.

Granzier and colleagues (2009a) conducted a study where colour matching was performed for four different conditions: surface to surface, display to display, surface to display and display to surface. During every stage of the experiment the four observers were fully aware of the origin of the presented colours. The authors reported that matches were most variable in the surface to display condition and that the overall systematic error was greatest in this condition. Both these findings were significantly different from those obtained for any of the other conditions. Granzier et al. suggest therefore, that the visual system treats the two types of colour-origin differently. However, it remains unclear why matching performance was so poor in the surface to display and not in the display to surface condition, which was cross-media as well. Matching performance was generally good for the surface to surface and the display to display condition and the authors concluded that performance is not dependent on the medium however, but that a mixture of self-luminous and reflective media causes confusion when presented simultaneously.

7.2.3 Comparing lightness matching across media

Another study that investigated whether observers' performance is medium dependent was carried out by Agostini and Bruno (1996), who compared lightness contrast in CRT and paper stimuli. Similar to the study by Kaufmann and O'Neill (1993) matching was performed within a medium and the subsequent comparison was conducted across the result from the different media. The matching consisted of selecting an identical looking sample out of an array of seven that were shown next to the test field. The results revealed that observers' performance (overall 156 observers participated in this study) was not consistent between paper and CRT matches. This was also found by Wu et al. (Wu, Wardman, & Luo, 2005), who studied lightness contrast using displayed colours and coloured fabric, but only for a group of nine observers. Wu and colleagues came to the same conclusions despite employing a very different methodology than Agostini and Bruno. In Wu's and colleagues' study matches were performed for two conditions (1) matching a displayed colour to a displayed test field by adjustment and (2) matching a coloured fabric to a displayed test field by adjustment of the same. For the data analysis Wu et al. compared cross-media (condition 2) with within-medium (condition 1) matching results.

In the studies by Kaufmann and O'Neill and Granzier et al. observers were aware of the origin of the presented colour and Granzier and colleagues pointed out that this might affect observers' performance significantly. However, it is not clear whether observers were aware of the origin of the presented stimuli in the study by Agostini and Bruno. Wu et al. purposely masked the surround of the colour stimuli in order to minimise available cues to their origin. If the awareness of the origin of the colours is essential for the visual system then the reported results of Wu et al. might be due to confusion of the visual system and misinterpretations of the stimuli.

Another comparison between perceived lightness of real and displayed surfaces was made by Schirillo et al. (Schirillo, Reeves, & Arend, 1990). Originally Gilchrist (1977, 1980) had studied the effect of spatial arrangement (i.e. depth information) on perceived lightness and had found a dependency between these factors. While Gilchrist had used exclusively real surfaces Schirillo et al. replicated and displayed Gilchrist's setup on a monitor. Schirillo and colleagues observed the same dependency as described by Gilchrist, although their effect was smaller in magnitude.

7.2.4 *Contrasting observers' performance for real-world and computer-simulated scenes*

Kaufmann and O'Neil (1993), Agostini and Bruno (1996) and Wu et al. (2005) presented their stimuli as isolated squares, and although the setup used by Granzier et al. (2009a) contained a selection of colourful household objects the actual test field/stimulus was of a simple geometrical shape. A more complex setup was created by Johnston and Curran (1996). Their study was motivated by shape-from-shading illusions ('three-dimensional curvature contrast' and 'illuminant-position effect on perceived curvature') that had been observed for computer-generated images. Johnston and Curran investigated whether these effects would also occur for real objects under real illuminants. Therefore, they created a three-dimensional scene consisting of curved matte white surfaces and images of exactly the same scene were also created. In a series of experiments observers (in total 67) were presented with the real-world and the rendered scene and their task was to decide which of the two central surfaces appeared more curved. The authors reported that (a) the same effects that had been observed for computer-generated images occurred also for real object scenes and (b) that the magnitude of these effects were equivalent regardless of the medium of presentation. However, Johnston and Curran did not draw general conclusions about the comparability of cross-media experiments because their scenes lacked complexity.

Instead of using a highly artificial setup Tatler et al. (Tatler, Gilchrist, & Land, 2005) presented their observers (in total 75) with natural scenes. More specifically, they showed six different real-world rooms (laboratory, an office, a waiting room, a seminar room, a dining room and a kitchen) and photographs of exactly the same rooms to two different groups of observers. The authors investigated the retention of object information from these scenes and the later stability of that information by asking questions about specific features in the scenes. (Observers filled in a questionnaire of nine questions about object presence, object shape, absolute position, relative distances between objects and object colour.) Tatler and colleagues used real scenes and photographs to verify the ecological validity of using two-dimensional images of three-dimensional scenes.

Despite the fact that the real-world scene extended over a larger field of view, covered a wider range of brightness and possessed real depth the authors found a high degree of consistency between the results of both viewing conditions. They therefore concluded that the use of displayed photographs is a valid method to study object and scene perception.

7.3 Summary

Generally, two concepts within cross-media studies should be distinguished (1) cross-media comparison and (2) cross-media matching and comparison. In (1), matching is performed using only one medium to obtain one set of data. A second set of data is obtained by completing exactly the same matching task but using a different medium. Subsequently, the two datasets are compared for consistency. Several of the studies discussed here (Agostini & Bruno, 1996; Gilchrist, 1977, 1980; Johnston & Curran, 1996; Kaufmann & O'Neill, 1993; Schirillo et al., 1990; Tatler et al., 2005) applied this concept and the majority of authors reported that observers' performances for real and displayed stimuli are comparable. Kaufmann and O'Neill did not find perfect consistency between media for all test stimuli but that might have been due to technical limitations of the display device. Only Agostini and Bruno came to the conclusion that matching yields different performances depending on the medium.

In (2), cross-media matching is performed to obtain one set of data. A second set of data is obtained by using only one medium. Subsequently, the two datasets are compared for consistency. This concept was applied by Granzier et al. and Wu et al. Granzier and colleagues found considerable variations for cross-media matching; surprisingly, only for the condition in which the test stimulus was a real surface and matching was performed on a monitor. In the study by Wu et al. cross-media matching was performed from fabric to display and not vice versa. However, the cross-media comparison revealed significant inconsistency.

Authors mainly concluded that the substitution of real objects by computer-generated and displayed objects was ecologically valid under the tested conditions. In order to generalise this statement more studies have to be conducted using a larger variety of stimuli and viewing conditions.

8 Methods

Experiments were carried out using real object stimuli as well as computer simulations. The first part (8.1) of this chapter will concentrate on the methods for the real-world (RW) experiments and the second part (8.2) on the computer simulations (CS).

8.1 Real world

8.1.1 Experimental setup

The RW experiments took place in a lighting booth sized 230 x 230 x 230 cm. The two side walls and the back wall were covered with black wool cloth to provide minimally-reflective surfaces and the floor was covered with dark blue carpet. An additional cloth divided the booth into two compartments. One compartment was used as a palette showroom and the other one contained a 3D scene. Each compartment was illuminated by a spotlight, which was mounted in an overhead position (figure 8.1.1-1). The spotlight that illuminated the palette was fixed so that the incident beam of light coincided with the normal of the palette surface. The exact position of the spotlight that illuminated the 3D scene is described in section 8.1.5.

8.1.2 Illuminants

Two low-voltage spotlights (Altman MR16 Micro Ellipses, 75 W, 36° reflectors) were used in the setup: one to illuminate the palette and the other to illuminate the 3D scene. Filters were placed in front of the spotlights to adjust their chromaticities, whereas a

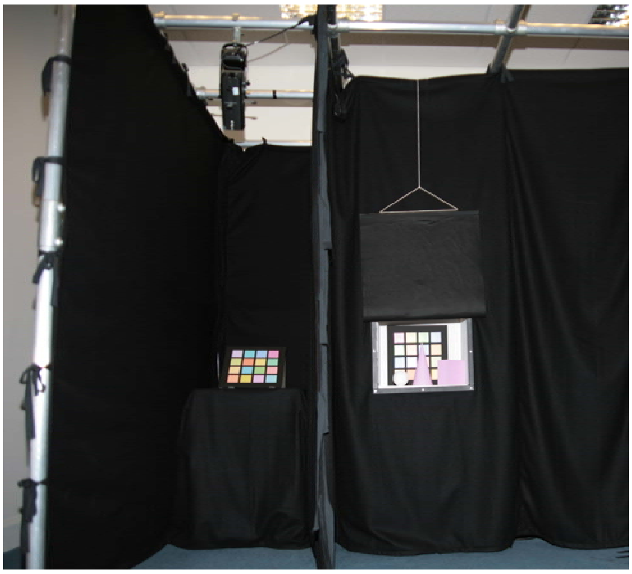


Figure 8.1.1-1. The lighting booth showing the palette (left) and 3D scene (right) compartment. The light source for the palette compartment is visible on the top left.

dimmer box (Betapack 2 by zero88) controlled the intensity. The spotlights were warmed up for at least 30 min before each data collection.

Four different LEE Filters (Lee Filters, 2008) were used to generate the experimental illuminants D1, Tun, D2 and Lily. Table 8.1.2-1 provides the CIE xyY chromaticity and luminance values as well as the colour temperatures of the experimental illuminants; figure 8.1.2-1 shows the coordinates of the illuminants in the CIE xy chromaticity diagram and figure 8.1.2-2 the transmission spectra of the illuminants. Measurements were taken with a PR650 SpectraScan® Colorimeter[1] (Photo Research, Inc., 2008) using a certified white reflectance standard (Labsphere®). The CIE recommends taking measurement under standardised illuminating and viewing conditions. All measurements reported in section 8.1 were taken under the (0/45)-condition, in which the incident beam of the spotlight coincided with the normal of the reflectance standard

[1] The colour accuracy of the PR650 is ±0.0015 CIE 1931 x, ±0.001 CIE 1931 y (0.006 CIE 1931 for CRT's typical). The luminance accuracy is ±2 % of the calculated luminance at 2856 K according to the manufacturer (Photo Research, Inc., 2008).

(0 deg) and the spectroradiometer measured the reflected light at 45 deg (Wyszecki & Stiles, 2000). The (0/45)-condition is illustrated in figure 8.1.2-3.

Label	Filter notation	CIE x	CIE y	CIE Y (cd/m^2)	Colour temperature (K)
D1	LEE Filter No. 201 Full C.T. Blue	0.370	0.377	84.52	4290
Tun	LEE Filter No. 204 Full C.T. Orange	0.517	0.425	152.90	2160
D2	LEE Filter No. 201 Full C.T. Blue	0.418	0.405	60.40	3350
Lily	LEE Filter No. 704 Lily	0.462	0.352	44.61	-

Table 8.1.2-1. The table provides the exact notation of the filters that were used to generate the four experimental illuminants (D1, Tun, D2 and Lily) and their CIE chromaticity and luminance values. Colour temperatures are listed in the last column. Lily does not have a colour temperature as it lies off the blackbody locus.

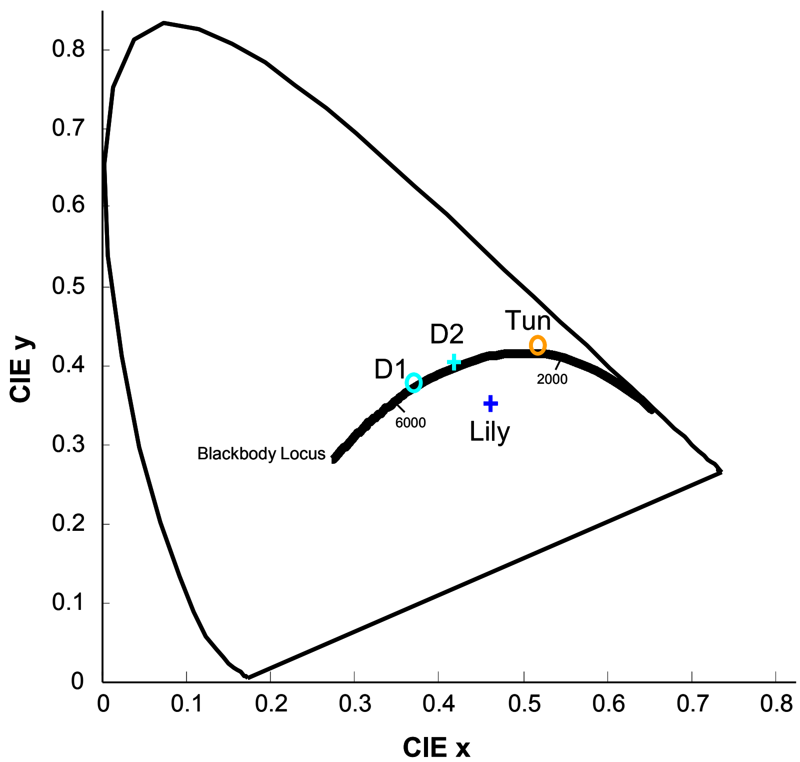


Figure 8.1.2-1. Coordinates of the experimental illuminants D1, Tun, D2 and Lily in the CIE xy chromaticity diagram.

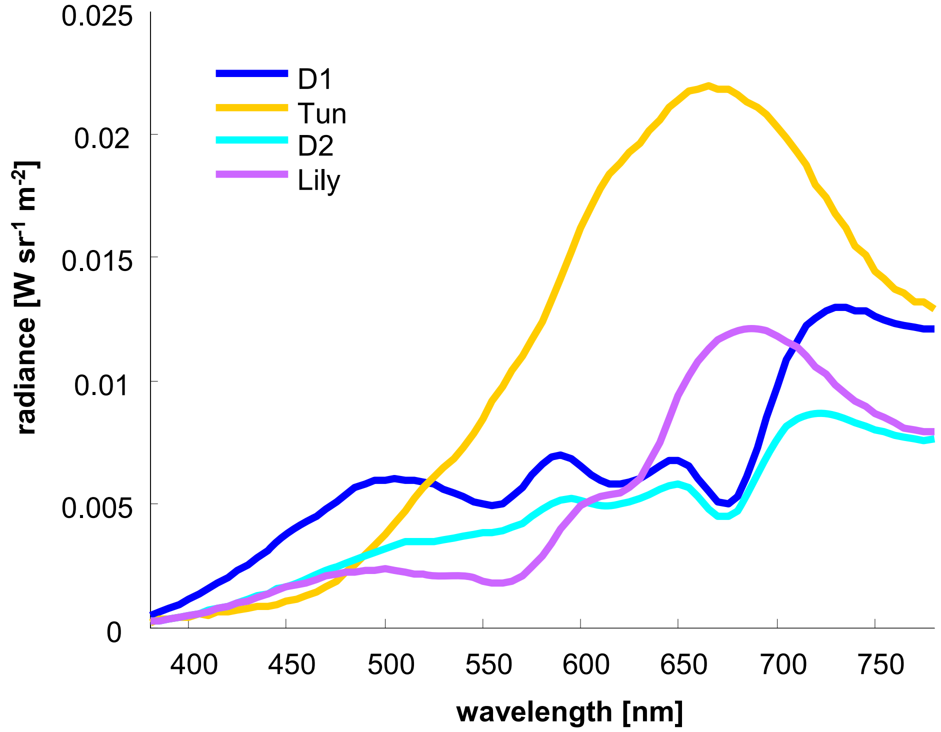


Figure 8.1.2-2. Transmission spectra of the four experimental illuminants.

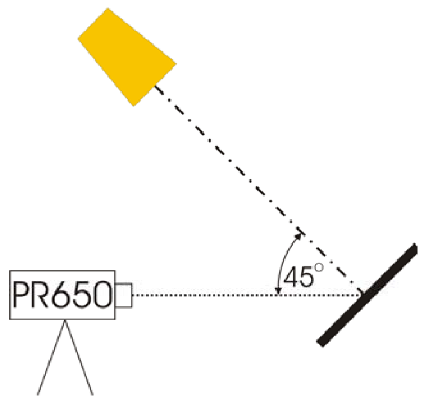


Figure 8.1.2-3. Measurements of the stimuli were taken under the (0/45) - standard illuminating and viewing condition.

8.1.3 Stimuli

Coloured papers from the NCS collection were used to create the RW stimuli (as pilot studies using NCS papers had already been conducted). Forty-eight colours from all over the colour circle were chosen. The overriding requirement was a balance between similarly coloured alternatives but at the same time discriminable from each other. This limited the number of usable papers. As a compromise between making swatches similarly saturated but having enough alternatives available all colour swatches had the same blackness value of 10 and three different chromaticness values of 30, 40 or 50. Furthermore, it was aimed to offer enough alternatives (thereby not making the task too easy), to allow a rotationally invariant layout of the palettes (which required a square number of samples) and to show reasonably large swatches. The 48 swatches were therefore divided into three groups that were arbitrary named 'Blue', 'Red' and 'Yellow'. Note that green swatches were part of the blue and the yellow colour group. From the 16 (4^2) colours of a colour group, two were determined as target colours and identified as B9 and B16 for the blue, R8 and R10 for the red and Y8 and Y10 for the yellow colour group. All six target colours had a chromaticness of 40, and hence, they differed form each other only in hue; they were never at the extreme of a colour group. The three colour groups into which the hue circle was divided are indicated in figure 8.1.3-1. In the blue colour group, colours range from chartreuse to blue, in the red group from blue via purple to orange and in the yellow group from orange to chartreuse. The exact notations of the experimental colours used are listed in table 8.1.3-1 and their surface reflectance functions are provided in table AP-1 (appendix).

One peculiarity has to be mentioned. Colour S1040 R90B was part of the red and the blue colour group, and thus, there were effectively only 47 different colours. However, this will not be considered in following descriptions, analyses, etc. Figure 8.1.3-2 shows the location of all colours in the NCS colour space.

Label	blue colour group	red colour group	yellow colour group
1	S1040 B	S1030 Y60R	S1030 G50Y
2	S1040 G	S1030 Y90R	S1030 G70Y
3	S1040 B30G	S1030 R10B	S1030 Y
4	S1030 B40G	S1030 R50B	S1030 Y20R
5	S1040 B70G	S1030 R70B	S1030 Y40R
6	S1040 B90G	S1040 Y60R	S1040 G20Y
7	S1050 G20Y	S1040 Y80R	S1040 G40Y
8	S1040 G10Y	**S1040 R**	**S1040 G60Y**
9	**S1040 B40G**	S1040 R20B	S1040 G80Y
10	S1030 B	**S1040 R40B**	**S1040 Y**
11	S1050 G	S1040 R60B	S1040 Y20R
12	S1050 B	S1040 R80B	S1040 Y30R
13	S1040 R90B	S1040 R90B	S1040 Y50R
14	S1040 B20G	S1050 Y70R	S1050 G30Y
15	S1050 R90B	S1050 R	S1050 G90Y
16	**S1040 B10G**	S1050 R30B	S1050 Y

Table 8.1.3-1. NCS notation of the selected colours. The highlighted notations are of the target colours.

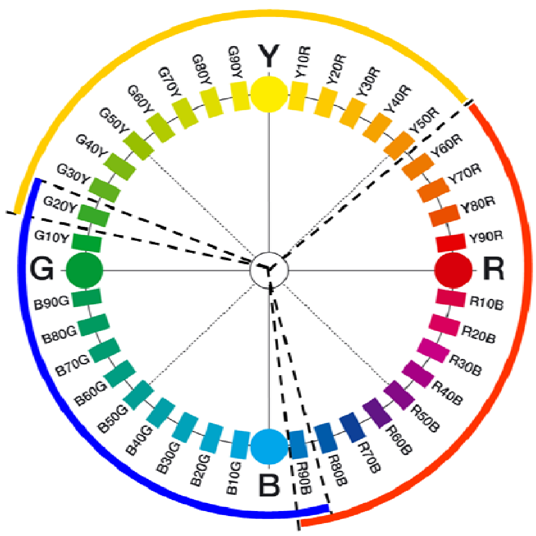


Figure 8.1.3-1. NCS colour circle. The yellow colour group covered the range from G20Y to Y50R, the red colour group covered the range from Y60R to R90B and the blue colour group covered the range from R90B to G20Y. Note that green samples were part of the yellow and the blue colour groups and that there was an overlap in hue between the blue and yellow, and the blue and red colour groups. Modified from (Scandinavian Colour Institute: Natural Colour System, 2004).

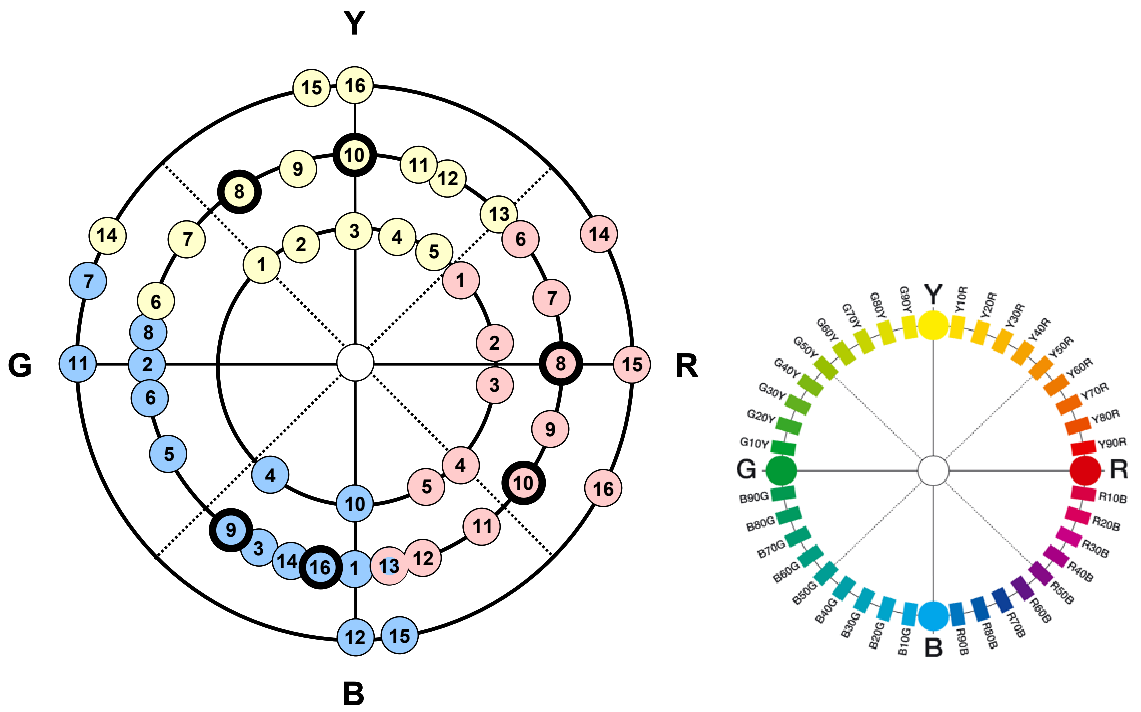


Figure 8.1.3-2. Left; each circle represents one of the selected NCS colours, the target colours are highlighted. The circles are colour-coded according to colour group (blue, red and yellow) and irrespective of the actual colour appearance of the paper samples. Notice that colour 13 (red circle with blue centre) was part of the red and the blue colour group. The location of the circles was determined by their chromaticness and hue. Colours with a chromaticness of 30, 40 and 50 lie on the inner, middle and outer circle, respectively and in terms of hue they were located according to the NCS colour circle (Scandinavian Colour Institute: Natural Colour System, 2004) (see right).

8.1.4 Learning and test palettes

Several palettes were created in order to present two-dimensional stimuli. The 2D stimuli were square paper swatches that had been cut from the selected NCS papers and stuck on the palettes.

All palettes were 30 x 30 cm in size and painted smoothly with matte blackboard paint (not providing texture), except one palette, which was painted with matte white paint.

All black palettes contained 16 colour swatches (4 x 4 layout). Each swatch was 5 x 5 cm and subtended approximately 1.8 deg of visual angle. The outer swatches were placed 3.5 cm away from the rim of the palette and all swatches were separated from each other by a 1 cm gap (figure 8.1.4-1 B). The swatches were arranged pseudo-randomly on the grid. The target colours were placed in different locations but never in a corner. The swatches surrounding the target colours also varied their positions between palettes.

Three different learning palettes, which were presented when observers were asked to learn an indicated target colour were created. Each of these palettes contained the six target colours together with ten other NCS papers selected randomly form all three colour groups (figure 8.1.4-1 A). The exact composition of the three learning palettes is provided in table AP-2 (appendix).

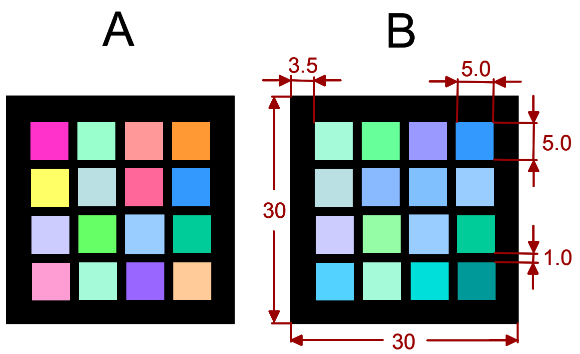


Figure 8.1.4-1. (A) Schematic of a learning palette. (B) Schematic of a test palette from the blue colour group. All dimensions are in cm. Learning and test palettes had the same dimensions.

Four test palettes, from which the colour matching was performed, were created for each colour group (12 palettes altogether). A test palette was composed of the two target colours and 14 colours from the same colour group. This variety of test palettes was produced to present the 16 colours of a group in different spatial arrangements and therefore minimise the effect of memorising positions of colour swatches. All palettes were used in any of the four rotational positions.

8.1.5 Three-dimensional scene

The 3D scene was set up in a small wooden chamber (height 35 cm, width 40 cm and depth 40 cm) whose walls and floor were painted with matte white paint. Two sides of the chamber were open, the front side to allow observation of the scene and the top for illumination purposes. There were three objects inside the chamber; on the left side in the foreground there was a white sphere (diameter 7 cm), in the centre a 20 cm tall cone (base diameter 10 cm) and on the right side in the middle a 12 x 12 x 6 cm sized box. A black palette containing a mixture of coloured swatches from all three colour groups but none of the target colours covered the back of the chamber. Figure 8.1.5-1 a photograph of the chamber and figure 8.1.5-2 shows a schematic. It had been shown in pilot studies that this disposition was appropriate for the purpose of the study.
The cone and the box were made from the paper of the target colour and subtended approximately 2.5 x 7.2 deg (cone) and the box 4.5 x 4.3 deg (box) of visual angle.

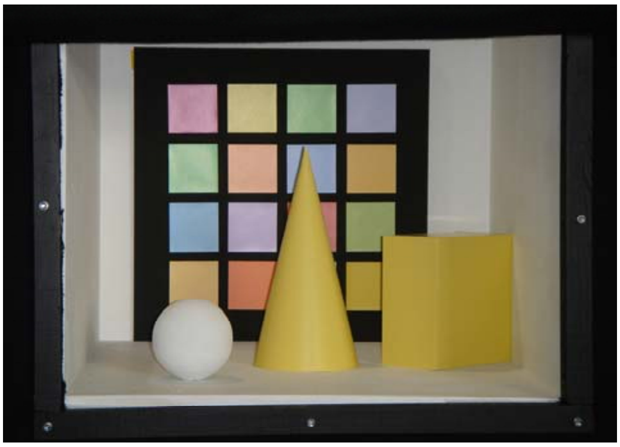


Figure 8.1.5-1. Photograph of the 3D scene. The target colour is presented by the cone and the box.

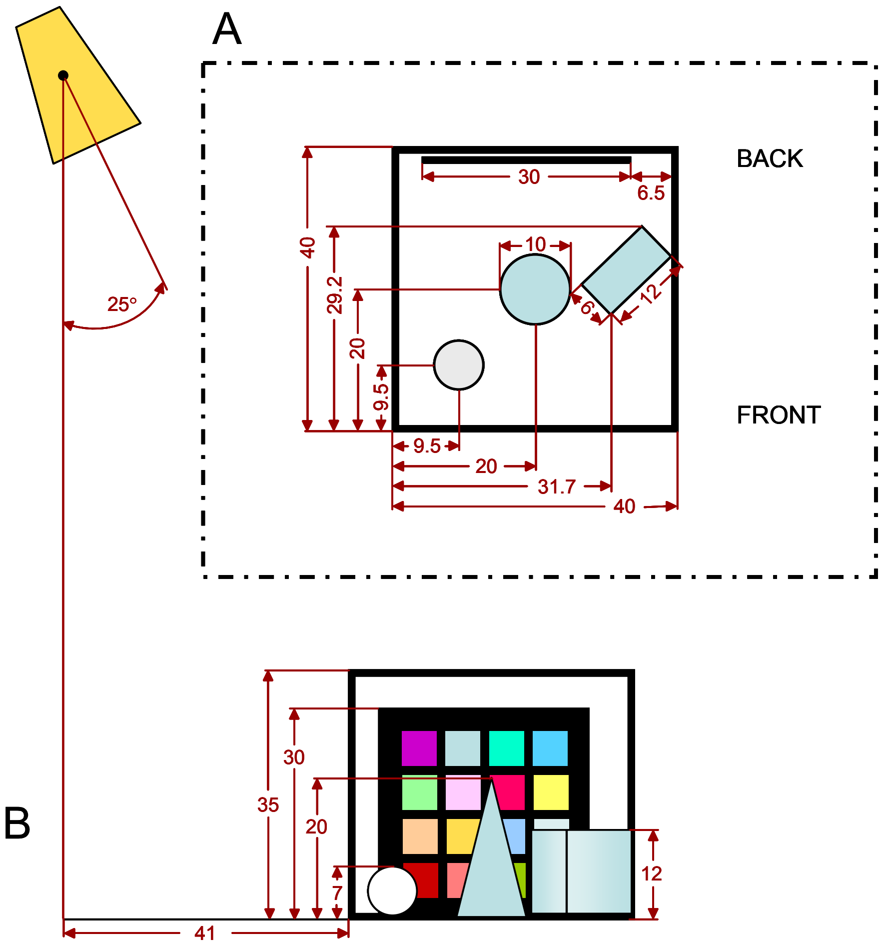


Figure 8.1.5-2. Schematic of the 3D scene setup. (A) shows a top view and (B) a front view of the layout. All dimensions are in cm.

8.1.6 Analysis of the stimuli

During the experiments the chosen colours were presented under different illuminants that strongly influenced their appearance. The similarity of colours of the same group under all experimental illuminant conditions was analysed using two different methods. The first analysis concentrated on metamerism and the second on perceived colour differences.

8.1.6.1 Metamerism-analysis

Two colour stimuli are metameric when they are perceptually identical despite different spectral compositions (Wyszecki & Stiles, 2000). In that case, both spectra lead to equal tristimulus values.

For the metamerism-analysis the tristimulus values of all colour swatches under each of the four illuminants (D1, Tun, D2, and Lily) were determined. Afterwards the tristimulus values of all swatches of a colour group were examined (a) for metamerism under the same illuminant and (b) for metamersim across an illuminant change. The main attention was directed towards the target colours as it was of most interest whether they were metameric to other colour swatches of their own colour group.

For (a) all 16 colours of a colour group were compared with each other, separately for all four illuminants. Then they were compared with each other under illuminant Tun and so on. These comparisons were completed for each colour group under each illuminant.

For (b) the target colours under one illuminant were compared to the other colours of the same colour group (that the target colour belonged to) under a different illuminant. For example, target colour Y10 under D1 was compared to all other colours that belonged to the yellow colour group under Tun. This search for metamers was carried out for the four illuminant change condition (D1 to Tun, Tun to D1, D2 to Lily, and Lily to D2) that were used in later experiments.

- **Methods**

The colour signals of all swatches were measured under the illuminants D1, D2, Tun and Lily. Measurements were taken with the PR650 under the (0/45)-condition as described earlier in this chapter.

In order to obtain the corresponding tristimulus values (XYZ), each colour signal (C_λ) was multiplied by the CIE 1931 2 deg XYZ Color Matching Functions. The factor k is a normalisation factor (Wyszecki & Stiles, 2000) and was set to be 1.

$$X = k\int_{\lambda} C_{\lambda} \cdot \bar{x}(\lambda) d\lambda \quad Y = k\int_{\lambda} C_{\lambda} \cdot \bar{y}(\lambda) d\lambda \quad Z = k\int_{\lambda} C_{\lambda} \cdot \bar{z}(\lambda) d\lambda \tag{8}$$

- **Results & Discussion**

As stated above, two (or more) colours must have the same tristimulus values to be metamers. When colour groups under the same illuminant were analysed (case (a)), no perfect coincidence of tristimulus values was found in any of the colour groups. In other words, there were no metameric colour pairs. The analysis across illuminants (case (b)) revealed no metamers either.

To summarise, the analysis showed that, based on the concept of metamerism, all colour swatches were distinguishable. This finding was consistent for both cases tested; under the same illuminant and across illuminants.

8.1.6.2 Colour differences-analysis

Another approach to decide whether colours are distinguishable is to focus on their perceived colour difference. The CIE L*a*b* colour space is an approximately perceptual uniform space where equally perceived colour differences are expressed by roughly the same distance (see section 3.3). Hence, analysing the distances between a target colour and the other colours of a colour group provides a measure of similarity. The shorter the distance between a target colour and an alternative colour is the more similar they are, and consequently, the larger the distance the easier they are to distinguish.

- **Methods**

The CIE xyY coordinates of all swatches under the four experimental illuminants were measured using the PR650 and under the illuminating and viewing condition described earlier. (These data are provided in table AP-3 in the appendix.) Then these coordinates were transformed into CIE L*a*b* applying equation 3. Afterwards, the colour difference ΔE_{ab} (equation 4) between each target colour and its alternatives was determined for each illuminant conditions individually ($6_{\text{target colours}}$ x $15_{\text{alternative swatches}}$ x $4_{\text{illuminants}} = 360_{\Delta E \text{ values}}$). The transformation required the specification of a white point for each illuminant condition. The white points used in this calculation coincided with the CIE xyY values of the experimental illuminants (table 8.1.2-2).

- **Results & Discussion**

Table AP-4, which is enclosed in the appendix, provides all calculated ΔE_{ab} values. In this section here only the representative data of target colour Y8 are presented in detail. The four plots in figure 8.1.6.2-1 show the distribution of the alternative swatches for target colour Y8 under the four experimental illuminants, arranged by the magnitude of their ΔE_{ab} values. The swatch with the smallest ΔE_{ab} value, and therefore, the most similar one to target colour Y8 was Y2, which is represented by the first bar on the left. Its ΔE_{ab} values under the different illuminants lay between approximately 7 and 8. As ΔE_{ab} values increase from left to right the alternative swatches change from being similar to Y8 to being clearly different. The most different colour swatch of this colour group was always Y13. Its ΔE_{ab} values varied roughly between 35 and 40 units. Although, the order of the swatches changed in some cases the overall distribution of the swatches was fairly stable under all illuminants.

The calculated colour differences for the other target colours followed a very similar pattern. However, the overall magnitude of ΔE_{ab} values differed between target colours. The ΔE_{ab} values for the target colours B16 and R8, for example, varied between 2.7 and 45.5, and 8.3 and 70.6, respectively.

The analysis of the colour differences was carried out in order to investigate whether every alternative swatch of a colour group was distinguishable from a corresponding target colour. Therefore, only the colours with the smallest ΔE_{ab} values were considered. Mahy et al. (Mahy, Eycken, & Oosterlinck, 1994) had evaluated uniform colour spaces and reported that the just noticeable difference was 2.4 ΔE_{ab} units for surface colours in the CIE L*a*b* colour space. That means, when the difference between two colours is smaller than 2.4 they are undistinguishable. Here, all calculated ΔE_{ab} values lay above this threshold except one. The difference between target colour B9 and the alternative swatch B3 under Tun was only 2.3 units and therefore, they were classified as undistinguishable. B9 and B3 had already attracted attention in the metamerism-analysis where they were metamers under D2. This time they showed striking similarity under Tun but it has also to be mentioned that the ΔE_{ab} value of B3 under D2 was only 2.8, barely above threshold. This is a further indication that B3 and B9 were indeed very similar.

The analysis of perceived colour differences referred to the situation where all swatches were under the same illuminant. A calculation of ΔE_{ab} values across illuminants was not possible because two colour swatches under two different illuminants involve to two different white points. However, equation 4 does not account for two white points.

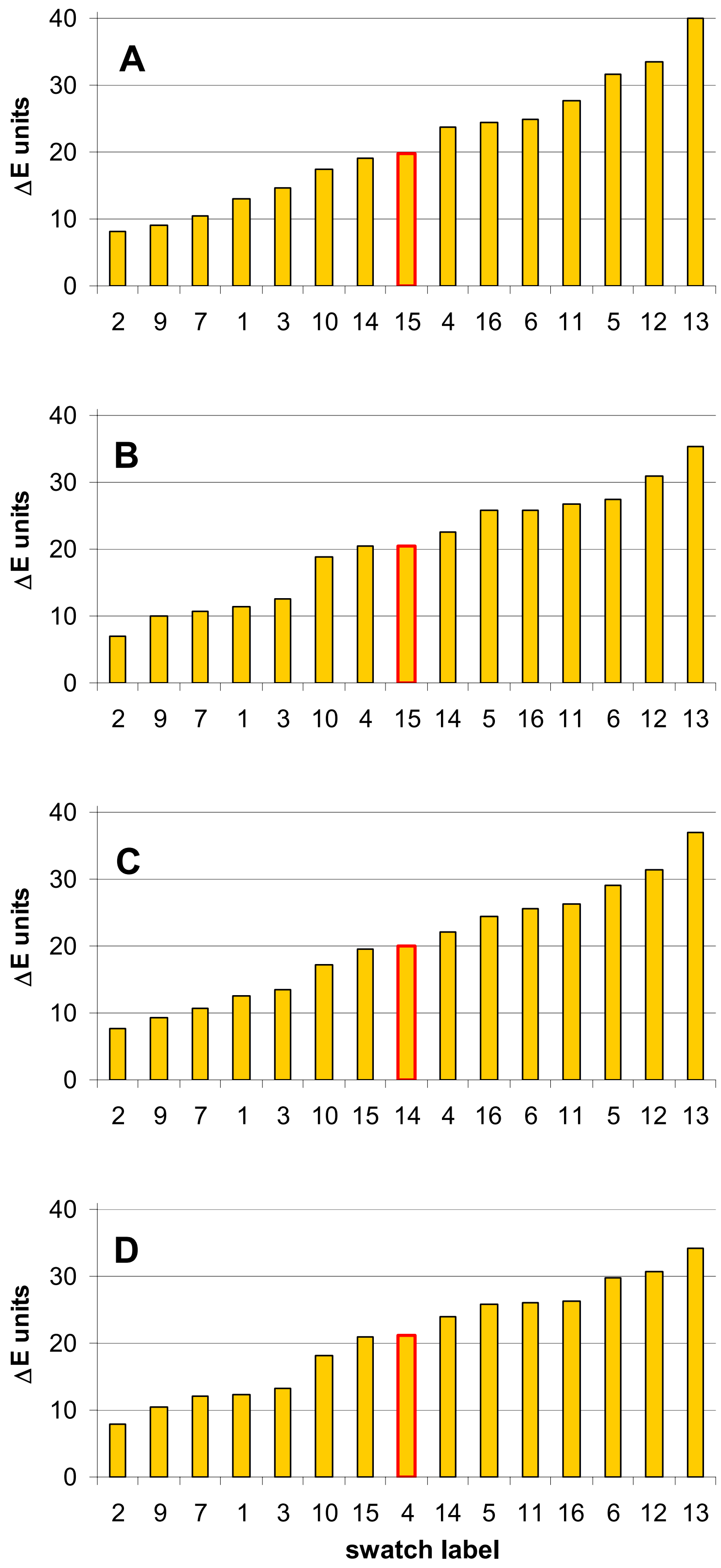


Figure 8.1.6.2-1. Colour differences (ΔE_{ab})between target colour Y8 and its 15 alternative colour swatches under illuminant D1 (A), Tun (B), D2 (C), and Lily (D). The median values are highlighted.

Figure 8.1.6.2-2 shows the distribution of all 48 colour swatches in the CIE a*b* chromaticity diagram under illuminant D1, Tun, D2, and Lily. The 16 colour swatches of a colour group build a cloud around the target colours, which are indicated by the filled circles. The closer an alternative colour swatch is to a target colour the more similar are these two colours. This statement is valid even though the CIE a*b* chromaticity diagram does not include the luminance component because luminance values of colour swatches of the same group differed only by approximately 5 %. Although, the absolute position of the colour swatches changed under the different illuminants they maintained more or less their relative position to each other. Only the swatches under Tun spread over a larger area but still the distribution of all swatches was similar to the one of the other illuminants.

To summarise, under all tested conditions the target colours were distinguishable (in terms of colour differences) from the rest of the colour swatches of their group. In only one occasion a target colour and an alternative swatch were classified as undistinguishable.

8.1.6.3 Level of difficulty

As mentioned in the previous section the alternative colour swatches of different target colours were not distributed homogenously with respect to their colour differences. This section now focuses on the possible effect of this inhomogeneity on matching behaviour. Instead of analysing the distribution of alternative swatches of each target colour the analysis will be conducted per colour group. The actual selections of alternative colour swatches for two target colours of the same colour group were identical apart from the target colours themselves. For example, R8 was an alternative colour for R10 and vice versa.

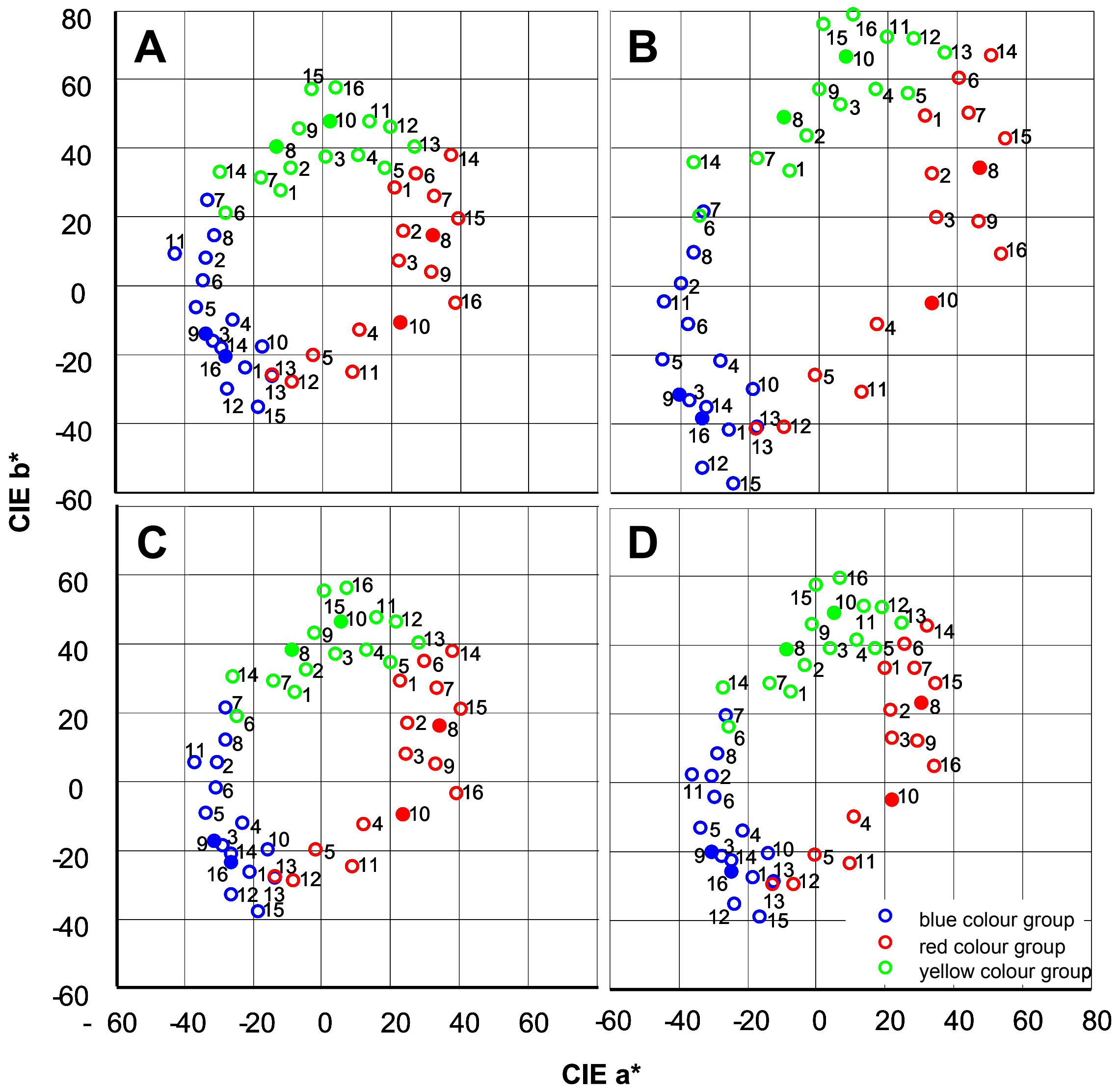


Figure 8.1.6.2-2. The circles represent the positions of all 48 colour swatches under illuminant D1 (A), Tun (B), D2(C), and Lily (D) in the CIE a*b* chromaticity diagram. The filled circles indicate the target colours.

The median ΔE_{ab} value of a distribution of alternative colours was selected as a measure of how widely these colours spread. The smaller the median ΔE value the smaller the spread of colours in colour space, i.e. the perceptually closer they were. Table 8.1.6.3-1 provides the median ΔE_{ab} values of the alternative swatches of all target colours and all four illuminants.

median ΔE_{ab} *values*	target colour					
illuminant	**B9**	**B16**	**R8**	**R10**	**Y8**	**Y10**
D1	16.9	14.8	21.4	27.3	19.9	17.6
Tun	15.8	14.0	19.9	29.4	20.6	14.9
D2	16.7	13.5	20.7	28.2	20.0	17.3
Lily	16.8	13.5	19.3	29.4	21.2	16.1

Table 8.1.6.3-1. This table provides a summary of all median ΔE_{ab} values of the selection of alternative colour swatches for each of the six target colours and for all four experimental illuminants.

Generally, the median ΔE_{ab} values were lowest for the two blue target colours and highest for the red ones. The yellow target colours had intermediate values. This means that the blue colours lay perceptually closest together and that the colours of the red colour group, in turn, were furthest away from each other. The yellow colour group was in-between the other two groups. This can be seen in all plots of figure 8.1.6.2-2. The effect that this may have on matching behaviour is best explained by an example. Refer to A of the same figure, target colour B16 was surrounded by several very similar looking swatches. Hence, it was likely that observers would get easily confused when matching B16. The distance between R10 and the most similar looking colour in turn was rather large, in other words, they were easily distinguishable. Hence, matching R10 would be less uncertain.

Based on these observations it became clear that the three colour groups had different levels of difficulty for matching. It was hypothesised that it would be easiest to match the red target colours, blue target colours would be most difficult to match and the yellow ones would have an intermediate level of difficulty.

8.2 Computer simulation

8.2.1 Setup

The experiments using computer generated stimuli took place in a room with minimally-reflective environment; the walls were painted black, the floor was covered with a dark blue carpet and the frame of the experimental monitor was covered with black cardboard. Apart from the monitor there were no further light sources present in the room during data collection.

8.2.2 Material

A 42-bit graphics card (ViSaGe – 256 MB: VSG7102.00F0) from Cambridge Research Systems (CRS) was used to generate the stimuli. The graphics card was controlled by MATLAB 7.0.4 (Mathworks®). The stimuli were presented on a calibrated (Garcia-Suarez 2009 unpublished) CRT monitor (Mitsubishi Diamond Pro 2070SB). The monitor output was checked periodically with a Minolta ColorCAL colorimeter, and the monitor was recalibrated when necessary (luminance variations > 5 % and chromaticity variations > 1 %). The accuracy of the Minolta ColorCAL colorimeter is ± 0.004 for the luminance as well as the chromaticity. These specifications refer to measurements taken under illuminant D65 at 40 cd/m^2 according to the manufacturer (Cambridge Research Systems, 2008).

The monitor was set to a resolution of 1024 x 768 pixels, a frame rate of 120 Hz and was warmed up for at least 30 minutes before data collection.

8.2.3 Monitor check

For a given colour the output of a monitor is not constant across the viewable area. Therefore, the monitor's luminance and chromaticity output was verified (a) at different locations all over the monitor and (b) within a smaller area that was used to display the stimuli.

(a) Measurements were taken at five different positions of the monitor: centre, top left corner, top right corner, bottom left corner, and bottom right corner. Three different colours, which covered a similar range of chromaticity and luminance values as the experimental colours, were displayed and CIE xyY values were determined using a PR650 spectroradiometer.
The chromaticity values were within the measuring tolerance of the PR650 for almost all measurements. Only in a few occasions this tolerance was exceeded but only by less than 1 %. However, the luminance values varied considerably between measuring points. The overall deviation at the five measuring points was as follows: centre, ±1.0 %; bottom left corner, ±3.6 %; bottom right corner, ±5.0 %; top left corner, ±8.0 %; top right corner, ±12.1 %.
To summarise, the monitor output can be described as homogeneous with respect to chromaticity and as heterogeneous with respect to luminance.

(b) The experimental stimuli occupied a smaller central area of the monitor. Therefore, a more detailed analysis of luminance variations in the relevant area was carried out and a new palette was created (RW and CS) consisting of the same colour (CIE x 0.51, y 0.38, Y 36.7) in all 16 positions. Now for each position a measurement was taken which allowed the direct comparison between RW and CS.

Measured luminance values were not constant either on the RW or the CS palette. However, within the measured area luminance varied unsystematically. The mean luminance of the RW palette was 36.6 cd/m^2 (SD 1.7 cd/m^2) and for the CS palette 35.2 cd/m^2 (SD 1.2 cd/m^2) and hence, the simulated swatches tended to be slightly darker. The perceptual colour difference (ΔE_{ab}), which was introduced by the pure luminance variations was approximately 1 unit, and thus insignificant. Based on these results it was decided that there was no need to compensate for luminance variations neither for the RW nor for the CS case.

8.2.4 Transformation of the experimental colours

The CIE xyY values of all 48 coloured paper swatches under the four experimental illuminants had already been specified. In order to simulate and display these 192 apparently different colours ($16_{colours}$ x $3_{colour\ groups}$ x $4_{illuminants}$ = 192) on the monitor their CIE xyY values were transformed into RGB values by using equation 5. The necessary matrix was determined during the calibration process of the monitor.

The number of displayable colours on a CRT monitor is limited by its physical properties. Only colours, which are inside a certain range called gamut are displayable. Table 8.2.4-1 provides an overview whether the computed RGB values lay inside the gamut of the monitor. It appeared that all 48 colours under Lily lay within the gamut of the monitor. All colours of the blue and the red colour group under D1 and D2 were also displayable. However, there were a number of yellow colours under D1 and D2 (six and two, respectively) that lay outside the gamut. Only three out of the 48 colours under Tun were displayable.

illuminant	blue colour group	red colour group	yellow colour group
D1	✓	✓	x_6
Tun	x_{13}	x_{16}	x_{16}
D2	✓	✓	x_2
Lily	✓	✓	✓

Table 8.2.4-1. RGB values were computed for all colours of the three colour groups under all four illuminants. ✓ indicates that all colours of a group lay inside the gamut. x indicates that some or all colours (the subscript specifies how many) of a colour group lay outside the gamut.

Under D1 and D2 only a few yellow colours lay just outside the gamut. More specifically, for all these colours only one of the three values (R, G or B) exceeded the limit. With minor adjustments it was possible to shift these colours into the displayable range. The effect that these adjustments had on the accuracy of the displayed colours is discussed in the following section. Almost all colours under Tun, in turn, were outside the gamut. In most cases R and B values exceeded the limit by a factor up to four. It was impossible to modify these colours to fit them inside the gamut. Consequently, only the experimental colours under D1, D2 and Lily were used for CS experiments.

8.2.5 Accuracy of the simulated colours

Each of the simulated colours was displayed in the centre of the screen and their CIE xyY values were measured with the PR650 (the position of the PR650 coincided with the normal of the monitor surface). These values were then compared to those of the RW colours.

The overall accuracy varied by ±0.8 % (SD 0.5 %) for CIE x values, ±0.6 % (SD 0.2 %) for CIE y values and ±2.7 % (SD 1.1 %) for CIE Y values. For more details see table 8.2.5-1.

illuminant	colour group	CIE x	CIE y	CIE Y
	blue	0.05	0.21	2.71
D1	red	0.92	0.58	1.07
	yellow	0.38	0.29	1.83
	blue	1.26	0.67	2.44
D2	red	0.69	0.62	3.77
	yellow	0.62	0.54	3.21
	blue	1.16	0.78	3.13
Lily	red	1.60	0.62	2.96
	yellow	0.74	0.65	2.92

Table 8.2.5-1. Differences for CIE x, y and Y values between RW and CS colours were calculated for each colour separately and then averaged for each colour group under each illuminant. Differences are given in percent.

The comparison between RW and CS colours was carried out in order to decide whether the small alterations were acceptable, i.e. whether the alterations could affect the outcome of experiments. As mentioned in section 8.2.4 some colours were adjusted in order to fit into the gamut. The comparison between these colours in RW and CS did not reveal significant larger variations than for not adjusted colours.

The accuracy of the chromaticity was high. Variations of the CIE xy values were clearly under 1 % and did not influence the perceived colour differences at a noticeable level. The overall accuracy of the luminance values was slightly worse but within an appropriate range. This decision was based on the perceived colour difference, which was only slightly affected, and technical factors. Despite warming up the spotlights and the monitor before data collection their output was not completely stable. Hence, inevitable variations in accuracy were introduced by the set up. Furthermore, it had already been shown that luminance values were not constant across the RW or the CS palette. Another source of inaccuracy was the PR650 with which the measurements were taken. However, chromaticity and luminance variations were larger than the measuring tolerance of the spectroradiometer. Taking into account these sources of inaccuracy it was decided that the CS colours were as accurate as they could possibly be and that there was no need to compensate for variations.

8.2.6 Learning and test palettes

In total nine learning ($3_{\text{learning palettes}}$ x $3_{\text{illuminants}}$ = $9_{\text{generated learning palettes}}$) and 36 test palettes ($4_{\text{test palettes}}$ x $3_{\text{colour groups}}$ x $3_{\text{illuminants}}$ = $36_{\text{generated learning palettes}}$) were generated. The array and distribution of colour swatches on the CS palettes was identical to the one on the RW palettes. All palettes were used in any of the four rotational positions, exactly as the RW palettes.

8.2.7 Three-dimensional scenes

Interactions of light and surfaces in natural scenes are very complex. Light is absorbed, refracted and reflected depending on the properties of the objects and surfaces it hits. In order to render a scene of the real world as accurate as possible these physical interactions have to be taken into account. Software that does this successfully is RADIANCE (Ward Larson & Shakespeare, 1998). RADIANCE is freely available and is used in areas such as computer graphics, lighting design and architecture. It uses backwards ray-tracing to generate images of 3D scenes. Backwards ray-tracing means that the path of the light is traced from the point of measurement (e.g. imaginary eye) into the scene and back to the light source. This technique reduces computation time significantly compared to “forward” ray-tracing that would compute all light paths from a light source, most of which never reach the imaginary eye.

For the purpose of the following experiments it was essential to reproduce the colours of the RW 3D scene described above as accurate as possible. High accuracy was achieved by following a special rendering technique used by Ruppertsberg and Bloj (2008), who used 81 wavebands (instead of only three) to create a hyperspectral image. Images were rendered under the assumption that all surfaces in the RW scene were Lambertian. This was a reasonable assumption as all surfaces in the RW scene had low gloss levels (15-20)[1].

One image was created for each target colour under D2 and Lily, resulting in 12 images ($6_{\text{target colour}}$ x $2_{\text{illuminants}}$ = 12_{images}). The colour accuracy of the rendered scene was verified by comparing measurements of two target colours under both illuminants from the RW scene with those from the CS scene. Table AP-5 (appendix) provides the exact CIE values of these measurements. Measurements were taken with the PR650 at four different points in the scene (figure 8.2.7-1). The chromaticity accuracy was similar to the one measured for the palettes (variations < 1 %). However, the luminance accuracy was slightly worse than for the palettes. Values deviated up to approximately 10 %, to be more specific, luminance values were consistently smaller for the rendered images i.e. the simulated scenes were slightly darker than the RW scene.

The size of the original RW 3D scene was 40.0 x 35.0 cm (width x heights) whereas the maximal visible area of the monitor was only 40.4 x 30.5 cm. Therefore, the images were not rendered in scale but smaller; the actual image size was 28.0 x 23.0 cm. In order to present the CS scene under the same visual angle as the RW scene the distance between monitor and observer was reduced.

[1] Standard NCS colour papers have a gloss level of 15 to 20 (Scandinavian Colour Institute: Natural Colour System, 2004) according to ISO (International Standards Organisation) 2813, 60 deg. A gloss level of 0 corresponds to a completely matt surface and a gloss level of 100 to a very glossy surface.

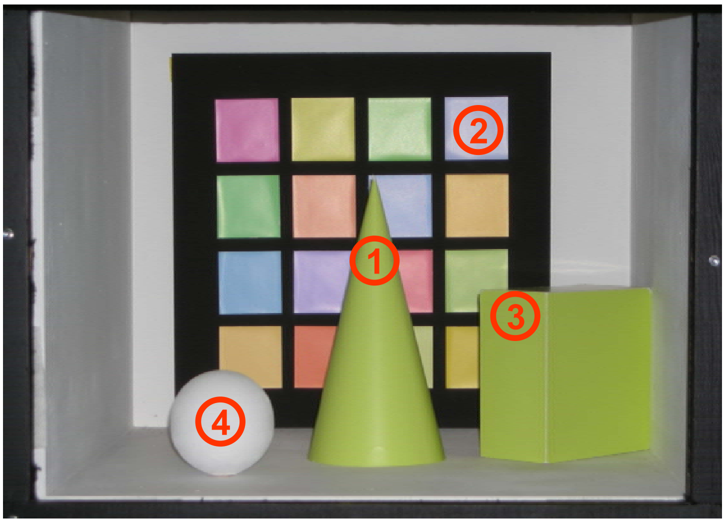


Figure 8.2.7-1. The red circles indicate the measuring points in the RW and CS scene.

9 Observer screening

The aim of all experiments reported in this thesis was to study particular aspects of colour perception of a sample of colour normal observers. Therefore, before taking part in experiments an observer screening was necessary because it had to be assured that observers had normal colour vision according to an internationally recognised colour test, the Farnsworth-Munsell 100-Hue Test and that their colour memory was satisfactory according to an ad-hoc participation test.

9.1 Farnsworth-Munsell 100-Hue Test

The Farnsworth-Munsell 100-Hue Test is widely used for assessing the ability to discriminate colour and detecting deficiencies in colour vision (Farnsworth, 1957). The test consists of 93 coloured chips which are arranged in four boxes. Each box contains colour chips that can be arranged such that they form a gradual transition in colour from the starting chip to the end chip. The start chip and the last are fixed inside the box, whereas the remaining chips are moveable. The task of the observer is to place the chips such that they form a gradual colour transition.

A total error score is calculated on the basis of misplacements. This score is a measure of whether an observer has a superior, average or low ability to discriminate colours or whether he/she has even defective colour vision. A total error score between 0 and 19 indicates superior colour discrimination, between 20 and 100 average colour discrimination and more than 100 indicates low colour discrimination or defective colour vision.

9.1.1 Observers

In total 119 persons were tested. Forty-five male observers (37.8%) were between 19 and 45 years old (mean age = 29.1). The 74 female observers (62.2%) were between 19 and 53 years old (mean age = 27.6). All observers had normal visual acuity or were corrected to normal, written informed consent was obtained from all observers.

9.1.2 Setup

The four boxes of the Farnsworth-Munsell 100 Hue Test were placed on a black cardboard on top of a table, to provide a uniform and minimally-reflective background. Apart from the room light the setup was illuminated by an anglepoise lamp fitted with a 60 W daylight bulb (mimicking D65).

One box at a time was opened by the experimenter and the chips were laid down on the black cardboard in random order. Observers were told that they had to place the chips back into the box so that they formed a gradual transition between the two fixed outer chips at either end. No time limit was set as accuracy was more important than speed. However, observers, who took very long (longer than 10 min per box), were encouraged to be more confident with their decisions.

9.1.3 Results

Thirty-three observers (27.73%) had superior, 78 (65.55%) average and eight (6.72%) lower (or defective) colour discrimination.

After several observers had completed the Farnsworth-Munsell 100-Hue Test and the participation test (the second stage of the observer screening) the experimenter noticed a relationship between observers' performance in these two tests. Observers with average colour discrimination but a relatively high total error score (between approximately 50 and 100) performed poorer in the participation test than observers with a lower total error score (between approximately 50 and 0). One reason for this could have been that their visual system did simply not allow them to perform any better. However, another reason could be in the amount of attention that was paid towards the task. While some observers completed the colour test with great care others did not. This carelessness was reflected in the relatively large number of misplacements between adjacent chips. Unfortunately, the participation test was then often completed with the same careless attitude, which led to rather disappointing results. These observers were excluded from any further experiments. In order to improve the time efficiency for the observer screening the maximum total error score to continue with the next stage was set to 50. In other words, now only observers with a total error score of less than 51 were run in the participation test.

As a result of the stricter criterion the number of observers that had to be excluded increased from 6.7% up to 21.0%. To summarise, 119 observers completed the Farnsworth-Munsell 100-Hue Test and 94 of them (32 male (34.0%) and 62 female (66.0 %) observers) passed this first stage according to the stricter criterion.

9.2 Participation test

The participation test was performed to ensure observers had good and roughly equal colour memory to obtain a homogeneous group of observers.

In most of the subsequent experiments observers' performance in colour constancy was tested. Colour constancy can be understood as a higher stage or a very sophisticated form of colour memory and hence, colour constancy is naturally limited by colour memory. It was expected that observers' level of colour constancy would be lower than their level of colour memory. Imagine an observer who correctly matches only three swatches in the participation test and one in a subsequent colour constancy experiment. It would be difficult if not impossible to determine whether this drop in performance is due to the increased difficulty of the task or only an insignificant variation. In order to be able to detect drops in performance an exclusion criterion was applied; observers needed to correctly match at least eight swatches out of 18, which represent 44.4% correct.

9.2.1 Observers

All 94 observers, who passed the colour vision test, were run in the participation test described below.

Eighty-three observers were randomly split into two groups where one group (45 observers) learned the target colours in the 2D and the other group (38 observers) in the 3D setup. A third group, which consisted of eleven observers, learned the target colours in the 2D as well as the 3D setup.

9.2.2 Procedure

The participation test was conducted in the lighting booth using the real-world setup. The materials are described in detail in section 8.1.

Observers adapted initially for 2 min to illuminant D1 by looking at the white palette. Then, a target colour was presented either on a learning palette (indicated by the experimenter) in the palette compartment or in the 3D scene by the cone and box. Observers had 20 sec to learn the colour. After the 20 sec the learning palette or 3D scene was taken from view and a test palette was presented under the same illuminant

D1. Observers were instructed to select the swatch they thought had been cut from the same piece of paper as the swatch (or the objects) they had focused on before. The same instructions for the matching were given for all colour memory and colour constancy experiments reported in this thesis. In none of the experiments was a time limit set for selection but all observers made their choice within 2 to 15 sec; no feedback was given. Figure 9.2.2-1 shows a trial sequence.

The task consisted of 18 trials in which each target colour was presented three times. This procedure was maintained throughout all colour memory and colour constancy experiments.

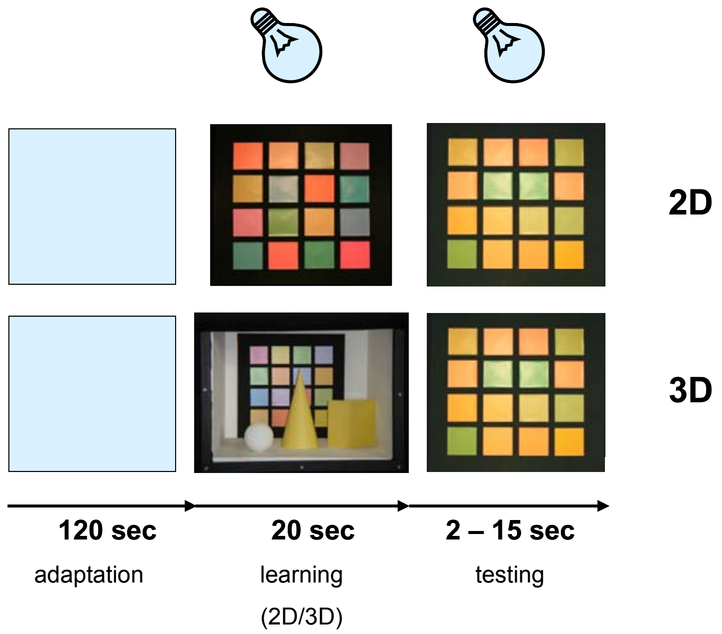


Figure 9.2.2-1. Trial sequence of the participation test; 2D task (top) and 3D task (bottom).

9.2.3 Results

Thirty-four observers (36.2%) were not able to correctly match the set minimum of eight out of 18 colours, i.e. more than one-third of observers were excluded from any further experiments.

The remaining 60 observers matched on average 9 to 10 (mean 9.5 SD 1.6) colour swatches correctly; 27 of them learned the target colour in 2D, 23 in 3D and ten in 2D as well as in 3D. The group that learned the target colours in the 2D setup (including the ten observers from the mixed group) selected the correct swatch in 58.3% (SD 1.6) of the cases and the 3D group (also including the ten observers from the mixed group) in 60.8% (SD 1.7).

In theory, it would have been necessary to establish also an upper exclusion criterion to allow detection of improved performance and to avoid a ceiling effect in the participation test. However, in practice none of the observers completed the participation test perfectly, i.e. correctly matching all 18 swatches. The highest number of correctly matched swatches was 14 and this was only achieved by a single observer. No observer was excluded because of outstanding performance.

9.2.4 Remark

The large number of dropouts shows how demanding the observer screening was. Approximately 50% of all observers failed to reach the set target for either the colour vision or the colour memory test. Although, the observer screening required more than 106 observer hours, this process was essential for the forthcoming experiments. Based on the results from the observer screening it was assumed (and later confirmed by statistical analysis) that observers formed a homogeneous group and that results between different observer groups were comparable.

The reason for this large number of dropouts lies in the choice of experimental colours. Only real surfaces were used and therefore the selection of experimental colours depended on the earlier described requirements (section 8.1.3) as well as on the samples actually provided by the manufacturer. The overall aim was to create perceptually balanced arrays of colours. The difficulty to obtain such arrays can be easily demonstrated. Imagine exchanging the two perceptually most different colours (with respect to the target colour) for two perceptually very similar ones. Just by exchanging two colours the task would become impossible to perform. On the other hand, providing an array were all colours have large perceptual differences to the target colours would make the task extremely easy. In the former example the array would lead to a floor and the latter one to a ceiling effect. These effects had been observed in pilot experiments in which perceptual coarser as well as finer colour spacing was used.

10 Colour memory

This chapter consists of three sections. Experiments are reported in sections 1 and 2, and various analyses are conducted in section 3 investigating shifts in colour memory.

10.1 Experiment 1 - Colour memory for delayed matching

As outlined in section 4.3 most studies that assess colour memory using successive matching paradigms report that colour memory deteriorates over time. It has been illustrated that the level of colour memory and its decline strongly depends on the experimental methods and the actual task observers are asked to accomplish. So far, no constant factor or measure has been established that would allow predicting the degree of deterioration.

In the present study the level of immediate colour memory has already been determined in the participation test, where observers learned target colours and matched them immediately after learning. In the following experiment, colour memory was assessed over a period of 2 min. The comparison between immediate and delayed colour memory allows quantifying the effect of time on memory. A delay period of 2 min was chosen to establish a performance baseline for subsequent colour constancy experiments, in which the same delay between learning and matching of a colour was applied.

10.1.1 Procedure

The colour memory experiment for delayed matching was conducted in the lighting booth using the real-world setup. The materials are described in detail in section 8.1.

Observers adapted initially for 2 min to illuminant D1 (which was used during the whole experiment) by looking at the white palette. Then, a target colour was presented either on a learning palette (indicated by the experimenter) or in the 3D scene by the cone and box. Observers had 20 sec to learn the colour. After the 20 sec the learning palette or 3D scene was taken from view and the white palette was shown for 2 min again. Afterwards the white palette was replaced by a test palette and observers selected the swatch they thought had been cut from the same piece of paper as the swatch (or the objects) they had focused on before. No time limit was set for selection but all observers made their choice within 2 to 15 sec; no feedback was given. A task consisted of 18 trials in which each target colour was presented three times.

Ten observers, who had passed the observer screening (chapter 9) took part in this experiment. Each observer completed two tasks; learning the target colour in 2D as well as in 3D.

10.1.2 Results

The last row of table 10.1.2-1 shows the percentages of correctly selected swatches for the delayed matching. The results are presented as hit rates. By a hit the selection of the correct swatch is understood and the hit rate is how often the correct swatch was selected, expressed as a percentage. The data is separated by colour group and whether a target colour was learned in 2D or 3D. This table also includes the data from the participation test of the same ten observers to facilitate the comparison between immediate and delayed matching.

	2D				3D			
matching	**blue**	**red**	**yellow**	∅	**blue**	**red**	**yellow**	∅
immediate	43.3%	68.3%	46.7%	52.8%	26.7%	78.3%	56.7%	53.9%
delayed	40.0%	68.3%	45.0%	51.1%	31.7%	60.0%	56.7%	49.5%

Table 10.1.2-1. Hit rates are presented for both delay periods. The data is separated by colour groups and whether a target colour was learned in 2D or in 3D. The highlighted columns indicate average hit rates over all colour groups.

A three-way repeated-measures analysis of variance (ANOVA) was conducted. The three factors were matching, which had two levels (immediate and delayed), learning environment, which had also two levels (2D and 3D), and colour group, which had three levels (blue, red, and yellow). The significance level was set at $p < 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed.

There was no significant main effect of matching (immediate-mean hit rate = 53.3% SE 2.1, delayed-mean hit rate = 50.3% SE 3.6) or learning environment (2D-mean hit rate = 51.9% SE 2.9, 3D-mean hit rate = 51.7% SE 3.3). A significant main effect was found for colour group $F(2,18) = 31.20$, $p < 0.01$. Pairwise comparison (Bonferroni corrected) revealed that the mean hit rate for the red colour group (68.8% SE 3.2) was significantly higher than for blue (mean hit rate = 35.4% SE 2.7) and yellow (mean hit rate = 51.3% SE 4.3), $p < 0.01$ and $p = 0.01$, respectively. The difference between the yellow and the blue colour groups was also significant $p = 0.01$ (figure 10.1.2-1).

None of the interactions between the factors was significant.

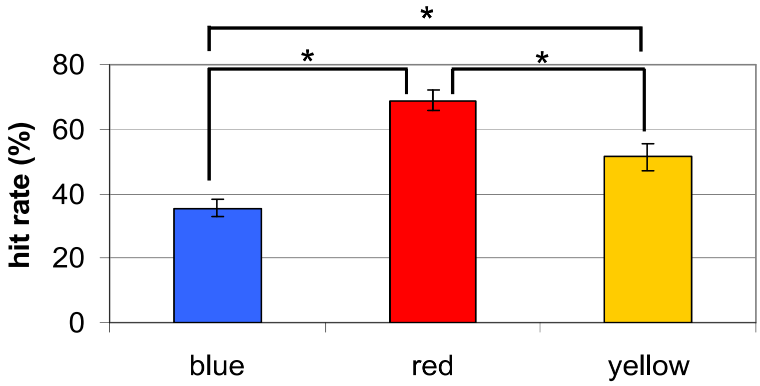


Figure 10.1.2-1. Observers' colour memory in the delayed task was best for the red and worst for the blue colour group. Intermediate performance was achieved for the yellow colour group. Error bars indicated ± 1 SE, * indicates a significant difference at $p < 0.05$.

10.1.3 Discussion

In the analysis observers colour memory for immediate and delayed matching was compared. The initial hypothesis was that performance would drop for the delayed matching. However, the hypothesis was not confirmed. It was found that observers' performance was independent of the time they waited to select a swatch from a test palette. Although this outcome contradicts most findings reported in literature (e.g. D'Ath et al., 2007; Francis & Irwin, 1998; Newhall et al., 1957; Perez-Carpinell et al., 1998a; Uchikawa, 1983) it was concluded that under this particular experimental conditions and with this specific procedure observers' ability to remember colours stayed stable over a period of 2 min.

Observers' performance was also not affected by the learning environment. Whether target colours were learned as part of a 3D scene or as flat 2D swatches on palettes did not influence the ability to recall a learned colour.

After theoretical considerations (see section 8.1.6.3) it was hypothesised that the three colour groups had different levels of difficulty. It was assumed that red would be the easiest colour group, in other words, that the hit rates for the two red target colours would be higher than those for other target colours. Yellow was expected to be the intermediate group and blue to be the most difficult one. The hypothesis regarding the levels of difficulty was confirmed by the analysis. Red was identified to be the easiest colour group, yellow the intermediate and blue the most difficult one. The lack of interactions stresses the robustness of this effect; the levels of difficulty were unaffected by the learning environment and the delay of matching.

10.1.4 Summary

The emphasis in this experiment was placed on investigating observers' ability to recall colours and on contrasting their performance for two delay periods. The results showed that under the given circumstances observers' colour memory did not drop over time.

10.2 Experiment 2 - Colour memory for 2D stimuli under different illuminants and presentation media

In this experiment it was investigated whether the degree of colour memory varies due to (a) the medium used for stimuli presentation and (b) the illuminant under which colours are viewed.

It is common practice in vision research to use computer-generated and displayed stimuli. Although these kinds of stimuli has been available for many years very few studies have investigated whether observers perform equally well for tasks that differ regarding the presentation medium of the stimuli and that are otherwise identical. Despite the small number of studies in this area most findings indicate that computer-generated and displayed stimuli may be a valid substitution for real object stimuli (Granzier et al., 2009a; Johnston & Curran, 1996; Tatler et al., 2005), but this remains to be confirmed for a wider range of experimental conditions and tasks.

Colour memory is frequently tested under daylight, which can be computer-simulated (D'Ath et al., 2007; Tarczali et al., 2006; Yendrikhovskij et al., 1999), created by a specific bulb or filter (Hamwi & Landis, 1955; Heider, 1972; Ling & Hurlbert, 2008; Newhall et al., 1957; Perez-Carpinell et al., 1998a; Siple & Springer, 1983) or be real daylight (Heider, 1972). However, colour memory has rarely been compared for different illuminants. Some studies investigating memory colours for familiar objects tested colour appearance under different illuminants (e.g. Olkkonen et al., 2008; Perez-Carpinell et al., 1998b). Observers matched the appearance of a test stimulus or selected a colour that represented best a mentioned object under different illuminants. In such a task it is unlikely that the illumination under which the object was first memorised coincides perfectly with the experimental illumination condition. Hence, an illumination change occurred between these two situations and therefore observers perform a colour constancy rather than a colour memory task. In order to study the effect of different illuminants on colour memory, abstract stimuli have to be used, whose colour appearance cannot be related to prior experiences.

In the present experiment colour memory was studied under several illuminants using real paper and displayed colour samples; all stimuli were 2D.

10.2.1 Procedure

Stimuli were presented on both presentation media, real-world papers and lights (RW) and computer-rendered and displayed images (CS). The materials are described in detail in chapter 9 including also details of how well RW and CS colours match (section 8.2.5). Tasks were completed for the RW setup under illuminant D2, Lily and Tun and for the CS under illuminant D1, D2 and Lily. Note that observers had already completed the RW task under illuminant D1 as part of the observer screening.

The procedure was identical to the one described for the participation test (section 9.2.2) and did not differ between the presentation modes.

Ten observers, who had passed the observer screening took part in this experiment. Each observer participated in each of the experimental conditions, which were six in total.

10.2.2 Results

Table 10.2.2-1 shows the hit rates for both presentation media and all illuminants, additionally the data is separated by colour groups. Note, that this table also includes results of the participation test (RW under D1). The last column provides the average hit rates across all colour groups.

presentation media	illuminant	colour groups: blue	red	yellow	Ø
RW	D1	36.7%	73.3%	50.0%	53.3%
	D2	30.0%	68.3%	73.3%	57.2%
	Lily	38.3%	76.7%	55.0%	56.7%
	Tun	31.7%	70.0%	50.0%	50.6%
CS	D1	41.7%	68.3%	55.0%	55.0%
	D2	36.7%	68.3%	50.0%	51.7%
	Lily	48.3%	70.0%	43.3%	53.9%

Table 10.2.2-1. Summary of the hit rates for the two presentation media under each of the experimental illuminants and each colour group. The highlighted column indicates the mean hit rates across all colour groups.

The data analysis was conducted in two steps. First, performances were analysed separately for the presentation media and then, a cross-media comparison was conducted.

(1) Individual presentation medium analysis

a. Presentation media RW

A two-way repeated-measures ANOVA was conducted. The factors were illuminant, which had four levels (D1, D2, Lily and Tun) and colour group, which had three levels (blue, red and yellow). All effects are reported as significant at $p < 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed for the factor colour group and the interaction illuminant/colour group. However, the assumption of sphericity had been violated for the factor illuminant, $p = 0.04$; therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

There was no main effect of illuminant. However, a significant main effect was found for colour group $F(2,18) = 27.68$, $p < 0.01$. Pairwise comparison (Bonferroni corrected) revealed that the hit rate for the blue colour group (mean hit rate = 34.2% SE 5.3) was significantly lower than for red (mean hit rate =

72.1% SE 4.0) and yellow (mean hit rate = 57.1 SE 3.2) $p < 0.01$ and $p = 0.02$, respectively. The difference between the red and the yellow colour group was not significant. The interaction was not significant.

b. Presentation media CS

A two-way repeated-measures ANOVA was conducted. The factors were illuminant, which had three levels (D1, D2 and Lily) and colour group, which had three levels (blue, red and yellow). All effects are reported as significant at $p < 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed.

There was a significant main effect of colour group $F(2,18) = 12.32$, $p < 0.01$. Pairwise comparison (Bonferroni corrected) revealed that the hit rate for the red colour group (mean hit rate = 68.9% SE 3.2) was significantly higher than for blue (mean hit rate = 42.2% SE 5.4) and yellow (mean hit rate = 49.4 SE 5.6) $p < 0.01$ and $p = 0.01$, respectively. The difference between the blue and the yellow colour group was not significant.

No significant main effect of illuminant was found. The interaction was also not significant.

(2) Cross-media comparison

A three-way repeated-measures ANOVA was conducted. The factors were presentation mode, which had two levels (RW and CS), illuminant, which had three levels (D1, D2 and Lily) and colour group, which had three levels (blue, red and yellow). All effects are reported as significant at $p < 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed.

There was a significant main effect of colour group $F(2,18) = 21.85$, $p < 0.01$. Pairwise comparison (Bonferroni corrected) revealed that the hit rate for the red colour group (mean hit rate = 70.8% SE 3.5) was significantly higher than for blue (mean hit rate = 38.6% SE 5.1) and yellow (mean hit rate = 54.4 SE 3.9) $p < 0.01$ and $p = 0.02$, respectively. The difference between the blue and the yellow colour group was not significant.

No significant main effect of presentation mode (RW-mean hit rate = 55.7% SE 3.2, and CS-mean hit rate = 53.5% SE 3.7) or illuminant (D1-mean hit rate = 54.2% SE 3.8, D2-mean hit rate = 54.4% SE 3.2, and Lily-mean hit rate = 55.3% SE 4.3) was revealed (figure 10.2.2-1). None of the interactions was significant.

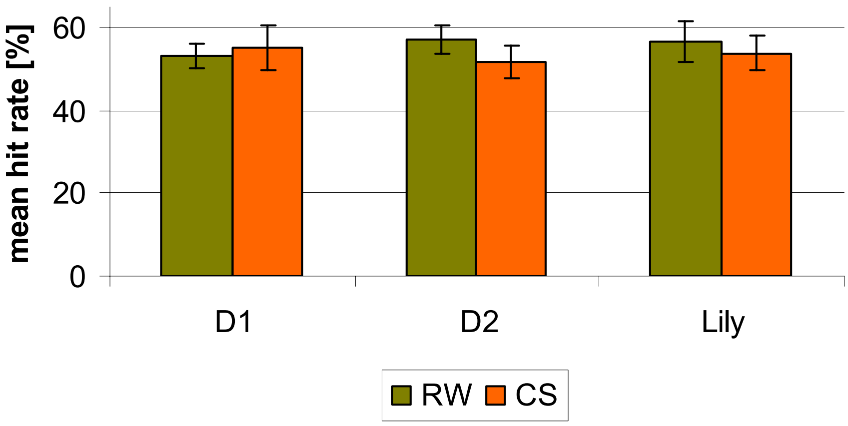


Figure 10.2.2-1. Observers' colour memory was unaffected by the illuminant and the presentation medium. Error bars indicate ±1 SE.

10.2.3 Discussion

2D stimuli were shown under four and three illuminants in the RW and CS setup, respectively, and it was revealed that observers' colour memory performance was equal under all of them. In other words, performance was independent of the illuminant. Whether the task was completed under a typical daylight (RW and CS) or a highly artificial light (RW and CS) did not change the degree of colour memory observers showed.

As expected colour group was a significant factor throughout the analyses. Although not all differences in performance for the three colour groups were significant, the overall order of the levels of difficulty (as described in section 8.1.6.3) stayed constant; red was the easiest colour group to match, followed by yellow and the blue.
The cross-media comparison showed that observers' performance was independent of the medium used to present the stimuli. In an earlier study Kaufmann and O'Neill (1993) compared CRT focal colours with those identified for real surface colours. Some of the focal colours did not coincide between the media; however, this deviation was mainly due to technical limitations of the CRT. Overall, the authors concluded that CRT and real surface focal colours are comparable. Granzier et al. (2009a) studied simultaneous colour matching within and between different presentation media. One of their main findings was that matching performance is not affected by the presentation media i.e. performance is independent of the media as long as stimulus presentation and matching takes place within the same medium. Also Johnston and Curran (1996) and Tatler et al. (2005) showed that computer-generated and displayed stimuli are valid substitutions for real object stimuli and this statement can be confirmed by the findings of the present experiment.

Within all the conducted colour memory experiments ((a) participation test, (b) colour memory for delayed matching and (c) colour memory for 2D stimuli under different illuminants and presentation media) observers selected the correct colour swatch in 50% to 60% of the trials. Whatever modification was applied on the procedure it did not affect observers' ability to recall a learned colour, which highlights the robustness of colour memory.

10.3 Shifts in colour memory

It has been shown that colour memory is not perfect and that remembered colours shift with respect to the original ones. Considering a wide range of literature on this topic it becomes evident that the direction[1] and the magnitude of these shifts are not generally predictable. Shifts seem to follow certain rules but only under very specific experimental conditions.

During the process of screening in the present study, observers completed a memory task (participation test). They learned target colours and were asked to match them immediately after by selecting a colour swatch from a given array. The motivation for the present analysis was to find out (a) which swatches were chosen when the original target colour was remembered erroneously and (b) whether the selection of alternatives followed a consistent pattern.

For (b) two different analyses were conducted. One was based on the attributes (hue, chroma and lightness) of the chosen alternatives and the other one on colour differences between the alternatives and the target colours. Shifts in colour memory are usually reported in terms of hue, chroma/saturation and lightness. Hence, the first analysis (attribute-analysis) was conducted to be able to compare the results of this memory task with the findings of other researchers. However, this analysis did not reveal a systematic pattern of erroneous matching and therefore the second analysis (colour-differences-analysis) was conducted.

[1] The remembered colour can shift in hue, saturation and lightness, the three directions/attributes of colour.

10.3.1 Data set

From the 60 observers who passed the participation test, 55 successfully completed further experiments. The results from the participation test of these 55 observers are included in this analysis. Note, ten of these observers completed a 2D as well as a 3D task.

During the participation test each target colour was shown three times to each observer and thus, the total number of trials per target colour is 195 ($65_{observer}$ x $3_{trials/observer}$); 1170 trials were completed overall (195_{trials} x $6_{target\ colours}$). The analysis is based on data from the participation test because it is the largest set of available data on a single colour memory experiment.

10.3.2 Analysis

(a) Perfect matching and the most frequently chosen alternatives

The data from the 2D and 3D task is analysed together because there was no difference in either performance between these tasks (see section 10.1.2).

The hit rates for individual target colours varied between 30.3% and 79.5%; overall the target colours were remembered accurately in 52.8% of the trials (table 10.3.2-1). Each target colour was surrounded by 15 alternative swatches but not all of these were considered by the observers. The total number of considered alternatives was as follows: for B16 eleven alternatives were considered, for B9 eleven, for R10 seven, for R8 ten, for Y10 eight and for Y8 ten. An inspection of the hit rates of each of the picked alternatives revealed that only three were generally selected with certain frequency. The three most frequently chosen alternatives of each target colour and their hit rates are listed and named in table 10.3.2-1. For the two target colours with the lowest hit rates (B16 and Y10) the three most frequently chosen alternatives had together higher hit rates than the actual target colours (Σ_1). In the last row of this table the sums of the hit rates for each of the target colours and its three most frequently chosen alternatives is presented (Σ_2). Whenever the hit rate for a target colour was low, then higher hit rates were observed for the three most frequently chosen alternatives. Thus, observers did not pick alternatives arbitrarily but were rather consistent in the mistakes they made. Overall, by considering the actual target colour and its three most frequently chosen alternatives the vast majority of matches are included in this analysis (between 81.5% and 94.4%; see last row of table 10.3.2-1).

hit rates [%]	**B16**	**B9**	**R10**	**R8**	**Y10**	**Y8**
tc	30.3	43.1	79.5	55.9	39.5	68.7
1st alt	28.2 $_{B12}$	19.0 $_{B3}$	7.2 $_{R9}$	20.5 $_{R15}$	36.4 $_{Y16}$	10.8 $_{Y9}$
2nd alt	15.4 $_{B3}$	9.7 $_{B12}$	4.6 $_{R16}$	15.4 $_{R9}$	12.3 $_{Y15}$	9.2 $_{Y7}$
3rd alt	9.2 $_{B14}$	9.7 $_{B16}$	3.1 $_{R4}$	2.6 $_{R2}$	6.2 $_{Y9}$	4.1 $_{Y2}$
Σ_1: 1st to 3rd alt	52.8	38.4	14.9	38.5	54.9	24.1
Σ_2: Σ_1 + tc	83.1	81.5	94.4	94.4	94.4	92.8

Table 10.3.2-1. The hit rates of the six target colours (tc) are in the first row. The hit rates and the labels of the most frequently (1st alt), the second (2nd alt) and third (3rd alt) most frequently chosen alternatives are listed in the second, third and fourth row, respectively. The subscript indicates the label of the alternative swatches. The fifth row shows the sums of the hit rates of the three most frequently chosen alternatives and the sums listed in the last row include the hit rates of the target colours. All hit rates are in percentage.

(b1) Attribute-analysis

This analysis was conducted using the CIE L*a*b* colour space formulae (Wyszecki & Stiles, 2000). First, all alternative swatches were classified with respect to their target colour. This means that they were classified as being either lighter or darker, higher or lower in chroma and different or the same in hue than the target colour.

Lightness difference was defined as: $\Delta L^* = L^*_{alt} - L^*_{tc}$ (9)

Chroma difference was defined as: $\Delta C = C_{alt} - C_{tc}$ where $C = \sqrt{\left(a^*\right)^2 + \left(b^*\right)^2}$ (10)

Hue difference was defined as: $\Delta H = \sqrt{\Delta E_{ab}{}^2 - \Delta L^2 - \Delta C^2}$ (11)

where alt indicates alternative and tc target colour. The colour difference ΔE_{ab} was calculated from equation 4 (see section 3.3). Table 10.3.2-2 gives an overview of the sign convention that was used for the classification of the alternatives and figure 10.3.2-1 illustrates further details. Note, that the palettes were not designed to provide uniformly distributed alternatives for hue, saturation and lightness. Therefore this analysis might not be appropriate.

			sign		**compared with the target colour the alternative is …**
Lightness			ΔL	+	lighter
				-	darker
Chroma			ΔC	+	higher
				-	lower
Hue sign (tc)	a*	b*			
I			ΔH	+	yellower
I (yellow-red)	+	+		-	redder
II			ΔH	+	redder
II (red-blue)	+	-		-	bluer
III			ΔH	+	bluer
III (blue-green)	-	-		-	greener
IV			ΔH	+	greener
IV (green-yellow)	-	+		-	yellower

Table 10.3.2-2. Scheme according to which the alternatives were classified.

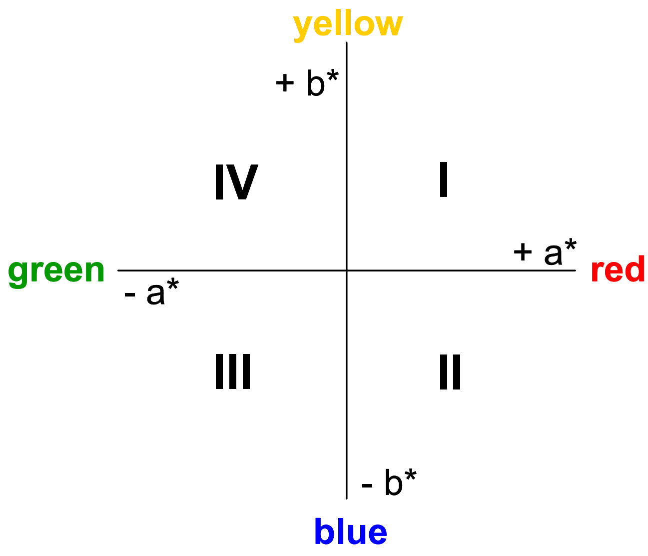


Figure 10.3.2-1. Sketch of the CIE 1976 L*a*b* colour space properties. The numbering of the quadrants runs in clockwise direction.

The classifications of all alternatives for target colour Y8 are listed in table 10.3.2-3. It shows that the 15 alternatives allowed observers to make mistakes for almost all combinations of colour attributes. For example, observers could select a swatch that was lighter, lower in chroma and greener than Y8, or select a darker colour that was higher in chroma and yellower. Here, only the classifications of the alternatives of Y8 are presented because it is representative for all other target colours.

As already reported, the three most frequently chosen alternatives for Y8 were Y9, Y7 and Y2; Y9 was darker, higher in chroma and yellower than the target colour, Y7 was darker, lower in chroma and greener and Y2 was lighter, lower in chroma and yellower. It becomes immediately clear that the question why these three colours were picked and not others cannot be answered by this analysis. It has often been reported that observers remember colours more saturated but here, two of the mentioned alternatives (Y2 and Y7) were lower in chroma, i.e. were less saturated than target colour Y8.

The complete absence of a pattern regarding the attributes of colour was equally observed for all target colours. In other words, observers were not generally misled by a certain attribute such as saturation.

Label	Lightness	Chroma	Hue
1	lighter	lower	greener
2	**lighter**	**lower**	**yellower**
3	lighter	lower	yellower
4	lighter	lower	yellower
5	lighter	lower	yellower
6	darker	lower	greener
7	**darker**	**lower**	**greener**
8	---	---	---
9	**lighter**	**higher**	**yellower**
10	lighter	higher	yellower
11	lighter	higher	yellower
12	darker	higher	yellower
13	darker	higher	yellower
14	darker	higher	greener
15	lighter	higher	yellower
16	lighter	higher	yellower

Table 10.3.2-3. All alternatives for target colour Y8 were classified in their attributes lightness, chroma and hue with respect to the target colour. The three most frequently chosen alternatives are highlighted.

(b2) Colour differences-analysis

Table 10.3.2-4 provides an overview of the three most frequently chosen alternatives and their corresponding colour differences (ΔE_{ab}) for all six target colours. For the target colours R10, R8 and Y8 the three most frequently chosen alternatives coincide with the three perceptually most similar swatches to the target colour, i.e. the alternatives with the smallest ΔE_{ab} values. For Y10 observers selected most frequently the alternatives with the smallest, second and fourth smallest ΔE_{ab} values. Note, that in this case the difference between the third and the fourth smallest ΔE_{ab} value was small (< 1 ΔE_{ab} unit). Similarly for target colour B9, alternative B3 had the smallest ΔE_{ab} value and B16 the fourth smallest. Here again, the difference between the third and the fourth smallest ΔE_{ab} value was small (< 1 ΔE_{ab} unit). B12 was frequently chosen as an alternative for B9 despite a rather large ΔE_{ab} value. Because B12 played also an important role in the selection of alternatives for target colour B16 this particularity is discussed below. Apart from alternative B12 the most frequently chosen alternatives for target colour B16 followed the same pattern as described before, B3 and B14 were the alternatives with the smallest ΔE_{ab} values.

It is striking that not necessarily the alternatives with the smallest ΔE_{ab} values were chosen most frequently. The reason therefore could lie in the non-uniformity of the CIE L*a*b* colour space (Kuehni, 2002; Mandic, Grgic, & Grgic, 2006; Pointer, 1981). A particular swatch may have a smaller ΔE_{ab} value than another but perhaps the other one is actually perceptually closer to the target colour.

Overall this analysis showed that observers' matching behaviour can be explained by perceptual colour differences between target colours and alternatives. Whenever observers did not remember a target colour accurately they selected those colour swatches that were perceptually closest to the target colour, in other words, observers committed the smallest mistakes possible from a perceptual colour difference point of view.

	target colour											
	B16		**B9**		**R10**		**R8**		**Y10**		**Y8**	
	alt	ΔE_{ab}	alt	ΔE_{ab}	alt	ΔE_{ab}	alt	ΔE_{ab}	alt	ΔE_{ab}	alt	ΔE_{ab}
1st	B12	11.5	B3	3.2	R9	17.2	R15	10.1	Y16	10.1	Y9	9.2
2nd	B3	5.8	B12	18.5	R16	18.2	R9	10.5	Y15	11.2	Y7	10.5
3rd	B14	3.6	B16	8.6	R4	13.0	R2	10.0	Y9	9.3	Y2	8.2

Table 10.3.2-4. List of the three most frequently chosen alternatives and their ΔE_{ab} values for each target colour. From top to bottom: most frequently chosen alternative, second most frequent and third most frequent.

What was so special about B12?

The swatch B12 was frequently chosen as an alternative for both blue target colours despite its rather large colour difference. The reason therefore may lie in the construction of the blue colour group. As described in section 8.1.6 the colours of the blue group lay perceptually close together. Within a colourful environment, as they were presented in during the learning phase, B16 and B9 (NCS notation S1040 B10G and S1040 B40G, respectively) appeared as almost perfectly blue; however, they were not. When observers were presented with a blue test palette they were confronted with numerous very similar looking swatches. It seems that they decided then to choose B12 because of the combination of two reasons: (1) B12 was the most bluish swatch of the selection and (2) it was highly saturated. In this particular case it appears that observers favoured a highly saturated alternative. However, as other considered blue alternatives were lower in saturation than the target colours, no general valid rule can be conveyed and this does therefore not contradict the findings reported in the attribute-analysis.

10.3.3 Summary

The shift between an original and the remembered colour is commonly described using the attributes hue, saturation and lightness. The overwhelming inconsistency of described shifts in literature, which was also demonstrated in the attribute-analysis, indicates that this method may not be appropriate. However, the analysis of colour differences led to a satisfying explanation regarding colour shifts in memory, at the least under the given experimental conditions. When the original/target colour could not be identified within the presented array (hence, the colour had shifted in memory) observers selected those colours that were perceptually most similar to the target colour.

Colour shifts have been analysed with data from the participation test because this was the largest set of data collected for a single colour memory condition. However, a colour differences-analysis has also been carried out for the colour memory experiment for delayed matching and for the colour memory experiment for 2D stimuli under different illuminants and presentation media. Overall the same tendencies were found as described above. When observers did not recall a target colour perfectly they selected alternatives that were perceptually most similar to the target colour. This kind of matching behaviour was found to occur under all experimental conditions. In other words, this consistency was observed under all illuminants and regardless of the presentation medium. However, the results were rather noisy because of the small number of trials per target colour per condition.

Generally, the colour differences-analysis appears to be appropriate to classify and quantify colour memory shifts when a matching by selection task is performed, because each colour sample of a presented array can be expressed in terms of colour difference with respect to the test colour.

11 Colour constancy for real surface and computer-simulated stimuli

This chapter consists of six sections. In section 1 colour constancy is assessed for 2D and 3D objects under various illuminant change conditions by using exclusively real stimuli. Target colours are learned either as flat swatches from a palette or as objects that are part of a 3D scene. The same experiment is then repeated using computer-generated and displayed images of exactly the same stimuli (section 2). The results of these experiments are directly compared in section 3. Section 4 describes an experiment in which observers' colour constancy performance with and without adaptation is contrasted. In section 5 a colour constancy index, that incorporates colour memory, is introduced and applied to experimental data. While analyses in sections 1 to 5 consider exclusively hit rates the last section (6) is about the selection of alternative swatches.

11.1 Experiment 1 – Testing colour constancy for 2D and 3D real objects under various illuminant change conditions

It has been argued that the dimensionality of stimuli in the environment in which they are presented is instrumental in achieving good colour constancy performance (Kraft & Brainard, 1999). However, nobody has actually compared 2D with 3D real-world setups, in which appearance changes correspond to true surface reflectance changes. As Kraft and Brainard's work (1999) has indicated providing a wide range of visual cues improves colour constancy, therefore it is proposed that learning a colour in a cue-rich environment (3D), where the target colour is represented by a 3D object will lead to better colour constancy performance than in a cue-poor environment (2D), where the target colour is a flat swatch.

Using real surface changes requires abandoning the frequently used matching by adjustment paradigm and employing a task, which resembles an everyday life situation, namely the selection of a colour swatch from alternatives.

Previous studies comparing colour constancy across diverse illuminant changes have drawn an inconclusive picture. It seems plausible to hypothesize that human colour constancy is best for illuminant changes that are experienced naturally; e.g. variations in natural daylight, as this is the light in which the visual system has evolved (see Shepard, 1992).

Brainard (1998) used two illuminants close to and a further nine off the blackbody locus, and concluded from his results that the visual system compensates equally well for illumination changes on and off the blackbody locus. However, Ruttiger et al. (Ruttiger, Mayser, Serey, & Sharpe, 2001) found actually higher colour constancy for red-green illuminant changes than for daylight changes. Delahunt and Brainard (2004) could not report a clear advantage of daylight illuminant changes over other illuminant changes. In their study, the highest colour constancy was found indeed for one of the two illuminants chosen from the blackbody locus but the second highest constancy was achieved for a green illuminant change. Daugirdiene et al. (2006) also compared colour constancy levels for on- and off-blackbody locus illuminants and did not find superior constancy for the on-blackbody locus illuminants. To summarize, there is no evidence that the visual system compensates more effectively for typical than for atypical illuminant changes.

It is here argued that the previously reported absence of benefit for natural/typical lighting conditions could be due to the kind of tasks and unnatural setups that have been used.

In the present experiment observers' colour constancy is tested for two conditions, where target colours are either learned from 2D swatches or from geometrical 3D objects. In order to ensure the generality of findings colour constancy is assessed under typical (shift on the blackbody locus) and atypical (shifts away and toward the blackbody locus) illuminant changes.

11.1.1 Procedure

This experiment took place in the lighting booth using the real-world setup (for details see section 8.1).

Observers adapted initially for 2 min to the prevailing illuminant by viewing the white palette. Observers learned the target colour either as a 2D swatch on a palette or as objects made of the target colours and embedded into the 3D setup. After a 20 sec learning period, the illuminant changed and observers adapted for 2 min to the new illuminant by viewing the white palette. Then, a test palette was shown and observers made their selection (figure 11.1.1-1). Each task consisted of 18 trials in which each target colour was presented three times. No time limit was set for selection and no feedback was given.

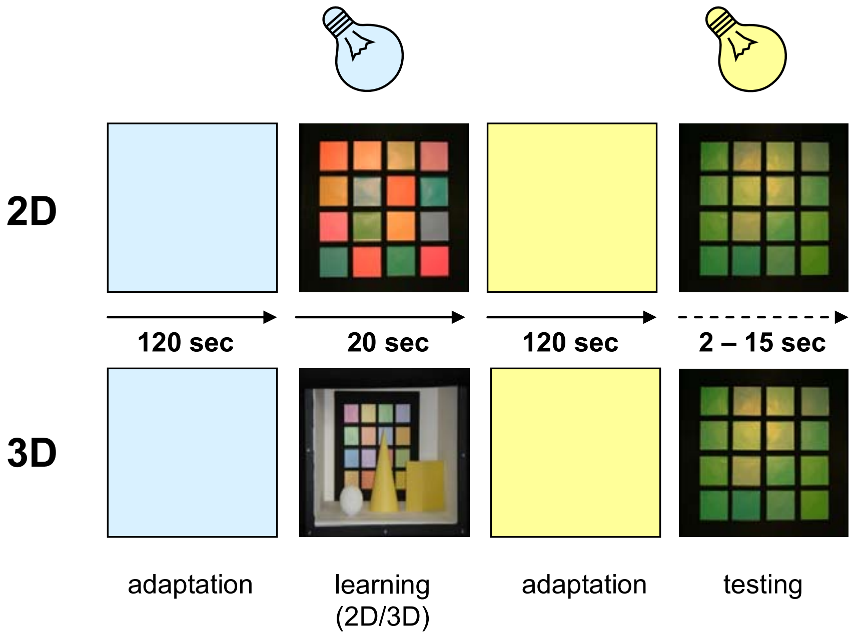


Figure 11.1.1-1. Trial sequence and timings for the 2D and 3D conditions.

Illuminant change conditions

Figure 11.1.1-2 provides an overview of the illuminant change conditions for which colour constancy was tested. As mentioned earlier, colour constancy was tested under various illuminant change conditions to ensure the generality of findings. There were four experimental illuminants; colour constancy however was solely assessed for illuminant changes between D1 and Tun, and D2 and Lily (and not for all possible illuminant change conditions) because of time management issues.

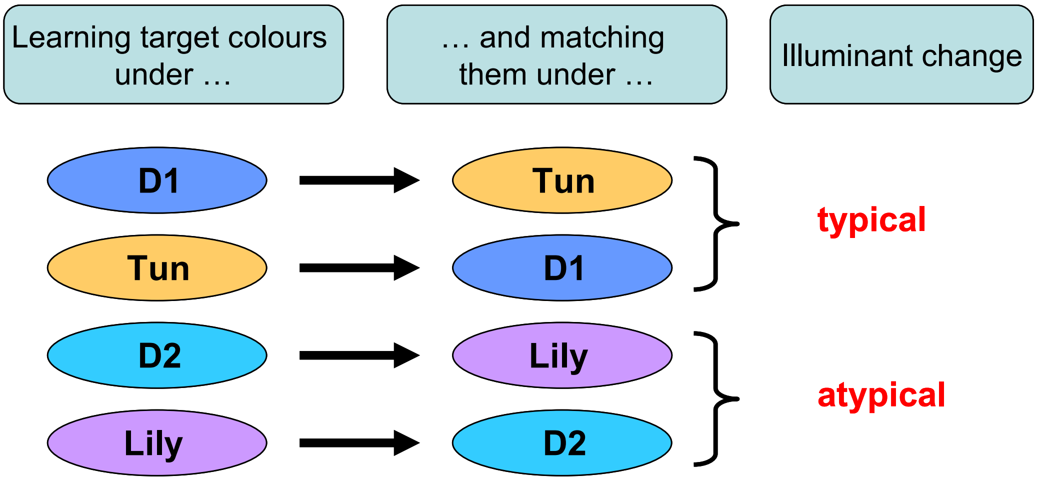


Figure 11.1.1-2. Colour constancy is tested for four different illuminant change conditions.

Observers

Seven observers were assigned to each of the four illuminant change conditions. Each of the 28 observers completed two tasks (2D and 3D) for the assigned illuminant change condition.

11.1.2 Results

Table 11.1.2-1 shows a summary of the hit rates for the two learning environments and the four illuminant change conditions tested. The data have also been separated for the three colour groups.

	2D				3D			
illuminant change condition	**blue**	**red**	**yellow**	**Ø**	**blue**	**red**	**yellow**	**Ø**
D1 to Tun	9.5%	28.4%	28.5%	22.1%	16.5%	52.0%	47.5%	38.7%
Tun to D1	42.6%	11.9%	23.8%	26.1%	23.8%	19.0%	45.2%	29.3%
D2 to Lily	28.6%	59.5%	33.3%	40.5%	45.2%	40.5%	28.6%	38.1%
Lily to D2	11.9%	26.2%	21.4%	19.8%	19.0%	42.9%	33.3%	31.7%
Ø	23.2%	31.5%	26.8%	**27.2%**	26.1%	38.6%	38.7%	**34.5%**

Table 11.1.2-1. Summary of hit rates for each colour group and illuminant change condition, separated whether a target colour was learned in 2D or in 3D. The highlighted row and columns indicate average hit rates across all illuminant change conditions and colour groups, respectively.

A three-way mixed ANOVA with two within-subject factors and one between-subject factor was conducted to analyse the data. The within-subject factor learning environment had two levels, 2D and 3D. The other factor was colour group and had three levels (blue, red, and yellow). The between-subject factor illuminant change condition had four levels: D1 to Tun, Tun to D1, D2 to Lily, and Lily to D2. The significance level was set at $p \le 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed.

There was a significant main effect of learning environment $F(1,24) = 5.71$, $p = 0.025$. Observers' performance improved significantly when target colours were learned as part of a 3D setup (2D-mean hit rate = 27.2% SE 2.2, and 3D-mean hit rate = 34.5% SE 2.9) (figure 11.1.2-1). Another significant main effect was found for colour group

F(2,48) = 3.19, p = 0.05, with 24.7% (SE 2.9) for the blue, 35.1% (SE 3.4) for the red, and 32.7% (SE 3.3) for the yellow colour group. Tukey's LSD post-hoc test revealed that only the difference between the blue and the red mean hit rates was significant (p = 0.034).

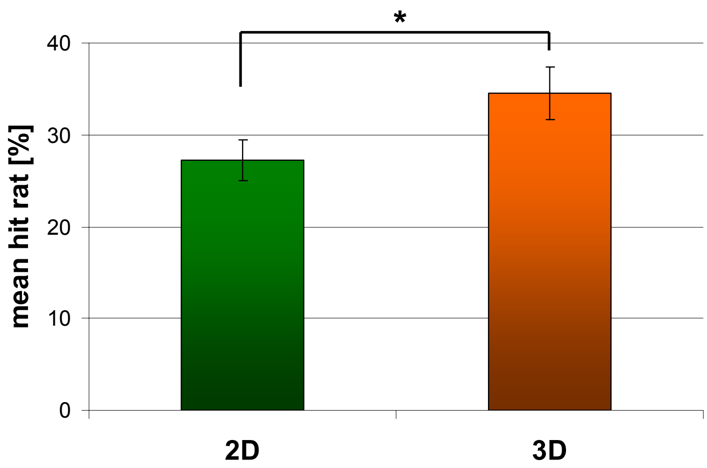


Figure 11.1.2-1. Significant main effect of learning environment. Error bars indicate ± 1 SE and * indicates a significant difference at $p < 0.05$.

No significant main effect of illuminant change condition was found. There was neither a significant interaction between learning environment and illuminant change condition, nor learning environment and colour group. However, there was a significant interaction between colour group and illuminant change condition F(6,48) = 3.69, p = 0.004 (figure 11.1.2-2). The interaction between all three factors (learning environment, illuminant change condition and colour group) was also significant F(6,48) = 3.52, p = 0.006 (figure 11.1.2-3).

A further three-way mixed ANOVA with two within-subject factors and one between-subject factor was conducted to explicitly study whether the nature of an illuminant change may influence the results, i.e. typical vs. atypical illuminant changes. The within-subject factors were the same as in the mixed ANOVA reported above: learning environment and colour group. For the between-subject factor, the data was regrouped into two illuminant changes (typical and atypical). The dataset of the typical illuminant change consisted of D1 to Tun and Tun to D1, whereas the atypical illuminant change contained the data of D2 to Lily and Lily to D2.

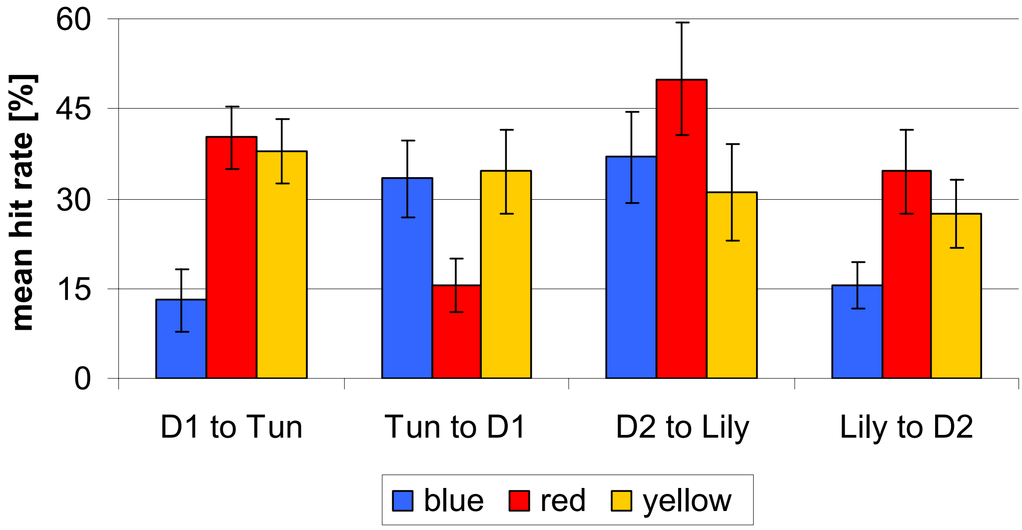


Figure 11.1.2-2. This figure shows the interaction between colour group and illuminant change conditions. Error bars indicate ± 1 SE.

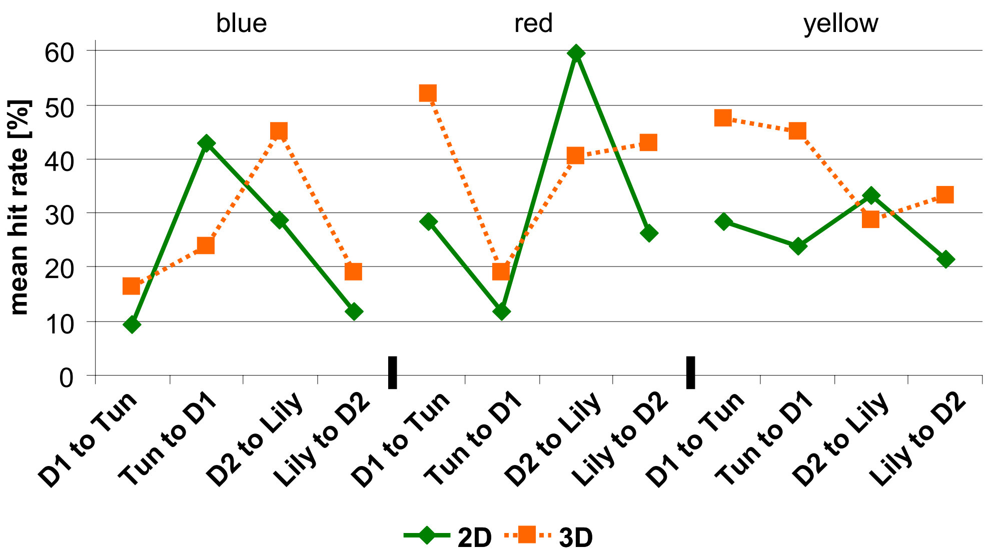


Figure 11.1.2-3. This figure shows the interaction between learning environment (2D and 3D), illuminant change condition (D1 to Tun, Tun to D1, D2 to Lily and Lily to D2) and colour group (blue, red and yellow).

As before a main effect of learning environment F(1,26) = 5.10, p = 0.033 was found. However, neither colour group nor illuminant change (figure 11.1.2-4) were significant factors F(2,52) = 2.59, p = 0.084 and F(1,26) = 0.63, p = 0.43, respectively. Also none of the interactions were significant.

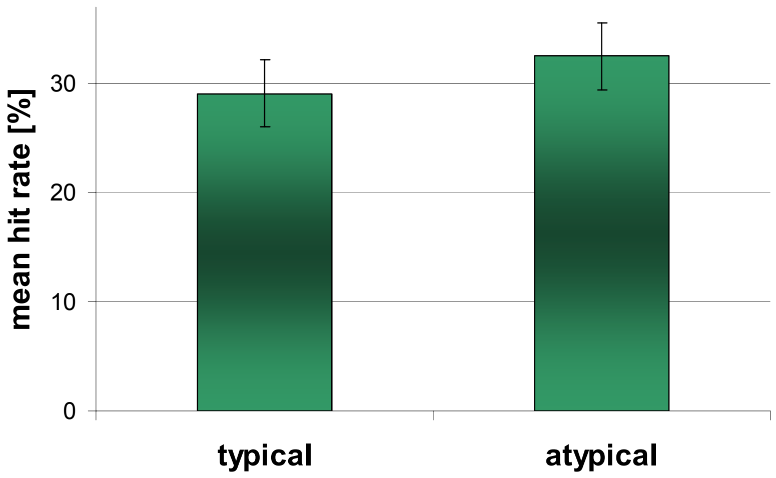


Figure 11.1.2-4. The difference in performance for typical and atypical illuminant changes did not reach a significant level. Error bars indicate ± 1 SE.

11.1.3 Discussion

In this experiment colour constancy was tested for various illuminant change conditions and observers' performance for 2D and 3D stimuli was contrasted.

Learning environment

It was revealed that observers achieved higher levels of colour constancy when target colours were learned as part of a 3D scene than as 2D swatches. This effect did not depend on the illuminant change conditions. Individual contributions of specific cues were not explicitly tested, although it was evident that the 3D scene contained more illuminant cues than the 2D palette setup.

Despite decreasing hit rates form 2D to 3D in the illuminant change condition D2 to Lily, and only a slight increase in the illuminant change condition Tun to D1 the difference between the two learning environments was significant. This effect was strengthened by the absence of two-way interactions with the factor learning environment. In other words, the difference between hit rates in the illuminant change conditions D2 to Lily was not significant and hence, observers' performance was similar for 2D and 3D.

More specifically, the hit rate for 3D was comparable with those of the other illuminant change conditions. However, observers performed for some reason particularly well in the 2D condition, almost twice as good as for other illuminant change conditions. As a consequence of the observer screening, only observers with roughly similar colour memory took part in the experiment and observers were assigned arbitrarily to one of illuminant change conditions to obtain balanced groups. Hence, this outstanding performance is unlikely to be due to extremely capable observers being concentrated in on illuminant change group.

In the experiment in which colour memory was assessed under various illuminants (section 11.2) it was shown that colour memory is independent of the illuminant. Therefore, observers' performance in the D2 to Lily illuminant change condition was also not due to superior memorability of colours under illuminant D2. The argument that matching under illuminant Lily was particularly easy does not hold either. If this would have been the case, observers had also shown extremely good performance for the 3D learning environment. However, here there performance was good but not outstanding. To conclude, it remains unclear why matching was apparently so much easier in the 2D learning environment for the D2 to Lily than for any other illuminant change condition.

The target colour was presented either by the cone and the box in the 3D scene or by a swatch on a 2D palette. Thus for the 3D learning environment not only the scene is 3D but also the object from which the test colour is learned. The 2D test patch was not simply embedded in a 3D environment (this would have been similar to the setup used in Kraft & Brainard (1999)) but the test colour was displayed by a 3D object. In previous work, de Almeida Fiadeiro, Nascimento, & Foster (2002) showed that observers were equally good at detecting a material change in real 3D scenes as in their 2D planar projections. The authors go on to interpret this as evidence that 3D cues play a limited role in surface colour perception. In the present experiment it was ensured that the test object provided cues that were fully consistent with a 3D object in a 3D scene including shading, shadows and mutual illumination. If no difference had been found between the two experimental setups then it would be evident that dimensionality (2D versus 3D) had no effect on colour constancy. But that was not the case. It was established that, for the given experimental conditions, colour constancy improves when the target colour is learned as part of a 3D object.

Stimulus size and visual field size

In total, the cone and the box subtended an area that was approximately nine times larger than the area of a swatch. It could be argued that the larger area of the target colour in the 3D scene might have led to higher levels of colour constancy. According to the available literature the accuracy of colour memory stabilizes for stimulus sizes equal or larger than 1 deg (Abramov & Gordon, 2005; Nerger, Volbrecht, & Ayde, 1995). Considering also that the rod-free area of the fovea extends over an angle of 1.7 deg (Wandell, 1995) it can be concluded that the difference in stimulus size was not responsible for the results because the swatches, the smallest stimuli, already subtended a visual angle of 1.8 deg. On a related note, it has been shown that colour constancy increases with size of adaptation field (Hansen et al., 2007). Hansen et al. (2007) compared a large adapting field size (64 x 45 deg) with a smaller one (10 x 8 deg), which differed in area by a factor of 36. The scenes used in this study (2D: 11 x 8 deg and 3D: 16 x 14 deg) were similar in size to the smaller adapting field size and the difference in area between these two scenes (2D and 3D) corresponds to a factor of

2.54. In the present experiment adaptation always took place by viewing the white palette (11 x 8 deg), it is suggested that this combined with the negligible change in field size between 2D and 3D rules out the possibility that the 3D advantage was due to an increase in visual field size.

Colour groups

It was not surprising that the factor colour group was significant and that the red colour group was overall easiest to match, followed by the yellow and blue colour group. However, this overall finding is slightly misleading as the level of difficulty depended on the illuminant change condition (figure 11.1.2-2). The theoretically and empirically established order of the levels of difficulty of the three colour groups (from easiest to most difficult: red, yellow, blue) was retained only for the illuminant change condition D1 to Tun and Lily to D2 but not for Tun to D1 and D2 to Lily. For the illuminant change condition Tun to D1 the performance for the blue and yellow colour groups was similar, whereas the red colour group was much more difficult. For the illuminant change condition D2 to Lily the red colour group was again easiest to match followed by the blue and yellow group.

This finding cannot be explained by colour differences as they stayed approximately constant under the different illuminants. Furthermore, it has been shown in the colour memory experiment (section 10.2) that the level of difficulty was independent of the illuminant. It therefore remains unclear why the level of difficulty altered under two illuminant change conditions.

Illuminant change condition

The question of whether the visual system compensates more effectively for illuminant changes that are part of daily life (typical changes) than for rare (atypical) changes has already been addressed in earlier studies (e.g. Brainard, 1998; Daugirdiene et al., 2006; Delahunt & Brainard, 2004; Hansen et al., 2007; Ruttiger et al., 2001; Schultz et al., 2006). In none of these studies have researches found an improved level of colour constancy when testing under a typical illuminant change in comparison to an atypical one. It has earlier been argued that this might have been due to the kind of tasks and setups used in these studies. However, the results of the present experiment are in line with these previous findings despite a natural task and the use of a real-world setup. No significant effect of illuminant change condition was found, nor was there an effect when the four illuminant change conditions were grouped into typical and atypical illuminant change conditions. Hence, the visual system does not seem to compensate more effectively for frequently experienced illuminant changes.

11.1.4 Summary

The initial hypothesis that learning a target colour in a cue-rich (3D) environment will lead to better colour constancy performance than learning it in a cue-poor environment (2D) has been confirmed. Colour constancy, similar to colour memory, appears to be independent of the illuminants under which it is tested.

11.2 Experiment 2 – Testing colour constancy using computer-generated and displayed 2D and 3D stimuli

Up to this point all colour constancy tasks were conducted using exclusively RW stimuli. The most significant finding so far is that colour constancy improves for 3D objects. As outlined in chapter 7 and already discussed in section 11.2 it needs to be clarified whether computer-generated and displayed stimuli are valid substitutions for real-object stimuli in vision research. Therefore, two factors were investigated in the present experiment: (a) whether the effect of improved colour constancy for 3D object could be reproduced by using computer-simulated stimuli and (b) whether observers' performance would be similar for RW and CS stimuli.

11.2.1 Procedure

Exclusively computer-generated and displayed stimuli were presented in this experiment; respective materials are described in section 8.2.

The procedure was identical to the one described in section 11.1.1 with the only difference that all stimuli were computer-generated and displayed instead of being made of real surfaces. Colour constancy was tested for the illuminant change conditions D2 to Lily and Lily to D2.

Seven observers were assigned to each illuminant change condition. Each of the 14 observers completed two tasks (2D and 3D) for the assigned condition.

11.2.2 Results

Table 11.2.2-1 provides a summary of the hit rates for the two learning environments and both illuminant change conditions. The data have also been separated for the three colour groups.

A three-way mixed ANOVA with two within-subject factors and one between-subject factor was conducted. The within-subject factor learning environment had two levels, 2D and 3D. The other factor was colour group and had three levels (blue, red, and yellow). The between-subject factor illuminant change condition had two levels (D2 to Lily and Lily to D2). The significance level was set at $p < 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed.

illuminant change condition	2D blue	2D red	2D yellow	2D Ø	3D blue	3D red	3D yellow	3D Ø
D2 to Lily	4.8%	11.9%	2.4%	6.4%	16.7%	28.6%	16.7%	20.6%
Lily to D2	11.9%	16.7%	2.4%	10.3%	7.1%	23.8%	16.7%	15.9%
Ø	8.4%	14.3%	2.4%	**8.4%**	11.9%	26.2%	16.7%	**18.3%**

Table 11.2.2-1. Summary of hit rates for each colour group and illuminant change condition, separated whether a target colour was learned in 2D or in 3D. The highlighted row and columns indicate average hit rates across both illuminant change conditions and all colour groups, respectively.

There was a significant main effect of learning environment $F(1,12) = 11.23$, $p = 0.006$. Observers' performance improved significantly when target colours were learned as part of the displayed 3D scene (2D-mean hit rate = 8.4% SE 1.7 and 3D-mean hit rate = 18.3% SE 2.5). Another significant main effect was found for colour group $F(2,24) = 6.75$, $p = 0.005$. Tukey's LSD post-hoc test showed that the mean hit rate of the red colour group (20.2% SE 2.4) was significantly higher than for the blue (10.1% SE 2.3) and yellow (9.5% SE 2.7) colour groups $p = 0.001$ and $p = 0.015$, respectively. The difference between blue and yellow was not significant.

However, there was no significant main effect of illuminant change condition (figure 11.2.2-1) and none of the interactions were significant.

11.2.3 Discussion

As before, the environment in which target colours were learned influenced significantly observers' performance, which was best when target colours were learned as part of the 3D scene. The 3D scene was rendered to be as similar as possible to the real scene, i.e. potential cues to the illuminant such as shadows and mutual illumination were preserved. However, the replicated scene differed from the real as all surfaces were rendered as being Lambertian although the real surfaces were slightly glossy. A further difference between the RW and CS scene emerged from the condition under which the scenes were viewed. The rendered images were presented binocularly but not stereoscopically, in other words, retinal disparity as a cue to depth was not longer available. Although the original three dimensions were still perceived because of monocular depth cues, a cue conflict occurred between them and the oculomotor cues to depth. While the monocular cues indicated that the viewed scene was 3D the state of accommodation and convergence was interpreted by the visual system as arsing from a 2D scene. If visual cues to depth contribute to colour constancy than this might have affected observers' performance. Despite these limitations better colour constancy was achieved for the 3D scene than for the 2D palettes.

Colour group was again a significant factor. While red was still the easiest group to match observers found it equally difficult to match the blue and yellow target colours. In contrast to the results reported for experiment 1 (section 11.1.2) the level of difficulty was independent of the illuminant change condition. Although this finding contradicts that from the previous experiment the results were in agreement regarding observers' performance for different illuminant change conditions. Whether observers learned a target colour under D2 and matched it under Lily or learned it under Lily and matched it under D2 did not affect their performance.

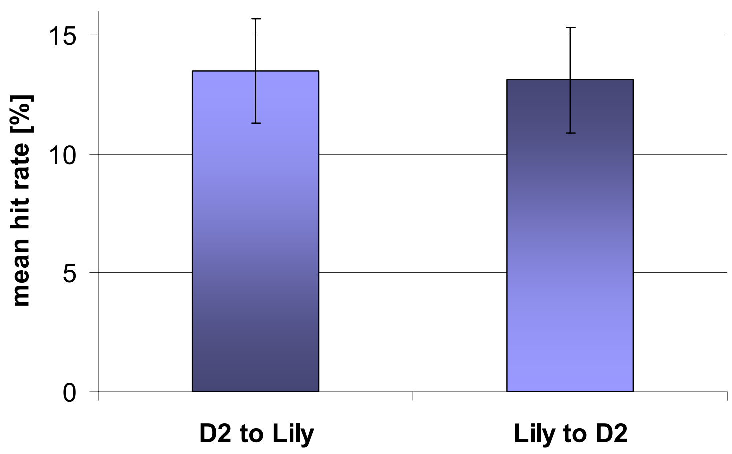


Figure 11.2.2-1. There was no significant difference in performance between the illuminant change conditions D2 to Lily and Lily to D2. Error bars indicate ± 1 SE.

11.3 Comparing observers' performance for RW and CS stimuli

Experiment 1 (section 11.1) was conducted using exclusively RW stimuli and in experiment 2 (section 11.2) solely CS were used. The results of these experiments are directly compared in this section. To facilitate the comparison table 11.3-1 shows the hit rates for the illuminant change conditions D2 to Lily and Lily to D2 for both presentation media. The data have also been separated for learning environments and colour groups.

		2D			3D		
presentation medium	**illuminant change condition**	**blue**	**red**	**yellow**	**blue**	**red**	**yellow**
RW	D2 to Lily	28.6%	59.5%	33.3%	45.2%	40.5%	28.6%
	Lily to D2	11.9%	26.2%	21.4%	19.0%	42.9%	33.3%
	Ø learning envir.			30.2%			34.9%
CS	D2 to Lily	4.8%	11.9%	2.4%	16.7%	28.6%	16.7%
	Lily to D2	11.9%	16.7%	2.4%	7.1%	23.8%	16.7%
	Ø learning envir.			8.4%			18.3%
	Ø colour group	14.3%	28.6%	14.9%	22.0%	33.9%	23.8%

Table 11.3-1. Summary of hit rates for the illuminant change condition D2 to Lily and Lily to D2 and both presentation media; the data were already shown in table 11.1.2-1 and 11.2.2-1. The highlighted numbers indicate mean hit rates.

A four-way mixed ANOVA with two within-subject factors and two between-subject factors was conducted. The within-subject factor learning environment had two levels, 2D and 3D. The other factor was colour group and had three levels; blue, red and yellow. The two between-subject factors were illuminant change condition, which had two levels (D2 to Lily and Lily to D2), and presentation medium, which had also two levels (RW and CS). The significance level was set at $p \leq 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed.

A significant main effect was found for learning environment F(1,24) = 6.97, p = 0.014. Observers achieved significantly higher hit rates when target colours were learned as part of a 3D scene (2D-mean hit rate = 19.2% SE 1.8 and 3D-mean hit rate = 26.6% SE 2.6). Another significant main effect was revealed for colour group F(2.48) = 7.31, p = 0.002. Tukey's LSD post-hoc test showed that the mean hit rate of the red colour group (31.3% SE 3.1) was significantly higher than for blue (18.2% SE 2.4) and yellow (19.3% SE 2.8) p = 0.004 and p = 0.005, respectively. The difference between blue and yellow was not significant.

From the between-subject factors only presentation medium showed a significant main effect F(1,24) = 30.25, p < 0.01. Observers' performance was significantly better when RW stimuli were used (RW-mean hit rate = 32.5% SE 2.5 and CS-mean hit rate = 13.3% SE 2.5).

None of the two-way interactions were significant and the only significant three-way interaction was between learning environment, presentation medium and illuminant change condition F(1,24) = 4.28, p = 0.049 (figure 11.3-1).

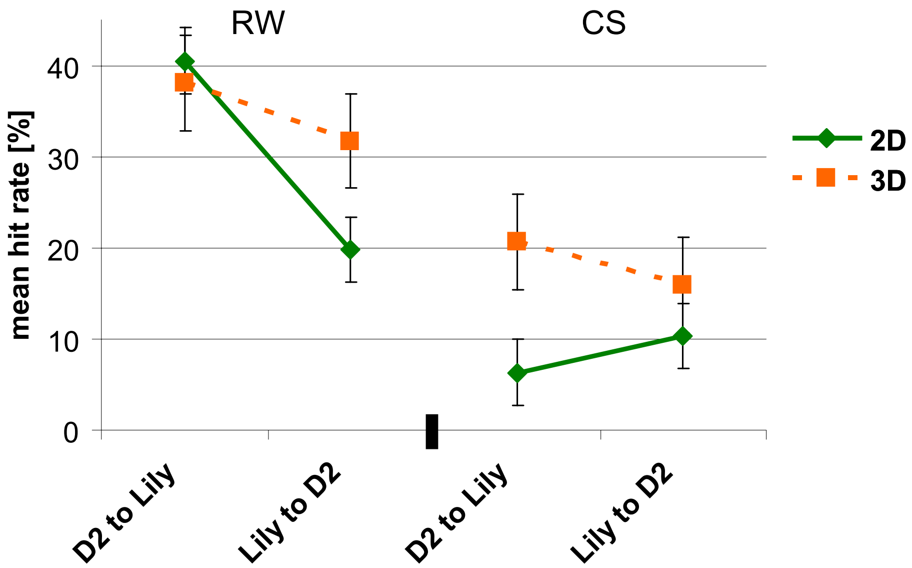


Figure 11.3-1. Three-way interaction between learning environment, illuminant change condition and presentation medium. Error bars indicate ± 1 SE.

11.3.1 Discussion

That learning a target colour from a 3D object leads to improved colour constancy was demonstrated for the RW and the CS setup, and as expected, the factor learning environment was also significant in this analysis. Similar to previous analyses, it was also revealed that observers' performance did depend on the colour group. However, the primary objective of this comparison was to find out whether observers' colour constancy performance depends on the medium that is used to present the stimuli.

Solely by looking at the hit rates it stands out that those achieved using CS stimuli are considerably lower than those for RW stimuli; subsequently this observation was confirmed by the ANOVA to be significant. Although the effect of learning environment was preserved for the displayed stimuli the overall performance dropped dramatically. In the 2D conditions the hit rate across illuminant change conditions and colour groups collapsed by almost 22% from 30.2% to 8.4% from RW to CS. A slightly lower deterioration was noticed for the 3D condition where performance dropped by 16.6% from 34.9% to 18.3%. This dramatic drop in performance may be due to the experimental procedure and to the number of available illuminant cues for both presentation media (this issue is discussed in more detail below).

All observers were naïve and had no experience with colour vision experiments. The instructions they were given were almost identical for both presentation media. The instructor explained the procedure carefully and a training trial was run when necessary, to ensure observers did understand the task. However, there was a crucial difference between the instructions for RW and CS setup. While observers were shown real palettes and spotlights in the RW setup, for CS they were told to imagine that real palettes under real illuminants were displayed. Inferring from observers' performance it appears that observers had generally a problem with imagining such a situation what affected then their performance negatively.

Earlier it was asserted that illuminant cues are crucial for the visual system to achieve colour constancy. The significant difference in performance for the two presentation media may be due to the reduced number of available illuminant cues. The loss of illuminant cues from RW to CS was larger for the 2D than the 3D condition. In the RW setup 2D condition light was reflected from the walls of the palette showroom despite covering them with black cloths. Furthermore, observers saw how the experimenter changed filters in front of the spotlight, i.e. how the illuminant was manipulated. Because illuminant cues were rare in the 2D scene this apparently minor cue may have become very important. In contrast, these kind of illuminant cues were completely absent from the CS environment. The simulated illuminant change was restricted to the dimensions of the palette and therefore, the only illuminant cue available arose from the overall shift colour swatches underwent. The rendered 3D scene contained the same illuminant cues as the RW scene. Both scenes were viewed binocularly but while the RW scene extended over three dimensions the rendered scene was only displayed on a two-dimensional monitor. In other words, monocular depth cues were preserved but binocular and oculomotor cues to depth were altered (section 11.2.3). Remember that the 3D scene was solely showed in the learning phase, subsequent matching was preformed from cue-poor 2D palettes.

This outcome is in sharp contrast to that of the colour memory experiment (section 10.2) where no differences in performance were found for RW and CS stimuli. For the colour memory task observer solely had to recognise colour swatches that were either reflective or self-luminous and otherwise identical. A colour constancy task is much more challenging for the visual system as an illuminant change has to be analysed and accounted for.

11.3.2 Summary

The effect that learning a target colour in 3D leads to higher colour constancy than learning it in 2D was found for both presentation media. However, the overall performance was severely decreased for CS and it is argued that this was not due to the quality of the CS stimuli but the conditions under which they were viewed.

11.4 Experiment 3 – Testing colour constancy with and without adaptation

Under natural conditions colour constancy occurs most often across time, i.e. surface appearance is not judged immediately after an illumination change. For example, if we go out to buy a shirt that matches a pair of trousers we have at home, we enter a shop look around and stay at least for several minutes under a constant illumination. In other words, we are normally well adapted when comparing a present surface colour with one from memory.

That adaptation plays an important role not only for colour constancy but for colour appearance in general, has already been established (e.g. Hunt, 1950; Jameson, Hurvich, & Varner, 1979; Webster, 1996; Wyszecki, 1986). Several studies have shown that chromatic adaptation to the spatial mean of a scene is almost complete (up to approximately 90%-95%) after about 1 to 2 min and that colour appearance is stable from this point onwards (Fairchild & Lennie, 1992; Fairchild & Reniff, 1995; Rinner & Gegenfurtner, 2000; Werner et al., 2000).
The present experiment was conducted to obtain a measure of the effect of adaptation under the given experimental conditions. Although emphasis was placed on the investigation of adaptation, observers' performance was also contrasted regarding the

learning environment of target colours. In other words, it was also studied whether learning target colours as part of a 3D scene leads to better colour constancy performance than learning them in a 2D environment.

11.4.1 Procedure

The colour constancy experiment with and without adaptation was conducted in the lighting booth using the real-world setup (for details see section 8.1).

This experiment consisted of four slightly different tasks. The differences lay in the presentation of the target colour (i.e. whether it was presented in 2D or in 3D) and in the selection (whether it was performed with or without adaptation).

Each task started with an initial 2 min adaptation to illuminant D1; during this time the observers looked at the white palette. Then a target colour was presented for 20 sec. After learning a target colour, the illuminant changed from D1 to Tun and in the without-adaptation condition a test palette was shown for selection immediately after. In the with-adaptation condition, the learning palette was replaced after 20 sec by the white palette, which was presented under illuminant Tun. The observers adapted to the new illuminant for 2 min before a test palette was shown for selection. Before the next trial, the observers adapted again for 2 min to the initial illuminant D1. Figure 11.4.1-1 shows a trial sequence for a 2D task with and without adaptation.

The same ten observers, who participated in the colour memory experiment for delayed matching (section 10.1) took also part in this experiment. Each observer completed all four tasks.

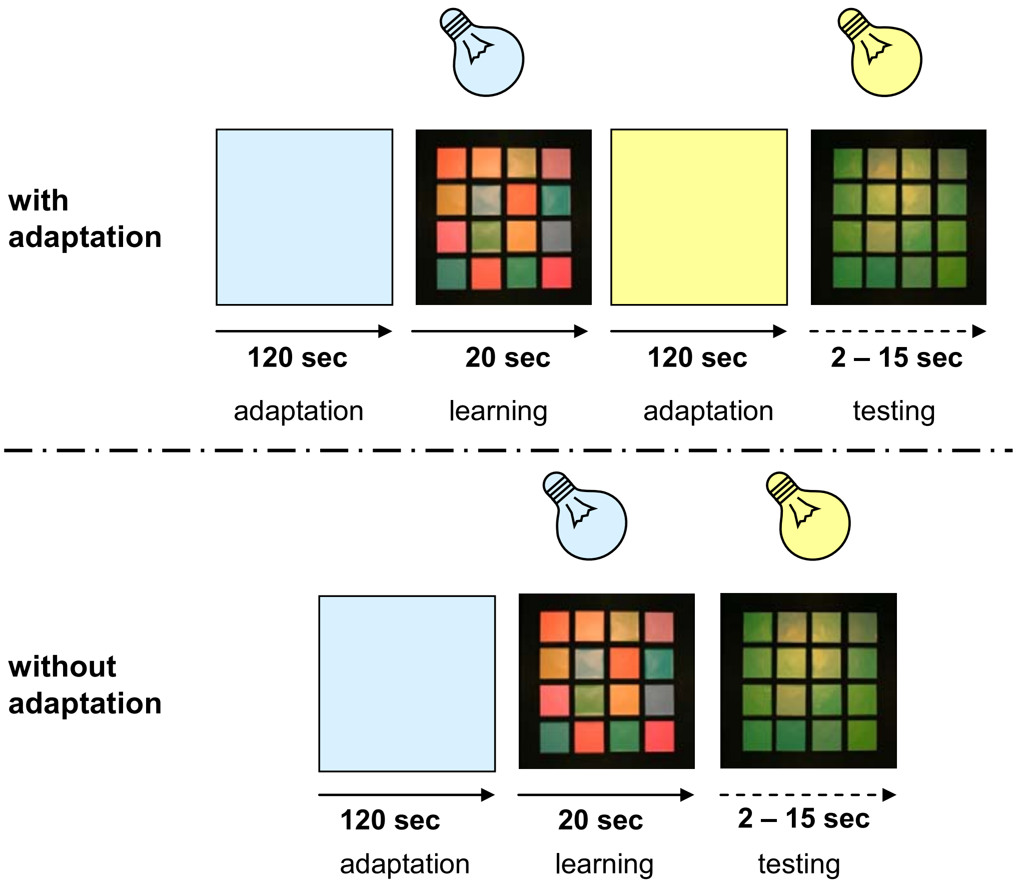


Figure 11.4.1-1. Trial sequence and timings for the colour constancy task with and without adaptation in the 2D condition. The illuminant changed from D1 to Tun.

11.4.2 Results

Table 11.4.2-1 shows the frequencies of correctly selected swatches for the two learning environments and both adaptation conditions. The data have also been separated for the three colour groups. These results have already been partially reported in section 11.1.

	2D				3D			
condition	**Blue**	**red**	**yellow**	**Ø**	**blue**	**red**	**yellow**	**Ø**
with adaptation	8.3%	29.9%	28.2%	22.2%	19.9%	49.6%	41.6%	37.0%
without adaptation	9.9%	19.9%	24.9%	18.2%	21.6%	18.3%	11.7%	17.2%
Ø	9.1%	24.9%	26.6%	**20.2%**	20.8%	34.0%	26.7%	**27.1%**

Table 11.4.2-1. Summary of the hit rates for each adaptation condition and colour group, separated whether a target colour was learned in 2D or in 3D. The highlighted row and columns indicate average hit rates across colour groups and adaptation condition, respectively.

A three-way repeated-measures ANOVA was conducted. The factors were adaptation condition, which had two levels (with and without) and learning environment, which had also two levels (2D and 3D). The third factor was colour group and had three levels (blue, red and yellow). The significance level was set at $p \leq 0.05$. Homogeneity of variance was examined using Mauchly's Test of sphericity and could be assumed.

There were significant main effects of adaptation condition $F(1,9) = 10.83$, $p = 0.009$ (with-adaptation-mean hit rate = 29.6% SE 2.4 and without-adaptation-mean hit rate = 17.7% SE 2.6) and learning environment $F(1,9) = 5.70$, $p = 0.041$ (2D-mean hit rate = 20.2% SE 2.2 and 3D-mean hit rate = 27.1% SE 2.2). A significant main effect was also revealed for colour group $F(2,18) = 5.80$, $p = 0.011$. Tukey's LSD post-hoc test revealed that the mean hit rate for the blue colour group (14.0% SE 3.0) was significantly lower than for red (29.4% SE 3.2) and yellow (26.6% SE 3.2) $p = 0.02$ and $p = 0.022$, respectively. The difference between red and yellow was not significant.

The interaction between adaptation condition and colour group was significant $F(2,18) = 5.53$, $p = 0.013$. This effect was further investigated using paired t-tests. Observers' colour constancy performance improved significantly with adaptation for the colour groups red and yellow, $p < 0.01$ and $p = 0.011$, respectively, and stayed unchanged for blue (figure 11.4.2-1).

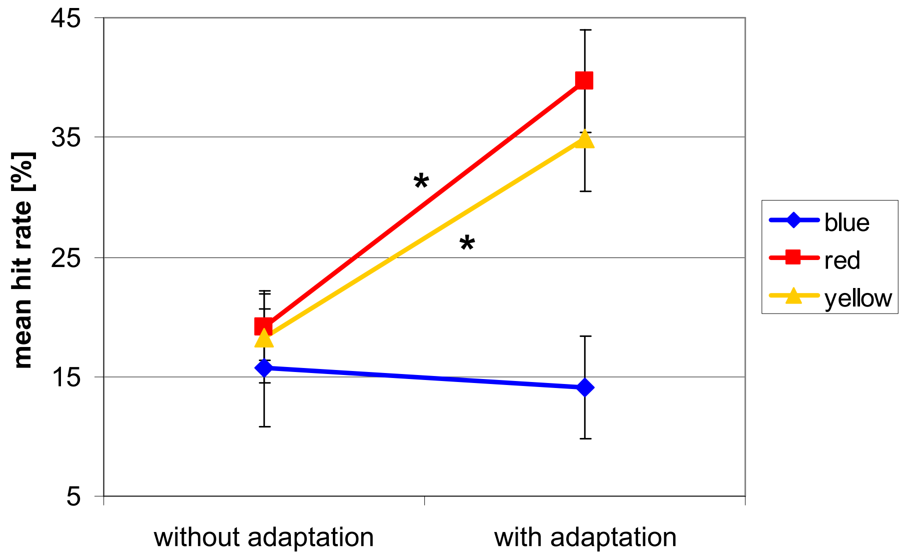


Figure 11.4.2-1. This figure shows the interaction between adaptation condition and colour group. Error bars indicate ± 1 SE and * indicates a significant difference at $p \leq 0.05$.

The interaction between adaptation condition and learning environment was also significant $F(1,9) = 14.70$, $p = 0.004$. This effect was further investigated using paired t-tests. Observers' colour constancy performance was significantly better when target colours were learned as part of the 3D setup after adaptation $p < 0.01$ (figure 11.4.2-2). Other interactions were not significant.

11.4.3 Discussion

In this experiment colour constancy was tested with and without adaptation. Tasks were conducted to obtain a measure of the effect of adaptation on observers' performance under the particular experimental conditions.

The results of the tasks without adaptation showed that the observers selected on average the correct swatch in 17.7% of the cases. When adaptation was enabled, the observers made the correct choice on average in 29.6% of the cases. In other words, colour constancy performance dropped by almost 12% when no time for adaptation was given. This drop confirmed the subjective impression of the observers, who described the task with adaptation as difficult and the task without adaptation as almost impossible. They all mentioned that all the swatches looked very similar when they selected immediately after the illuminant change.

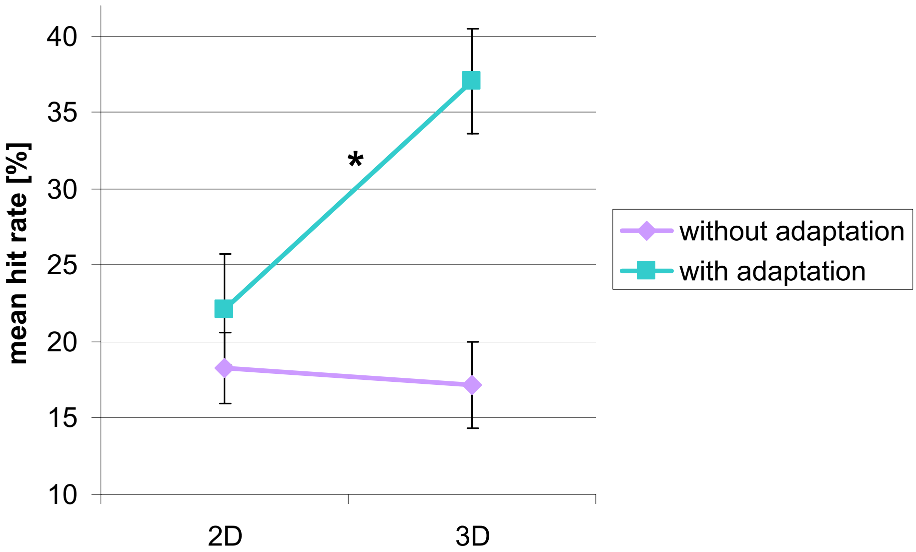


Figure 11.4.2-2. This figure shows the interaction between adaptation condition and learning environment. Error bars indicate ± 1 SE and * indicates a significant difference at $p < 0.05$.

As before learning environment was a significant factor; observers' performance improved for 3D stimuli. However, it has to be emphasised that this effect was entirely due to the performance in the with-adaptation condition. The performance in the without-adaptation condition even deteriorated when target colours were learned in 3D, although this deterioration was not significant (figure11.4.2-2). Overall, it appears that performance in the without-adaptation condition was dominated by the lack of adaptation.

The analysis showed that observers' performance was dependent on colour groups; red yielded the highest hit rate and blue the lowest. The hit rate for the yellow colour group lay between those of the other two groups. This confirmed the hypothesis developed earlier (see section 8.1.6.3). However, the interaction between adaptation condition and colour group showed that the matching of yellow and red target colours benefited from adaptation and that the blue target colour were equally difficult to match in both adaptation conditions (figure 11.4.2-1).

It was demonstrated that observers' colour constancy performance benefits significantly from adaption. However, as adaptation occurs over time colour memory becomes an important issue. The colour memory experiment for immediate and delayed matching (section 10.1) had shown that colour memory does not deteriorate over a gap of 2 min under these particular experimental conditions. Therefore it can be concluded that delayed matching in the with-adaptation condition was not compromised by a deteriorated colour memory.

11.4.4 Summary

This experiment underlined the importance of adaptation for colour constancy and allowed quantifying the effect. Consequently, all colour constancy experiments were conducted using adaptation periods of 2 min.

11.5 Colour constancy index

Previously used ratios to quantify colour constancy have been discussed in section 5.5. Under the present experimental conditions however the Brunswik ratio or any modified versions were not applicable for different reasons: (1) under no circumstances could the perceptual shift equal the physical shift, i.e. it was impossible to perform a chromaticity match, indicative of a complete lack of colour constancy, and (2) the colour constancy index calculated for any patch in the test palette would have a high value. A target colour was learned under one illuminant and when it was presented under a different one for matching, the alternative swatches on the test palettes were perceptually and colorimetrically close to the target colour. Hence, it was impossible to choose a swatch that was very different from the target colour or even to select a swatch with the same chromaticity as the target colour under the first illuminant, as such patches were not available on the palette. Regardless which alternative swatches had been selected, the colour constancy ratio would have indicated almost perfect colour constancy.

A Brunswik ratio was also not applied because it does not take into account the effect of colour memory. Recently, Ling and Hurlbert (2008) introduced a new colour constancy index, which incorporates the effect of colour memory as well as the physical shift of a test stimulus. They separated the perceptual shift into two components, a memory and a pure constancy shift. To compute the colour constancy index the memory shift is subtracted from the perceptual shift. Therefore, the index accounts only for the perceptual shift caused by the illuminant change and the physical shift of the test stimulus.

Usually, researchers limit the matches in a colour constancy task to a line in colour space, which is equivalent to the illuminant change direction. By doing this they avoid dealing with different directions in a formula that is entirely based on distances

(Brunswick ratio). Because alternative swatches in the present study are organized as a cloud, distances alone are not indicative enough. Up to now results have been presented in the form of hit rates, where only the correct instances are taken into account. The reported hit rates, especially of the colour constancy tasks, may appear low (on average about 30%), but it must be remembered that the overall colour memory hit rate was approximately 50%; thus, performance dropped by only 20%. Therefore, the colour constancy performance must be set into context with the memory performance. In order to compute a colour constancy index (CI_i), the colour constancy performance of each observer (i) was normalized by their individual colour memory performance from the participation task.

$$CI_i = \frac{HR_{Ci}}{HR_{Mi}} \tag{12}$$

HR_{Ci} is the colour constancy hit rate and HR_{Mi} the colour memory hit rate of an observer. This index is 1 if the colour constancy hit rate equals the memory hit rate, indicating that the colour constancy performance was not compromised by the observer's memory. Such index drops to zero if the hit rate for a colour group was zero irrespective of the level of colour memory that was achieved in the colour memory test. Note that this index never assumes the upper limit of colour constancy to be at 100% unlike the Brunswick ratio. Here, colour constancy is normalised only by colour memory.

11.5.1 Normalised colour constancy performance

Colour constancy indices were calculated using equation 12 and are listed in table 11.5.1-1 for the two learning environments, the two presentation media and the illuminant change conditions tested. The data have also been separated for the three colour groups. The overall level of colour constancy is 0.44 and 0.65 for 2D and 3D, respectively. When RW and CS indices are averaged separately the following mean levels of colour constancy result: RW-2D 0.58, RW-3D 0.76, CS-2D 0.17 and CS-3D 0.41.

		2D				3D			
PM	**illuminant change condition**	**blue**	**red**	**yellow**	**Ø**	**blue**	**red**	**yellow**	**Ø**
RW	D1 to Tun	0.31	0.39	0.52	0.41	0.46	0.74	0.95	0.71
	Tun to D1	1.31	0.21	0.49	0.67	0.92	0.34	1.24	0.83
	D2 to Lily	0.86	0.86	0.78	0.83	1.57	0.58	0.70	0.95
	Lily to D2	0.21	0.41	0.46	0.36	0.37	0.60	0.71	0.56
CS	D2 to Lily	0.14	0.23	0.03	0.13	0.71	0.48	0.30	0.50
	Lily to D2	0.29	0.29	0.04	0.21	0.20	0.47	0.31	0.33
	Ø	0.52	0.40	0.39	**0.44**	0.70	0.54	0.70	**0.65**

Table 11.5.1-1. Summary of colour constancy indices for the two learning environments, both presentation media (PM) and the illuminant change conditions tested. The data have also been separated for the three colour groups. The highlighted row and columns indicate mean hit rates across colour groups and illuminant change conditions, respectively. The bold and highlighted numbers indicate the mean hit rates for the 2D and 3D learning environment.

11.5.2 Colour constancy index by Ling and Hurlbert

Colour constancy indices were also computed using the index proposed by Ling and Hurlbert (2008) (for details see section 5.5):

$$CCI = 1 - \frac{\dfrac{\left(\overrightarrow{S_c} - \overrightarrow{S_m}\right) \cdot \left(-\overrightarrow{S_p}\right)}{\left\|\overrightarrow{S_p}\right\|}}{\left\|\overrightarrow{S_p}\right\|} \quad (7)$$

<u>Example</u>: Computing a colour constancy index for the following incident: matching the target colour Y8 in the Tun to D1 illuminant change condition (RW 2D); learning took place under Tun and matching under D1. The observer matched the target colour perfectly in the participation test (colour memory component; see section 10.2); the memory shift ($\overrightarrow{S_m}$) is therefore 0. The physical shift ($\overrightarrow{S_p}$) swatch Y8 undergoes is the distance in the CIE u'v' colour space between Y8 under Tun and Y8 under D1. In this example $\overrightarrow{S_p}$ = [-0.0670, -0.0155]. The observer selected Y15 and therefore the constancy shift ($\overrightarrow{S_c}$) is [0.0197, 0.0077]; resulting: CCI = 0.6954 ≈ 0.7.

Colour constancy indices were computed for each colour constancy trial and then averaged across observers. Table 11.5.2-1 provides a summary of colour constancy indices for each colour group, illuminant change condition and presentation medium. The data have also bee separated for the learning environment. The overall level of colour constancy is 0.91 and 0.96 for 2D and 3D, respectively. When RW and CS indices are averaged separately the following mean levels of colour constancy result: RW-2D 0.92, RW-3D 0.95, CS-2D 0.90 and CS-3D 0.96.

PM	illuminant change condition	2D blue	2D red	2D yellow	2D Ø	3D blue	3D red	3D yellow	3D Ø
RW	D1 to Tun	0.99	0.94	0.99	0.97	1.00	0.92	0.96	0.96
	Tun to D1	1.01	0.82	0.93	0.92	1.02	0.89	1.01	0.98
	D2 to Lily	1.02	0.86	0.92	0.93	0.98	0.83	0.89	0.90
	Lily to D2	0.90	0.77	0.91	0.86	0.99	0.92	1.01	0.98
CS	D2 to Lily	1.08	0.84	0.83	0.92	1.11	0.80	0.90	0.93
	Lily to D2	0.95	0.93	0.73	0.87	1.05	1.00	0.91	0.99
	Ø	0.99	0.86	0.89	**0.91**	1.03	0.90	0.95	**0.96**

Table 11.5.2-1. Colour constancy indices computed using the index proposed by ling and Hurlbert (2008). Summary of colour constancy indices for the two learning environments, both presentation media (PM) and the illuminant change conditions tested. The data have also been separated for the three colour groups. The highlighted row and columns indicate mean hit rates across colour groups and illuminant change conditions, respectively. The bold and highlighted numbers indicate the mean hit rates for the 2D and 3D learning environment.

11.5.3 Discussion

Normalised colour constancy performance

The indices reflect well the overall findings of the colour constancy experiments. Higher levels of constancy were achieved for the 3D learning environment than for the 2D and the dramatic drop of performance from RW to CS is also reflected by the indices. The normalisation yields a more precise representation of observers' performance than solely the evaluation of hit rates. For example, colour constancy performance decreased for the illuminant change condition D2 to Lily (RW) from 2D to 3D when considering hit rates only (see section 11.1.2). Taking observers' colour memory into account shows that for this illuminant change condition, as for all other, performance was better when target colours were learned in a 3D environment. This underlines the importance of colour memory for colour constancy, and that colour constancy cannot be evaluated without considering observers' ability to recall colours.

It stands out that colour constancy indices for the blue colour group are in many occasions higher than for the other groups and that may appear to contradict the earlier reported findings; however, this is not the case. If poor colour constancy performance is normalised by a rather poor level of colour memory, the resulting constancy index is moderate or even high. On the other hand, observers matched most successfully the red target colours in the memory task. Although performance for these target colours was for the constancy tasks better than for other target colours, hit rates were here generally much lower than for colour memory. Hence, normalising a moderate colour constancy performance by an almost perfect memory yields rather low colour constancy indices.

To summarise, the here introduced index evaluates observers' colour constancy in the context of colour memory. If colour constancy is seen as a sophisticated form of colour memory, then observers' ability to recall colour must not be ignored when determining the level of colour constancy.

Comparison between indices

Colour constancy indices were computed using the here introduced index and the one proposed by Ling and Hurlbert. It is evident that the latter indices are generally higher as the previously computed ones and show less variation. The normalised colour constancy performance indices varied between 0.03 and 1.57 and between 0.73 and 1.08 for Ling and Hurlbert's index (see table 11.5.1-1 and 11.5.2-1, respectively). Despite these differences both kinds of indices showed similar tendency: observers' performance was better for the 3D than the 2D setup, although Ling and Hurlbert's index does not reveal a significant difference in performance for real and computer-simulated stimuli. The average colour constancy index using Ling and Hurlbert's index was about 1, indicating that colour constancy was almost perfect. However, in the here reported experiments colour constancy was far from perfect. It becomes evident that the choice of index is of fundamental importance, and that the same experimental results evaluated with one or another index may lead to very different scientific conclusions.

The fundamental difference between these two kinds of colour constancy indices is, that the one introduced in this study is entirely based on performance whereas Ling and Hurlbert's is based on distances between matches in the CIE u'v' colour space. When incorporating colour memory into a colour constancy index one encounters the challenge that colour memory was assessed under a different illuminant under which then colour constancy was tested. The reference white point for the experimental colours is different in both situations. Using CIE u'v' colour space does not require the specification of a white point but it is arguable whether colour memory and colour constancy performances should be compared this way (ignoring the reference white point completely).

The normalised colour constancy performance index does not consider magnitudes and directions of perfect and imperfect matches in a specified colour space. It rather concentrates on the overall quality of performance, dividing observers' performance strictly in correct (perfect) and incorrect (imperfect) matches. This index might not be perfect as already smallest deviations from the perfect match are classified as incorrect match. However, it is generally applicable (without relying on the specification of physical and/or perceptual properties of the stimuli) as it requires solely a measure of (a) observers' ability to remember colours and (b) their colour constancy performance.

11.6 Shifts in colour constancy

Considering all colour constancy tasks reported in this chapter, observers selected in about 22% of the cases the perfectly matching swatch. For some conditions the correct swatch was picked in as few as 2.4% of the cases, in other conditions this went up to 40.5%. To summarise, in almost 80% of all matchings observers selected a swatch that was not cut from the same piece of paper as the swatches (or 3D objects) they had looked at before. In this section it is investigated whether erroneous matching was systematic. The analysis is conducted in a similar manner to the one described for shifts in colour memory (section 10.3).

11.6.1 Analysis

(a) Perfect matching and the most frequently chosen alternatives

In table AP-6 (appendix) the hit rates of all target colours and those of the three most frequently chosen alternatives are listed. The identification of the most frequently chosen alternatives is indicated by the subscript. The table comprises data from six illuminant change conditions (RW: D1 to Tun, Tun to D1, D2 to Lily and Lily to D2, and CS: D2 to Lily and Lily to D2). Table 11.6.1-1 shows a summary of averaged hit rates of the target colours and the three most frequently chosen alternatives for each illuminant change condition.

As already demonstrated in the analysis of section 10.3, by considering the actual target colour and its three most frequently chosen alternatives the majority of matches are included in the analysis. Thus, the remaining matches can be disregarded as they do not contribute significantly to the total of selected swatches. Although it appears that the analysis includes between 61.4% and 91.7% of all matches, in some incidents the hit rates of the target colour and its three most frequently chosen alternatives added up to only 42.9% (target colour B16, 3D condition, D2 to Lily, CS; see table AP-5). On the other hand, for some target colours not more than three alternatives were considered in total (e.g. target colour Y8, 2D condition, D2 to Lily, RW), in other words, the hit rates added up to 100%.

presentation medium	**illuminant change condition**	**Σ**
RW	D1 to Tun	87.2%
	Tun to D1	62.2%
	D2 to Lily	91.7%
	Lily to D2	86.3%
CS	D2 to Lily	69.9%
	Lily to D2	61.4%

Table 11.6.1-1. Summary of mean hit rates of all target colours and the three most frequently chosen alternatives.

Table 11.6.1-2 provides the number of considered alternatives for each target colour, learning environment and illuminant change condition. On average 6 alternatives were considered. The largest number of alternatives was considered for illuminant change condition Lily to D2 when stimuli were displayed (mean = 9). This was the same illuminant change condition that had the lowest overall hit rate for the target colour and the three most frequently chosen alternatives. However, it seems that there is no systematic relation between the sum of hit rates, as presented above, and the number of considered alternatives. For example, the sum of hit rates for Tun to D1 was low but on average observers did not consider more than 5 alternatives. Lily to D2 (RW) had a high sum of hit rates and the number of considered alternatives was also rather high (8).

Note that the largest number of alternatives was picked for the illuminant change condition Lily to D2 for RW and CS. No other illuminant change condition led to so much variation during the process of matching.

illuminant change condition	LE	target colours B16	B9	R10	R8	Y10	Y8	Ø
D1 to Tun – RW	2D	6	6	6	5	8	7	6
	3D	9	7	3	5	9	4	
Tun to D1 – RW	2D	4	5	3	5	4	6	5
	3D	5	6	6	6	5	3	
D2 to Lily – RW	2D	5	6	2	3	3	3	5
	3D	5	6	6	7	4	4	
Lily to D2 – RW	2D	8	9	4	7	5	7	8
	3D	11	7	9	9	7	9	
D2 to Lily – CS	2D	5	7	4	5	6	7	5
	3D	6	5	4	5	4	4	
Lily to D2 – CS	2D	9	11	9	8	9	9	9
	3D	13	10	9	8	5	8	

Table 11.6.1-2. Summary of the number of considered alternatives for each target colour, both learning environments (LE) and all illuminant change conditions. The highlighted column lists the mean number of considered alternatives across learning environment for each illuminant condition.

(b) Colour-differences analysis

It was demonstrated that colour differences between target colours and their alternatives serve well to explain observers' matching behaviour for the colour memory tasks. It seems that colours undergo rather small shifts in memory and observers therefore select colours that are perceptually similar to the target colour when it cannot be recalled exactly.

The same analysis was applied to the data of the six colour constancy tasks reported above, to see whether erroneous matching follows a similar pattern. However, it was revealed that colour differences cannot explain matching behaviour for the colour constancy tasks; not even a tendency could be observed. Although, alternatives with small ΔE_{ab} values were often chosen, observers also considered swatches that differed from the respective target colour by more than 20 ΔE_{ab} units. This occurred not only in some isolated occasions but frequently for all target colours and illuminant change conditions.

To summarise, this analysis provided an explanation of observers' matching behaviour for colour memory tasks however, it failed for colour constancy.

(c) Further analyses

(i) The attribute-analysis, which had not led to satisfying results by the determination of colour memory shifts, was also applied on the colour constancy data. Again, this method was proved to be inappropriate for this kind of data analysis.

(ii) All observers were naïve and inexperienced regarding colour constancy experiments. As mentioned above, it appears that some observers might not have understood the task perfectly. Theoretically there is a possibility that observers tried to perform a chromaticity match instead of surface match. Practically that was impossible because none of the test palettes contained a swatch that was perceptually similar to the target colour under the initial illuminant.

(d) In order to quantify the magnitude of imperfect matching an average colour difference was determined. The colour difference (ΔE_{ab}) of each selected alternative was multiplied by the frequency with which it was chosen and averaged across observers. Table 11.6.1-3 shows the mean colour difference for both learning environments, all illuminant change conditions and colour groups. Overall, selected alternatives for the 2D conditions were perceptually more different from the target colours than those selected in the 3D condition. In other words, smaller mistakes were made when target colours were learned as part of a 3D scene. It seems that the 3D learning environment leads not only to an improvement regarding prefect matches but also induces a smaller range of mistakes.

LE	illuminant change condition	colour groups blue	red	yellow	Ø	Ø
2D	D1 to Tun - RW	15.60	17.45	11.83	14.96	
	Tun to D1 - RW	4.46	15.92	8.86	9.75	
	D2 to Lily - RW	9.12	4.82	7.75	7.23	
	D2 to Lily - RW	12.63	13.99	14.98	13.87	
	Lily to D2 - CS	12.68	11.68	9.73	11.36	
	Lily to D2 - CS	11.63	17.58	15.63	14.95	**12.02**
3D	D1 to Tun - RW	12.97	12.62	8.89	11.49	
	Tun to D1 - RW	6.29	14.30	5.55	8.71	
	D2 to Lily - RW	5.27	9.34	9.91	8.17	
	D2 to Lily - RW	9.77	15.77	13.98	13.17	
	Lily to D2 - CS	8.76	9.40	6.38	8.18	
	Lily to D2 - CS	16.22	15.40	11.89	14.50	**10.70**
	Ø	10.45	13.19	10.45		

Table 11.6.1-3. Summary of mean colour differences between the selected alternatives and the respective target colours. The highlighted row and column indicate mean ΔE_{ab} values across colour groups and illuminant change conditions, respectively. The last column provides the mean ΔE_{ab} values for both learning environments across all colour groups and illuminant change conditions.

11.6.2 Summary

A series of analyses was carried out but none of them led to a satisfying result, i.e. none of them revealed a consistent pattern of how alternatives were selected. It was expected that erroneous matches in the colour constancy tasks would follow a pattern comparable to that discovered for colour memory. However, the analyses indicate that erroneous matching was not systematic and it remains unclear what motivated observers to select some swatches and others hardly ever.

Although target colours were not matched perfectly in almost 80% of all trials the mean colour difference of all selected alternatives (12.02 and 10.70 ΔE_{ab} units for the 2D and 3D learning environment, respectively) indicate that observers did not select alternatives arbitrarily.

12 Determination of colour categories for experimental RW and CS colours

The colour categories for experimental real-world (RW) and computer-simulated (CS) colours were determined for two different reasons. First, after comparing observers' colour memory and colour constancy performances for RW and CS stimuli, the present experiment provides a further opportunity to contrast performances across media. It has been argued (chapter 7) that a wider range of stimuli and tasks should be used to investigate whether the substitution of RW by CS stimuli is valid.

The second reason for conducting this experiment rises from the results of the previous colour constancy experiments. It was noticed that some target colours were apparently much more difficult to match than others. The level of difficulty of the three colour groups has already been discussed (section 8.1.6.3). However, numerous variations could not be explained by colour differences. The determination of the colour categories of the experimental colours is expected to deliver explanations for these variations.

Note that the concepts of assigning a colour to a term or a category are here used synonymously.

This chapter comprises seven sections. Section 1 describes the experimental procedure and section 2 is about the general categorisation of the experimental colours. In sections 3, 4 and 5 different analyses are conducted to investigate the obtained data. In section 6 observers' colour constancy performance using a matching by selection and a colour naming paradigm is contrasted. In the last section (7) the effect of colour categorisation on colour matching is analysed.

12.1 Procedure

Colour categories were determined for all experimental RW and CS colour swatches (for materials see chapter 8). Each task started with a 2 min adapting to the prevailing illuminant while viewing the white palette. Then three test palettes, one of each colour group, were put consecutively in view and observers assigned one of the eleven basic colour terms[1] to each swatch; no time limit was set. Each observer repeated this task five times for each illuminant (D1, D2, Tun and Lily for RW and D1, D2 and Lily for CS) and presentation medium, resulting 1680 colour namings per observer. The total number of namings in this experiment mounts consequently up to 10080 (1680 x $6_{\text{observers}}$).

RW: $48_{\text{colour swatches}}$ x $4_{\text{illuminants}}$ x $5_{\text{repetitions}}$ = 960_{namings}

CS: $48_{\text{colour swatches}}$ x $3_{\text{illuminants}}$ x $5_{\text{repetitions}}$ = 720_{namings}

Observers

Six observers, who had already participated in one of the previously reported experiments, completed the colour naming experiment.

12.2 Used colour terms

From the eleven available colour terms two were never used, namely white and black. Figure 12.2-1 shows how often each colour term was used summed across illuminants. The distribution does not vary significantly between naming RW and CS swatches. For both presentation media almost half of the swatches were named as either green or

[1] The eleven basic colour terms are purple, blue, green, yellow, orange, red, pink, brown, white, grey and black (Berlin & Kay, 1969).

orange, the terms red and brown were hardly ever used. Analyses of individual observer naming behaviour showed very similar distributions as the one in the figure. Figure 12.2-2 shows the distribution of colour names for each illuminant and presentation media.

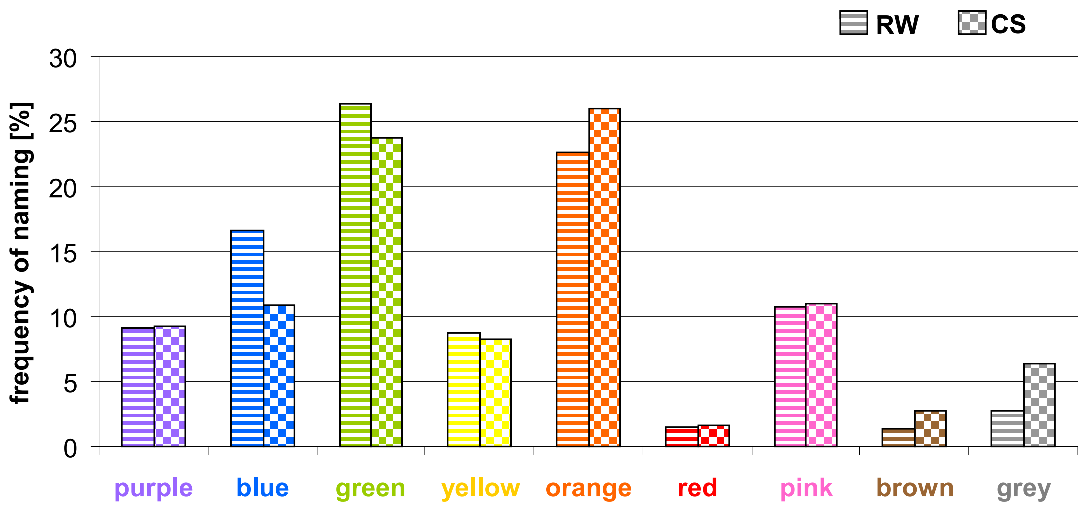


Figure 12.2-1. Summary of colour namings for all illuminant conditions and both presentation media.

12.2.1 Discussion

As can be seen in figure 12.2.1-1 (reprint of figure 6.3-1) colour categories vary considerably in size; green is by far the largest category. This means that no other category includes so many hue and lightness steps (Munsell colours were used to determine categories for this plot). Blue, purple and pink are medium sized categories, and orange, yellow, red and brown occupy rather small areas in the colour space.

The experimental colours used in this study spread more or less evenly across the hue circle of the NCS space and their approximate locations in the Munsell space are shown in figure 12.2.1-2. A visual analysis of this figure reveals that the majority of swatches lay within the boundaries of the green category. This is in agreement with the

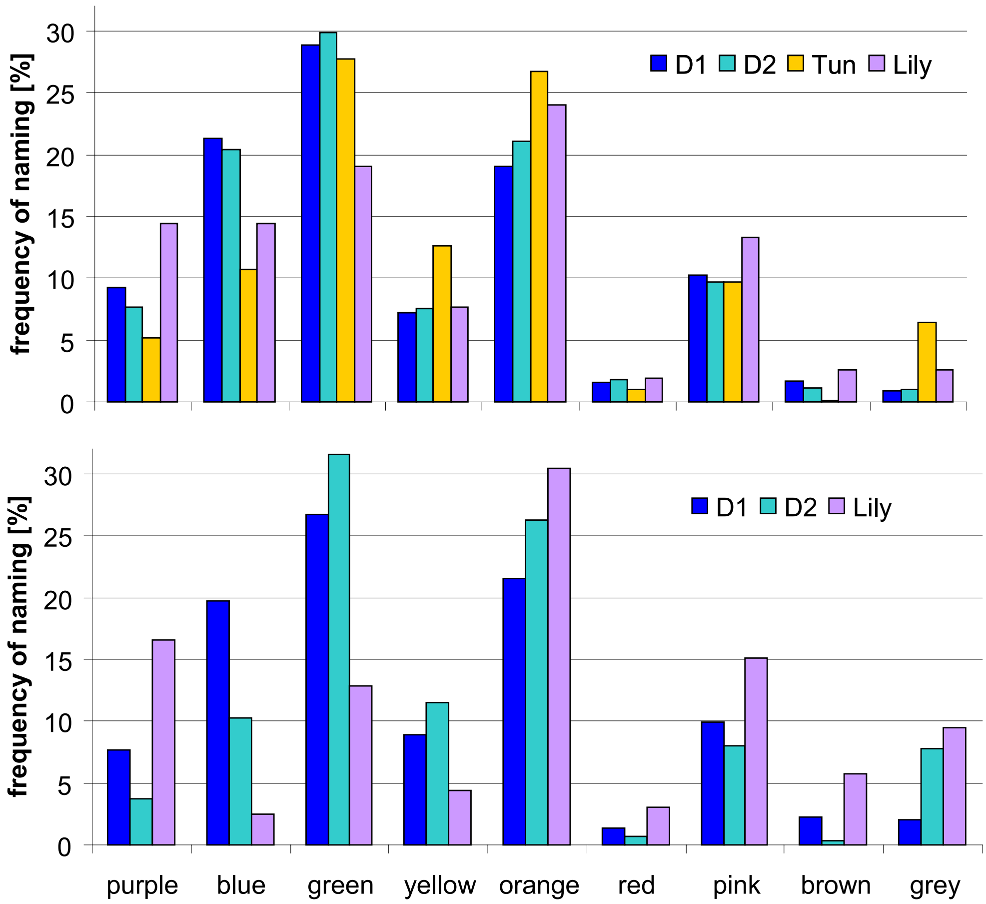


Figure 12.2-2. Top: colour naming of RW swatches under the illuminants D1, D2, Tun and Lily. Bottom: colour naming of CS swatches under the illuminants D1, D2 and Lily.

actual colour naming as green was one of the most frequently used terms. Fewer swatches lay in the blue and purple category and the terms purple and blue were also used less frequently to describe swatches. The frequencies with which the terms yellow and pink were used were also in agreement with the actual sizes of the categories. Very few swatches were named either red or brown and as the figure indicates, these are also two of the smallest categories.

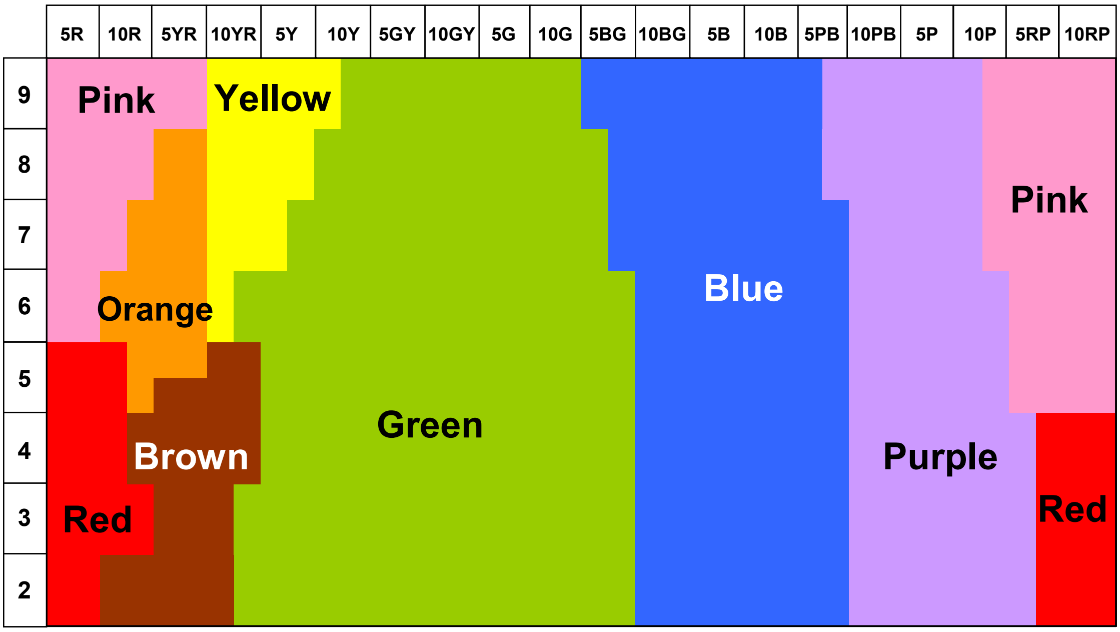


Figure 12.2.1-1. Colour categories in English. (Data redrawn from Roberson et al. (2000). A set of 160 fully saturated Munsell colour chips was used to determine the colour categories. The array consisted of 20 evenly spread hues (levels 5 and 10) and eight levels of lightness (value 2 – 9).

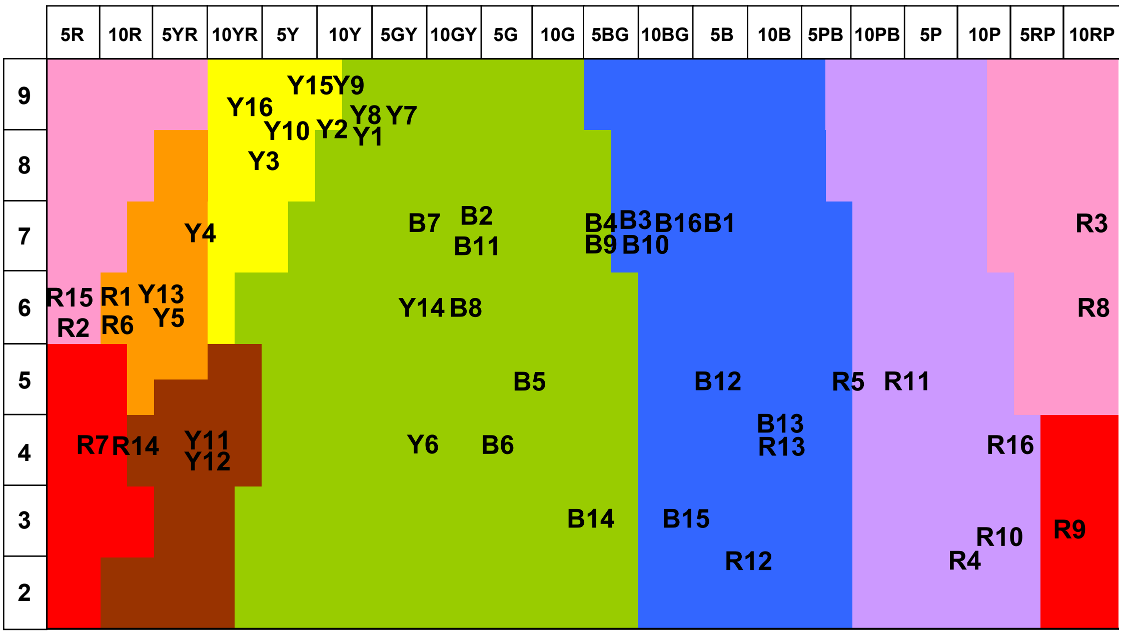


Figure 12.2.1-2. Approximate location of the experimental NCS colours within the colour categories as determined by Roberson et al. (2000). B, R and Y are the abbreviations for blue, red and yellow colour group, respectively.

However, there is a significant disagreement between the frequency with which the term orange was used and the size of the corresponding colour category. Overall, orange was the second most frequently used term while the orange category is one of the smallest. As stated above, colour swatches spread evenly across the hue circle (see figure 9.1.3-1) and this effect was therefore not due to a biased selection of swatches.

Figure 13.2-2 cannot explain the effect entirely but shows that the frequency with which the term orange was used depended considerably on the illuminant. For example, more swatches were called orange under Tun (RW only) and Lily than under D1 and D2.

To summarise, the frequency with which colour terms are used to describe a more or less homogeneous selection of swatches reflects roughly the size of the corresponding categories. Overall it appears that RW and CS swatches are named similarly. However, naming appears to depend, to a certain extent, on the illuminant under which swatches are presented. (This dependency is further analysed in the following sections.)

12.3 Consistency of colour naming

Every swatch was named five times by each observer but not every swatch was consistently (five times) assigned the same colour term. For example, one observer named a particular swatch for the same illuminant and presentation medium twice blue and thrice green. In such a case the predominant colour term was chosen (in the example it was green) to describe the overall colour appearance of this swatch. Detailed information about predominant colour terms of individual observers is provided in table AP-7 (appendix).

In order to quantify the degree of consistency, a consistency score (Co_i) was introduced and calculated for each colour swatch (i), under each illuminant and for both presentation media individually:

$$Co_i = \frac{No_{pre.c.t.i}}{No_{rep.}} \quad (13)$$

where $No_{pre.c.t.i}$ is how often the predominant colour term was used and $No_{rep.}$ is the number of completed tasks; which was here always five. The calculated scores were averaged for each observer. Table 12.3-1 shows a summary of consistency scores for each observer and experimental condition. The consistency scores were on average 0.86, ranging from 0.80 to 0.90.

		Observer						
presentation medium	**illuminant**	**1**	**2**	**3**	**4**	**5**	**6**	**Ø**
RW	D1	0.80	0.84	0.81	0.83	0.90	0.90	0.85
	D2	0.88	0.87	0.89	0.82	0.91	0.90	0.88
	Tun	0.84	0.80	0.86	0.84	0.91	0.89	0.86
	Lily	0.88	0.85	0.83	0.80	0.85	0.90	0.85
CS	D1	0.82	0.90	0.88	0.85	0.92	0.90	0.88
	D2	0.86	0.88	0.86	0.89	0.90	0.90	0.88
	Lily	0.84	0.86	0.88	0.85	0.88	0.82	0.85
	Ø	0.84	0.86	0.86	0.84	0.90	0.89	

Table 12.3-1. Summary of consistency scores for each observer under each illuminant and for both presentation media. The highlighted row and column indicate the mean scores over all illuminants and observers, respectively.

12.3.1 Statistical analysis

(a) RW

- A Friedman's four-way ANOVA[1] was conducted on the factor illuminant (levels: D1, D2, Tun and Lily). No main effect of was found.
- A Friedman's six-way ANOVA was conducted on the factor observer (levels: observer 1 to 6). There was a main effect χ^2 = 13.06, p = 0.023, which was further investigated by a series of Wilcoxon Signed Ranks Tests. None of the paired comparison reached a significant level but as can be seen in figure 12.3.1-1 the consistency score for observer 4 was lower than for any other observer.

(b) CS

- A Friedman's three-way ANOVA was conducted on the factor illuminant (levels: D1, D2 and Lily). No main effect of was found.
- A Friedman's six-way ANOVA was conducted on the factor observer (levels: observer 1 to 6). There was no significant main effect.

(c) Comparison across presentation media

- A Friedman's two-way ANOVA was conducted on the factor presentation medium (levels: RW and CS). The main effect did not reach a significant level χ^2 = 1.47, p = 0.225 (RW-mean Co = 0.86 SD 0.05 and CS-mean Co = 0.87 SD 0.04).

[1] Friedman's analysis of variance is a non-parametric test similar to the parametric repeated measures ANOVA (Siegel & Castellan, 1988).

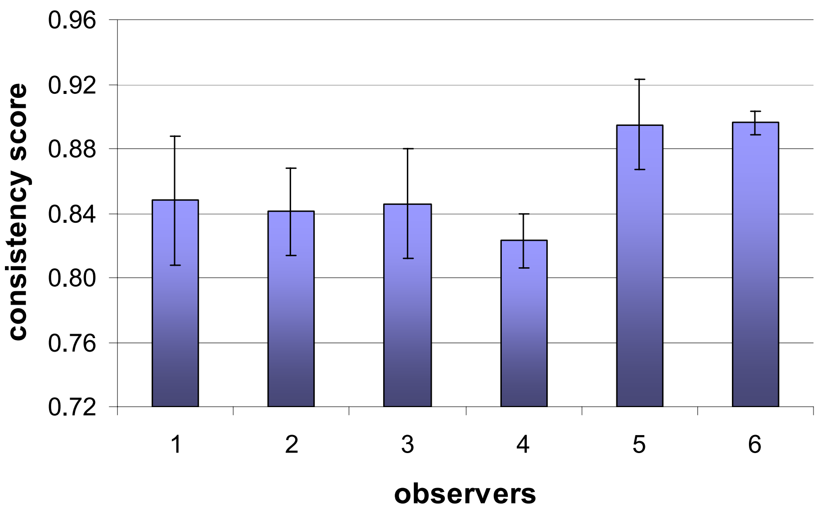


Figure 13.3.1-1. None of the paired comparisons between observers reached a significant level. However, the consistency scores for observer 5 and 6 were considerably higher than for all other observers ($p = 0.068$). There was only one exception, the difference between observer 5 and 1 was statistically less pronounced. Error bars indicate ± 1 SD.

12.3.2 Discussion

A visual analysis of table 12.3-1 reveals already that consistency scores did not undergo major fluctuations. Statistical analyses confirmed that none of the differences was significant. In other words, the consistency with which swatches were assigned a colour term was independent of illuminant, presentation medium and observer. Observers' performance was not perfect but an overall consistency score of 0.86 indicates that they were reasonably certain about the category a swatch belonged to.

12.4 Colour naming within an illuminant condition and across presentation media

The consistency score does not reveal whether swatches were named equally. For example, although colour swatches were named on average with the same consistency under illuminant D2 for both presentation media (0.88), they were not necessarily assigned the same colour term. A certain swatch might have been named blue with a consistency of 0.88 under D2 when shown as a paper swatch and green with exactly the same consistency when displayed on a monitor. Therefore it is necessary to analyse in how many cases individual RW and CS swatches were assigned exactly the same term.

Table 12.4-1 shows an example of how this analysis was conducted. Here the predominant colour terms of all swatches (RW and CS) of the blue colour group (under D1; one observer) are listed and compared whether they are identical for a particular swatch. The only swatches that were not assigned the same term for RW and CS were B5 and B10. Hence, in 87.5% of the cases swatches were named identically for RW and CS; henceforward this will be called the percentage of agreement.

This analysis was then repeated for the other two colour groups, the illuminants D2 and Lily and the remaining five observers. The resulting percentages of agreement were subsequently averaged to obtain a single value for each observer and illuminant condition. Table 12.4-2 provides a summary of these data. The overall percentage of agreement was 82.3% for D1, 72.9% for D2 and 64.6% for Lily.

label	RW	CS	naming
B1	blue	blue	same
B2	green	green	same
B3	blue	blue	same
B4	blue	blue	same
B5	green	blue	different
B6	green	green	same
B7	green	green	same
B8	green	green	same
B9	blue	blue	same
B10	purple	grey	different
B11	green	green	same
B12	blue	blue	same
B13	purple	purple	same
B14	blue	blue	same
B15	Blue	blue	same
B16	Blue	blue	same

Table 12.4-1. Naming of colour swatches (one observer) of the blue colour group, which were presented either under the real or simulated illuminant D1. The last column indicates whether swatches were assigned the same colour term.

	illuminant		
observer	D1	D2	Lily
1	68.75	68.75	54.17
2	89.58	70.83	77.08
3	81.25	89.58	77.08
4	85.42	64.58	47.92
5	81.25	68.75	66.67
6	87.50	75.00	64.58
Ø	82.29	72.92	64.58

Table 12.4-2. Summary of percentages of agreement across media for each observer and illuminant condition.

12.4.1 Statistical analysis

- A Friedman's three-way ANOVA was conducted on the factor illuminant (levels: D1, D2 and Lily). There was a main effect $\chi^2 = 7.91$, $p = 0.019$, which was further investigated by a series of Wilcoxon Signed Ranks Tests. Paired comparison revealed a significant difference between D1 and Lily $p = 0.027$ (D1-mean percentage of agreement = 82.3% SD 7.4 and Lily-mean percentage of agreement = 64.6% SD 11.9). Other differences did not reach a significant level (D2-mean percentage of agreement = 72.9% SD 8.8) (figure 12.4.1-1).

- A Friedman's six-way ANOVA was conducted on the factor observer (levels: observer 1 to 6). There was no significant main effect.

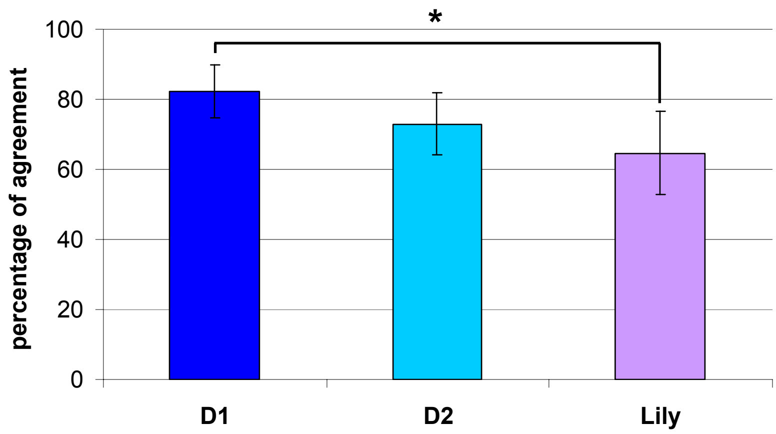


Figure 12.4.1-1. Mean percentages of agreement for the illuminants D1, D2 and Lily. Error bars indicate ± 1 SD and * indicates a significant difference at $p \leq 0.05$.

12.4.2 Discussion

The analysis of colour naming within an illuminant condition and across presentation media was conducted to investigate whether real and computer-simulated swatches are assigned identical colour terms.

In 82.3% of the cases RW and CS swatches under D1 were assigned the same colour term; in 72.9% and 64.6% of the cases when presented under D2 and Lily, respectively. Despite high consistency scores for all illuminants and both presentation media, the steadiness among observers when naming RW and CS swatches was significantly higher under D1 than Lily (but similar for D1 and D2, and D2 and Lily). This is so far the first time that observers' performance was dependent on the illuminant. Such an effect has neither been found for the previously reported colour memory nor colour constancy experiments.

Despite this difference in performance it has to be emphasised that in the worst case, naming RW and CS swatches under Lily, still about two third of the RW and CS swatch namings were in agreement. Regarding the comparison between performances for RW and CS stimuli, this means that overall observers' performance in this task was comparable for both presentation media.

To summarise, the cross-media comparison of colour naming showed that the overall performance of observers for paper and displayed stimuli is similar, although not perfect. In other words, colour swatches that are either reflective or self-luminous and otherwise identical are not always named identically.

12.5 Colour naming across illuminant conditions

While colour naming within illuminant conditions and across presentation media was analysed in the previous section, the present analysis concentrates on comparisons across illuminant conditions within a presentation medium.

Colour naming was compared between all possible combinations of illuminant condition, i.e. it was determined how often swatches were assigned the same colour term when presented under different illuminants. As observers adapted to each illuminant before the actual task, it was expected that naming would not vary considerably between illuminants, i.e. rather high percentages of agreement were expected.

Table 12.5-1 provides a summary of percentages of agreement between illuminant conditions for each observer and presentation medium. The percentage of agreement was on average 67.4%. The statistical analyses concentrate on comparisons of percentages of agreement within a presentation medium. However, subsequently a direct comparison between RW and CS is conducted, including the illuminant conditions D1 – D2, D1 – Lily and D2 – Lily.

	RW						CS		
observer	**D1 - D2**	**D1 - Tun**	**D1 – Lily**	**D2 - Tun**	**D2 - Lily**	**Tun - Lily**	**D1 - D2**	**D1 - Lily**	**D2 - Lily**
1	91.67	70.83	72.92	72.92	75.00	70.83	70.83	43.75	45.83
2	89.58	70.83	68.75	77.08	66.67	66.67	68.75	66.67	60.42
3	87.50	72.92	81.25	77.08	81.25	72.92	89.58	66.67	64.58
4	77.08	58.33	62.50	60.42	60.42	58.33	68.75	33.33	31.25
5	97.92	68.75	62.50	79.17	58.33	64.58	77.08	50.00	47.92
6	91.67	75.00	54.17	75.00	54.17	62.50	75.00	47.92	41.67
Ø	89.24	69.44	67.01	73.61	65.97	65.97	75.00	51.39	48.61

Table 12.5-1. Summary of percentages of agreement between all possible illuminant conditions for each observer and presentation medium. The highlighted row indicates the mean percentage of agreement across all observers.

12.5.1 Statistical analysis

(a) RW

- A Friedman's six-way ANOVA was conducted on the factor illuminant condition (levels: D1-D2, D1-Tun, D1-Lily, D2, Tun, D2-Lily and Tun-Lily). There was a main effect χ^2 = 37.84, $p < 0.01$, which was further investigated by a series of Wilcoxon Signed Ranks Tests. All comparisons comprising the illuminant condition D1-D2 reached significant level at $p < 0.01$. None of the other comparisons were significant (figure 12.5.1-1). Mean percentages of agreement for the six illuminant change conditions were as follows: D1-D2 89.2% SD 9.0, D1-Tun 69.4% SD 8.8, D1-Lily 67.0% SD 11.9, D2-Tun 73.6% SD 11.8, D2-Lily 66.0% SD 12.0, Tun-Lily 66.0% SD 13.4.
- A Friedman's six-way ANOVA was conducted on the factor observer (levels: observer 1 to 6). There was a main effect χ^2 = 11.51, $p = 0.042$, which was

further investigated by a series of Wilcoxon Signed Ranks Tests. The percentage of agreement achieved by observer 4 (mean = 62.8% SD 7.1) was significantly lower than for observer 1 (mean = 75.7% SD 8.0), 2 (mean = 73.3% SD 8.9) and 3 (mean = 78.8% SD 5.7), p = 0.026, p = 0.027 and p = 0.027, respectively (figure 12.5.1-2). Mean percentages of agreement were 71.9% SD 14.6 and 68.8% SD 14.6 for observe 5 and 6, respectively.

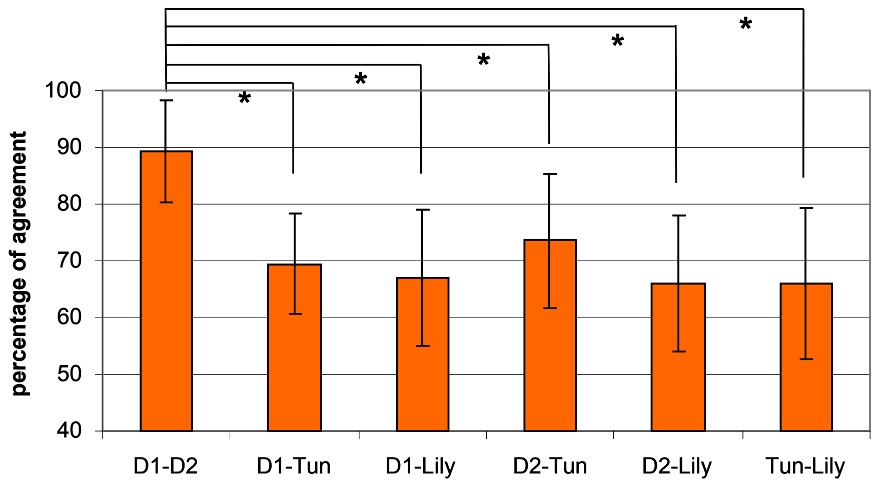
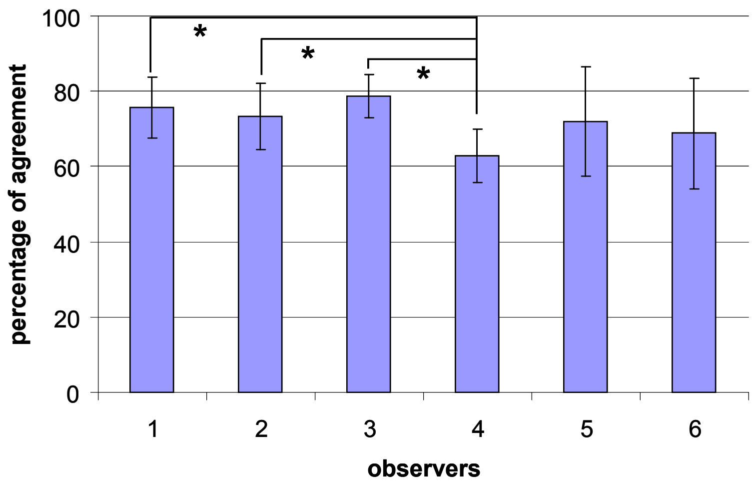


Figure 12.5.1-1. Mean percentages of agreement for all possible illuminant conditions using RW swatches. Error bars indicate ± 1 SD and * indicates a significant difference at $p < 0.05$.

Figure 12.5.1-2. Mean percentages of agreement for all six observers across



illuminant conditions. Error bars indicate ± 1 SD and * indicates a significant difference at $p < 0.05$.

(b) CS

- A Friedman's three-way ANOVA was conducted on the factor illuminant condition (levels: D1-D2, D1-Lily and D2-Lily). There was a main effect χ^2 = 17.15, $p < 0.01$, which was further investigated by a series of Wilcoxon Signed Ranks Tests. Both comparisons comprising the illuminant condition D1-D2 (mean = 75.0% SD 12.1) reached significant level at p = 0.001. The difference between D2-Lily (mean = 48.6% SD 21.8) and D1-Lily (mean = 51.4% SD 20.6) was not significant (figure 12.5.1-3).
- A Friedman's six-way ANOVA was conducted on the factor observer (levels: observer 1 to 6) and a significant main effect was found χ^2 = 11.17, p = 0.048. In subsequent Wilcoxon Signed Ranks Tests none of the paired comparison reached a significant level, but as can be seen in figure 12.5.1-4 the percentage of agreement achieved by observer 4 was lower than for other observers.

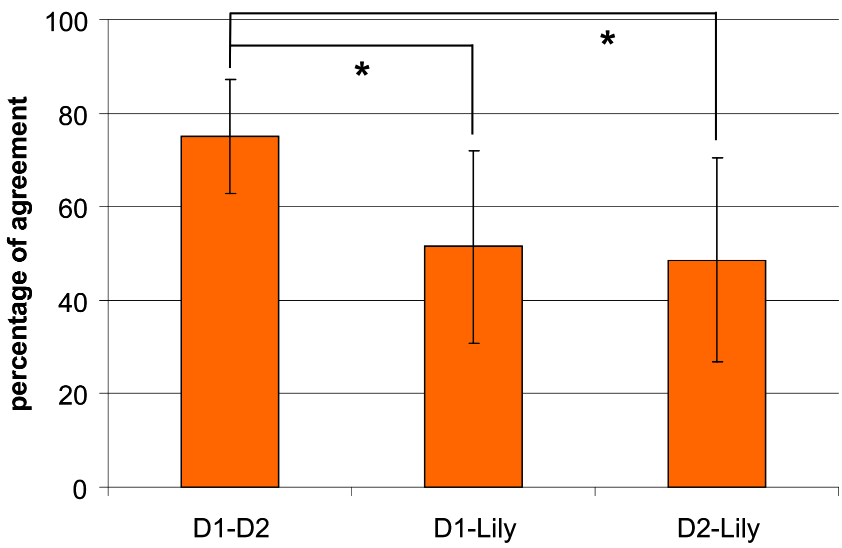


Figure 12.5.1-3. Mean percentages of agreement for the three possible illuminant conditions using CS swatches. Error bars indicate ± 1 SD and * indicates a significant difference at p < 0.05.

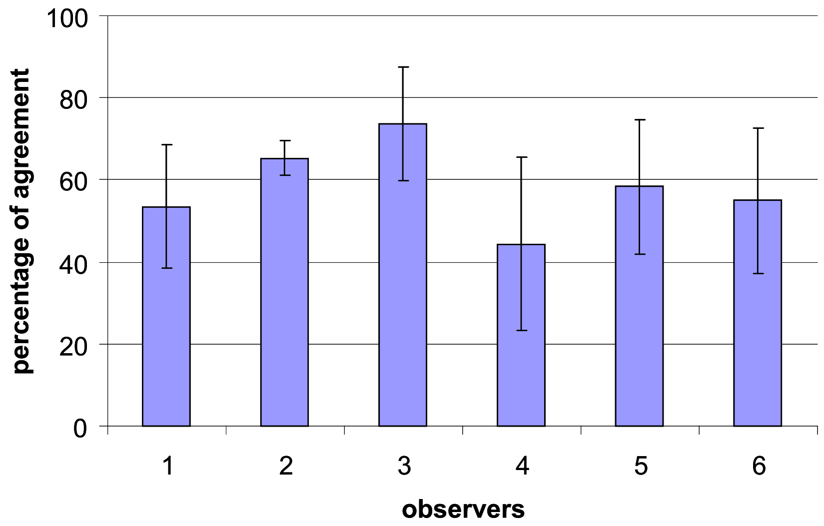


Figure 12.5.1-4. Although the main effect of observers was significant, none of the subsequent paired comparisons reached a significant level. Error bars indicate ± 1 SD.

(c) Comparison across presentation media

- A Friedman's two-way ANOVA was conducted on the factor presentation medium (levels: RW and CS). The main effect was significant $\chi^2 = 25.00$, $p < 0.01$, the percentage of agreement was significantly higher for naming RW than CS swatches (RW-mean = 74.1% SD 15.3 and CS-mean = 58.3% SD 21.9). Wilcoxon Signed Ranks Tests showed that the percentages of agreement for the illuminant conditions D1-D2, D1-Lily and D2-Lily were significantly higher for RW than for CS swatches $p = 0.003$, $p = 0.002$ and $p = 0.001$, respectively.

12.5.2 Discussion

Comparing colour naming of RW swatches under different illuminants shows that the highest agreement is achieved for swatches under D1 and D2. In this case almost 90% of all swatches were assigned the same term. The percentage of agreement achieved for the other five comparisons were significantly lower, ranging from approximately 66% to 74%. Initially it was hypothesised that the agreement would be consistently high for all comparisons because adaptation compensated for the changed illumination. However, the overall agreement for RW swatches was about 74% and therefore lower than expected. It is unlikely that this effect was due insufficient adaptation but that different illuminants had stronger effects on naming than anticipated.

The same trend was observed for the colour naming of CS swatches. The agreement between swatches under D1 and D2 was highest, followed by much lower percentages of agreement for the other two comparisons. However, the overall agreement between assigned colour terms for CS swatches was with approximately 58% significantly lower than RW swatches.

The colour categorisation experiment was conducted to investigate whether CS stimuli are a valid substitution for RW ones. First the consistency of colour naming was analysed and no difference in performance was noticed for the two kinds of stimuli. The second analysis showed that observers' performance in naming CS and RW swatches depended on the illuminant. In other words, for one illuminant CS stimuli were valid substitutions for RW stimuli but not for others. This statement however, has to be put into perspective because there is no classification of validity of substitution. One could argue that a substitution is only valid if both kinds of stimuli yield exactly the same performance. On the other hand, certain variation resulting from the different kinds of stimuli might be acceptable as well. However, if such variations are acceptable then their margin has to be defined.

The third analysis revealed that observers' performance was better for RW than for CS stimuli however, the results showed the same tendencies solely on overall different levels. Therefore under the particular experimental conditions no clear conclusions can be drawn of whether CS stimuli are valid substitutions for RW stimuli.

12.6 Measuring colour constancy by colour naming

In chapter 11 colour constancy was assessed using a matching by selection paradigm. A target colour was learned under one illuminant and then matched by selecting one swatch out of an array of 16 under a different illuminant. Observers' performance was expressed in hit rates, analysed and finally a colour constancy index was computed by normalising observers' colour constancy performance by their colour memory performance (assessed in chapter 9).

In the present experiment the experimental colours were named under various illuminants, i.e. it was tested whether the colour appearance of the 48 swatches was perceived as unchanged despite a change in illumination. Hence, colour constancy was assessed, but in contrast to chapter 11 it was now assessed with a colour naming instead of a matching by selection procedure. Colour naming procedures have already been used successfully by other researchers to determine colour constancy (e.g. Hansen et al., 2007; Smithson & Zaidi, 2004). Subsequently, data is analysed similarly as in chapter 11 and finally colour constancy indices are computed applying exactly the same equation as before. Consistency scores for RW swatches under D1 are used as colour memory baseline performance.

To summarise, the purpose here is to investigate whether observers' colour constancy performance achieved with a colour naming procedure is comparable to that achieved with a matching by selection paradigm. Generally it is expected that observers show a higher degree of colour constancy for the colour naming procedure for a simple reason. For the matching by selection procedure observers selected one swatch out of 16 whereas in the colour naming procedure they choose one term out of nine (see section 12.2).

12.6.1 Analysis

In section 12.5 the percentage of agreement of namings across illuminants was analysed. The present analysis focuses on the illuminant change conditions analogue to those reported in chapter 12, i.e. illuminant changes between D1 and Tun, and D2 and Lily. Percentages of agreement express the level of achieved colour constancy equivalent to hit rates earlier. Table 12.6.1-1 shows the percentages of agreement achieved by each observer for the relevant illuminant change conditions (these data was already presented in table 12.5-1).

	RW		CS
observer	**D1-Tun**	**D2-Lily**	**D2-Lily**
1	70.8	75.0	45.8
2	70.8	66.7	60.4
3	72.9	81.3	64.6
4	58.3	60.4	31.3
5	68.8	58.3	47.9
6	75.0	54.2	41.7
Ø	69.4	66.0	48.6

Table 12.6.1-1. Summary of percentages of agreement for the illuminant change conditions D1-Tun (RW) and D2-Lily (RW and CS). The highlighted row indicates the mean percentage of agreement across all observers.

A Friedman's two-way ANOVA (levels: D1-Tun and D2-Lily (both RW)) was conducted. It was revealed that the difference between the illuminant change condition D1-Tun and D2-Lily was not significant (D1-Tun mean = 69.4% SD 5.8 and D2-Lily mean = 66.0% SD 10.4).

Another Friedman's two-way ANOVA was conducted to compare observers' performance for the illuminant change condition D2-Lily for both presentation media (levels: RW-D2-Lily and CS-D2-Lily); a significant main effect was found $\chi^2 = 6.00$, $p = 0.014$ (CS-D2-Lily mean = 48.6% SD 12.3). This indicates that observers performed significantly better for RW than CS stimuli.

12.6.2 Computing a colour constancy index

Colour constancy indices were computed applying equation 12. Here observers' colour constancy performance (percentage of agreement) is normalised by their consistency score achieved for naming RW swatches under illuminant D1. Table 12.6.2-1 provides a summary of those consistency scores (in percentage), the percentages of agreement for all illuminant change conditions (repeated from above) and the resulting colour constancy indices.

	RW - D1	RW		CS	CCI		
observer	**Co**	**D1-Tun**	**D2-Lily**	**D2-Lily**	**D1-Tun$_{RW}$**	**D2-Lily$_{RW}$**	**D2-Lily$_{CS}$**
1	80.0	70.8	75.0	45.8	0.89	0.94	0.57
2	84.0	70.8	66.7	60.4	0.84	0.79	0.72
3	81.0	72.9	81.3	64.6	0.90	1.00	0.80
4	83.0	58.3	60.4	31.3	0.70	0.73	0.38
5	90.0	68.8	58.3	47.9	0.76	0.65	0.53
6	90.0	75.0	54.2	41.7	0.83	0.60	0.46
Ø	84.7	69.4	66.0	48.6	0.82	0.79	0.58

Table 12.6.2-1. Summary of observers' consistency scores (Co) (as percentage) for RW swatches under illuminant D1 (second column), percentages of agreement for all illuminant change conditions (columns 3 to 5) and the computed colour constancy indices (columns 6 to 8). The highlighted row indicates mean values across all observers.

12.6.3 Discussion

This analysis aimed to investigate the comparability of colour constancy performances obtained by the colour naming and matching by selection procedure. Note that the present indices are compared to those of the 2D condition only, as colour naming was tested for swatches and not for 3D objects. For the present experiment it was found that observers' colour constancy performance for RW stimuli was independent of the illuminant change condition. In other words, the percentage of agreement between colour naming under D1 and Tun was similar to the one under D2 and Lily. It was also revealed that observers achieved a higher percentage of agreement for RW than for CS swatches. Both findings are in line with previous findings (see section 11.3), underlining their generality.

As expected the level of colour constancy for colour naming was considerably higher than for the selection task. The reason for this lies not only in the reduced number of available ‘alternatives’ (16 in selection and 9 in the naming task) but also in the actual procedure. Colour naming was repeated five times under the same condition and then the predominant colour appearance was determined for each swatch. Any sign of uncertainty an observer might have had about a swatch was removed by this process. Although the absolute level of colour constancy was higher for the colour naming than the selection procedure, overall the same tendencies and effects were revealed. This finding underlines that both methods serve well for the investigation of colour constancy.

The drop in performance for RW and CS stimuli occurred for both methods, while performance decreased by 12.5% in the naming task it had previously decreased by almost 22% (2D stimuli). During the selection task observers adapted to changed illuminant conditions for 2 min and it was assumed that chromatic adaptation was almost complete after this time span. Observers adapted also for 2 min to the prevailing illuminant during the naming procedure but the actual naming took several minutes during which the observers continued adapting. It might be possible that the less pronounced drop in performance was due to a more complete state of adaptation.

Colour constancy index

As the overall performance for the colour naming procedure was better than for the selection task, it is not surprising that colour constancy indices were also higher. The relation between performances for the different tasks is preserved; indices for both RW conditions are similar and considerably lower for the CS condition.

To summarise, measuring colour constancy by colour naming leads to the same performance pattern as a matching by selection procedure. The only difference is that the overall level of colour constancy is higher for the colour naming procedure.

12.7 Investigating the effect of colour categorisation on colour constancy

In section 11.1 to 11.4 colour constancy performance was presented as hit rates per colour group, despite the fact that matching was performed for single target colours. The reason for it lies in the experimental design. At the beginning it had been decided (a) to use 48 colours and to divide them into three colour groups and (b) that each task should consist of 18 trials. This meant that the same colours had to be shown over and over again, this repetition could potentially introduce a learning effect. In order to minimise the learning effect (especially for the target colours) two colours of each colour group were designated as target colours. Hence, each target colour was presented three times per task ($6_{\text{target colours}}$ x $3_{\text{repetitions}}$ = 18_{trials}), instead of presenting only one target colour per colour group six times per task. Theoretically the two target colours of a group were similarly easy or difficult to match (see section 8.1.6.3), and should therefore yield similar results.

However, this assumption could not be confirmed by empirical data. In some conditions, observers matched one target colour correctly almost constantly whereas the other one was hardly ever matched correctly. Sometimes the opposite effect was observed, and in other occasions again, both target colours were matched equally well or poorly. This effect could not be explained by colour differences within a colour group but might have been due to colour categorisation effects.

As outlined earlier (section 5.1.2), in order to recognise a colour despite and illumination and/or context change (i.e. to be colour constant) it is important that this colour does not change colour category due to the illumination change. The present analysis investigates whether target colours were assigned different colour categories under the learning and the matching illuminant and whether this affected their matching performance.

It is expected that target colours, which are assigned the same term under the learning and the matching illuminant, are easier to recognise, and therefore to match, than target colours that underwent a categorical change.

12.7.1 Analysis

Three out of the six observers who completed the colour categorisation experiment had previously participated in one of the colour constancy experiments. Two had participated in the first colour constancy experiment (RW stimuli), one for the illuminant change condition D1 to Tun and the other for Tun to D1 (section 11.1). The third observer had completed a colour constancy task (CS stimuli) for the D2 to Lily illuminant change condition (section 11.2).

Because each of three observers completed colour constancy tasks for different illuminant change conditions the analysis is conducted individually.

For the analysis relevant data is listed in tables; in the first column are the target colours, in the second and third are the colour terms a respective target colour was assigned under the learning and the matching illuminant, respectively. In the fourth column it is indicated whether these terms were identical, and the mean hit rates (averaged for the 2D and 3D task) are shown in the last column

Observer 1

Observer 1 was assigned to the illuminant change condition Tun to D1 (RW); learning took place under Tun and matching under D1 (table 12.7.1-1). Half of the target colours (B9, R8 and R10) were assigned different terms under the learning and matching illuminant whereas the other half (B16, Y8 and Y10) was named identically. However, R8 and R10 (different categories across illuminants) were correctly matched in the same number of trials as Y8 and Y10, which were assigned the same term across illuminants. A similar situation was revealed for the blue target colours. While B9 was named green under Tun and blue under D1, B16 was constantly named blue. However, matching performance was poor for both of them.

obs. 1	**Tun** [learning illuminant]	**D1** [matching illuminant]	**category**	**HR**
B9	green	blue	different	0.17
B16	blue	blue	same	0.00
R8	orange	pink	different	0.50
R10	pink	purple	different	0.50
Y8	green	green	same	0.50
Y10	yellow	yellow	same	0.50

Table 12.7.1-1. Assigned colour terms for all target colours under the learning and matching illuminant. Observer's performance for the illuminant change condition Tun to D1 is shown in the last column; hit rates (HR) are presented as relative frequencies.

Observer 2

Observer 2 was assigned to the illuminant change condition D1 to Tun (RW); learning took place under D1 and matching under Tun (table 12.7.1-2). Four target colours (B16, R8, R10 and Y8) were assigned the same term under both illuminants. However, matching performance was not consistent, reaching from 0.0 up to 0.83. The two target colours that were named differently under both illuminants (B9 and Y10) did not yield similar hit rates.

obs. 2	D1 [learning illuminant]	Tun [matching illuminant]	category	HR
B9	blue	green	different	0.00
B16	blue	blue	same	0.00
R8	pink	pink	same	0.83
R10	pink	pink	same	0.33
Y8	green	green	same	0.17
Y10	green	orange	different	0.67

Table 12.7.1-2. Assigned colour terms for all target colours under the learning and matching illuminant. Observer's performance for the illuminant change condition D1 to Tun is shown in the last column; hit rates (HR) are presented as relative frequencies.

Observer 3

Observer 3 was assigned to the illuminant change condition D2 to Lily (CS); learning took place under D2 and matching under Lily (table 12.7.1-3). All target colours were named differently under both illuminants, except R10. The overall performance of this observer was poor irrespective of the target colour but note, that performance for this illuminant change condition for displayed stimuli was generally very poor.

obs. 3	D2 [learning illuminant]	Lily [matching illuminant]	category	HR
B9	green	grey	different	0.00
B16	blue	grey	different	0.17
R8	orange	pink	different	0.33
R10	pink	pink	same	0.17
Y8	green	brown	different	0.00
Y10	yellow	orange	different	0.00

Table 12.7.1-3. Assigned colour terms for all target colours under the learning and matching illuminant. Observer's performance for the illuminant change condition D2 to Lily for CS stimuli is shown in the last column; hit rates (HR) are presented as relative frequencies.

12.7.2 Discussion

The present analyses of observers' colour constancy performance was analysed taking into account the colour categories of the target colours. It was expected that target colours that were assigned the same colour term under the learning and the matching illuminant were easier to match than target colours that changed category.

Hit rates for the two blue target colours were generally very low and it appears that whether a target colour was assigned the same term under both illuminants did not affect observers' matching performance. The difference between hit rates for R8 and R10 (observers 2) is not due to categorical effects as neither of them changed category. However, this difference can also not be explained by colour differences as they were very similar for both target colours. In contrast, Y10 changed category but was correctly matched more often than Y8, which was assigned the same category. Something similar occurred for the red target colours matched by observer 3. R8 changed category but was matched correctly more often then R10, which did not undergo a category change.

It is evident that the irregularities in observers' matching behaviour cannot be explained by colour categorisation. To conclude, the hypothesis that colours that are consistently assigned the same colour name are matched correctly more frequently cannot be confirmed.

13 General discussion

In the present study a series of colour memory, colour constancy and colour categorisation experiments were conducted and emphasis was placed on (a) the investigation of colour constancy for two-dimensional and three-dimensional objects and (b) the comparison of observers' performance for real and computer-simulated stimuli.

Colour constancy for 2D and 3D objects

At the beginning every observer passed through a screening process to ensure that only colour normal observers with roughly equal colour memory participated in the study. Additionally, the obtained data from the participation test also provided a performance baseline for subsequent colour constancy experiments. Further colour memory experiments were carried out for a similar reason, to provide fundamental information about observers' ability to remember colours for the following analyses of observers' colour constancy performance. Colour constancy was assessed for various illuminant change conditions and it was essential to ensure that differences in observers' colour constancy performance were not due to differences in memorability of colours under specific illuminants. The colour memory experiment for 2D stimuli under different illuminants (section 10.2) had shown that observers remembered colours equally well under all experimental illuminants.

The colour memory experiment for delayed matching was conducted to establish a measure of deterioration of colour memory over a period of 2 min as most studies on this topic had reported such weakening (e.g. Collins, 1932; D'Ath et al., 2007; Epps & Kaya, 2004; Francis & Irwin, 1998; Jin & Shevell, 1996; Newhall et al., 1957; Perez-Carpinell et al., 1998a; Uchikawa, 1983). Surprisingly it was found that, under the particular experimental conditions used, colour memory did not deteriorate. This finding was essential for the analysis of the colour constancy experiments as they were designed to allow adapting for 2 min after each illuminant change. Without the colour memory experiment it would have been impossible to distinguish between the real colour constancy and the colour memory component in performance. The outcome of the colour memory experiment for delayed matching showed that colour constancy was not compromised by colour memory.

Observers' colour constancy performance for 2D and 3D stimuli was compared directly. The initial hypothesis that colour constancy would be better when colours are learned as part of a 3D than a 2D setup was confirmed. This effect was revealed for all constancy experiments using real-world (RW) and computer-simulated (CS) stimuli, underlining its robustness. It has to be highlighted that the 3D setup led only in the colour constancy, but not in the colour memory experiments, to an improvement in performance.

The difference in dimensionality was limited to the learning environment as matching was always performed from 2D palettes. It appears that the cue-rich 3D environment provided the observer with more precise information about the actual surface reflectance of the target colour, which subsequently facilitated the recognition of the same surface under changed illumination and context when compared to learning in the cue-poor palette environment.

Performance for real-world and computer-simulated stimuli

The investigation and comparison of observers' performance for RW and CS stimuli was another major objective in this study. Colour memory, colour constancy and colour categorisation experiments have been conducted using both kinds of stimuli (paper and displayed samples), i.e. using both presentation media. While 2D and 3D stimuli were used for testing colour constancy, observers' colour memory and the categorisation of all experimental colours was assessed solely with 2D stimuli.

The colour memory experiment was completed under different illuminants and it was revealed that observers' performance was independent of the illuminant and the presentation media. Whether coloured paper swatches were matched under, for example, D1 or Lily did not affect observers' performance, and observers' ability to recall colours was also unaffected by the choice of presentation media. As computer-simulated and real-world stimuli elicit comparable results it can be concluded that they are mutually exchangeable for memory tasks employing 2D stimuli.

Colour constancy experiments were completed for 2D and 3D, RW and CS stimuli. Although improved performance was noticed for both presentation media when target colours were learned as part of a 3D setup, the overall performance was much lower for CS than RW stimuli. It has been argued that this might have been due to a combination of suboptimal viewing conditions and the reduction of illuminant cues from RW to CS. It would be interesting to equate viewing conditions as far as possible for both presentation media and to repeat these experiments. It is likely that performances for RW and CS stimuli would be more similar. However, under the present circumstances it is not possible to draw conclusions about whether CS stimuli are a valid substitution for RW stimuli in colour constancy studies due to limitations of the experimental design.

All 48 experimental RW and CS colour swatches were named repeatedly under different illuminants and observers' naming behaviour was then evaluated in a series of analyses. It was found that observers' colour naming was not identical for the two presentation media and that the degree of discrepancy varied between analyses. For example, in some conditions CS and RW stimuli elicit similar results but not in others.

Although previous studies mainly concluded that computer-generated and displayed stimuli are valid substitutions for real-world stimuli (Granzier et al., 2009a; Johnston & Curran, 1996; Schirillo et al., 1990; Tatler et al., 2005) the present study cannot confirm this statement. The diversity of outcomes for different tasks does not allow drawing general conclusions.

Matching task

The original RW setup consisted exclusively of real surfaces. Consequently, only a limited range of surface colours could be presented for matching, which, for this reason, had to be made by selection. In other studies achromatic and chromatic settings are popular whenever observers are asked to judge surface appearance, because settings are continuously adjustable over a wide range of possibilities. The main drawback is that this kind of matching is not natural. In everyday life we see a particular surface colour and we have to decide whether it is the same colour that we remember. In such a situation a decision has to be made and a surface colour cannot be adjusted until it matches our memory. In this study, a more natural setting was replicated by providing a fixed set of alternatives, including the correct answer. Observers had merely to recognize the correct swatch.

In the studies by Brainard (1998) and Kraft and Brainard (1999) observers adjusted the appearance of the test patch by varying its illumination, which was independent of the scene illumination. In the experiment described by de Almeida et al. (2004) observers adjusted the appearance of a test-object that was projected into the scene.

A far more natural task was used by Nascimento et al. (2005) and Zaidi and Bostic (2008) where observers judged whether an appearance change was entirely due to an illumination change, or had to indicate an odd test object in the setup, respectively. A slightly different but equally natural procedure was employed by Granzier et al. (2009b). The authors asked observers to assign terms to presented coloured sheets of paper. Ling and Hurlbert (2006) used a matching by selection paradigm, however, all so-called surface colours were generated entirely by illuminating white surfaces. Observers found this setting artificial and confusing, and were unable to perform surface matches. In the present context matching tasks are referred to as being natural when the appearance of stimuli is judged without actively manipulating them.

In the colour memory and colour constancy experiments reported in this work matching was also performed by selection but all appearance changes corresponded to true surface reflectance changes. Observers experienced the setup and the procedure as natural and had no difficulty in performing their tasks; for RW as well as CS stimuli. The only time they had difficulties was during the CS colour constancy tasks. As discussed earlier (section 11.3.1), the illuminant and palette changes appeared artificial and observers therefore had problems performing the task. This is an indication that the naturalness of the task affected observers' performance. A potential drawback of a matching by selection paradigm is that, occasionally, observers have to compromise between swatches as the perfect match seems to be not available. However, it is argued that this is an important, and at the same time inevitable, component of a natural task.

Comparison with other studies

A colour constancy index was proposed, which is based entirely on performance (observers' colour constancy performance is normalised by their individual colour memory performance). Computed indices were on average between 0.58 and 0.76 for RW 2D and 3D, respectively, and between 0.17 and 0.41 for CS 2D and 3D, respectively. These indices allow comparing the results of the present study with those from other studies that used 3D stimuli. The levels of colour constancy reported in studies using real surfaces were between 0.11 and 0.83 (Kraft & Brainard, 1999), 0.81 and 0.93 (de Almeida et al., 2004), 0.61 and 0.84 (Ling & Hurlbert, 2006), and around 0.8 (Nascimento et al., 2005). Therefore, the level of colour constancy achieved in this study lies in the same range and is comparable to these earlier studies despite a different experimental approach and the computation of an alternative colour constancy index. A direct comparison with the studies by Zaidi and Bostic (2008) and Granzier et al. (Granzier et al., 2009b) is not easily possible as they expressed their results differently.

Chance-level

All test palettes consisted of 16 swatches and strictly speaking the chance-level of selecting the correct swatch was therefore 1/16 (6.25%). However, this is only a theoretical value as observers did never consider all available alternatives. Although they sometimes had difficulties selecting the perfectly matching swatch they did not pick alternatives at random. Hence selection was practically never performed out of 16, but a smaller number of alternatives. Which and how many alternatives were considered, depended strongly on observers, target colour and the illuminants under which stimuli were presented.

For example, for target colour R10 in the illuminant change condition D2 to Lily (RW, colour constancy task, 2D condition) observers considered only two alternatives besides the correct swatch. Thus, the selection effectively took place between three swatches and not 16; the chance-level was 33.3%. In illuminant change condition Tun to D1 (RW, colour constancy task, 3D condition) observers made their selection out of eight swatches instead of out of 16 and the effective chance-level was therefore 12.5%. That illustrates that a chance-level of 6.25% was solely a theoretical value, and for all RW and most CS tasks observers performed well above this level. Performance only sank below chance-level for some CS target colours in the 2D condition, but note that the overall performance in this condition was generally only slightly above the chance-level of 6.25%.

If observers' performance in relation to the chance-level had been of interest, the chance-level would have been calculated for each observer, target colour and experimental conditions individually.

Summary

The present study had two major objectives: (1) to investigate whether colour constancy improves for 3D in contrast to 2D setups and (2) to investigate whether computer-simulated stimuli are valid substitutions for real-world stimuli. Regarding (1), it has been demonstrated that learning a colour as part of a 3D setup improves colour constancy. Regarding (2), the investigation of observers' performance for RW and CS stimuli however led to ambiguous results, and the issue of whether CS stimuli are valid substitutions for RW stimuli needs to be further addressed.

References

Abramov, I., & Gordon, J. (2005). Seeing unique hues. *Journal of the Optical Society of America, A, Optics, Image Science, & Vision, 22*(10), 2143-2153.

Agostini, T., & Bruno, N. (1996). Lightness contrast in CRT and paper-and-illuminant displays. *Perception and Psychophysics, 58*(2), 250-258.

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience, 25*, 339-379.

Alfvin, R. L., & Fairchild, M. D. (1997). Observer variability in metameric color matches using color reproduction media. *Color Research and Application, 22*(3), 174-188.

Amano, K., Foster, D. H., & Nascimento, S. M. C. (2006). Color constancy in natural scenes with and without an explicit illuminant cue. *Visual Neuroscience, 23*(3-4), 351-356.

Amano, K., Uchikawa, K., & Kuriki, I. (2002). Characteristics of color memory for natural scenes. *Journal of the Optical Society of America, 19*(8), 1501-1514.

Arend, L. E. (1993). How much does illuminant color affect unattributed colors? *Journal of the Optical Society of America A, 10*(10), 2134-2147.

Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A, 3*, 1743-1751.

Arend, L. E., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: papers with diverse Munsell values. *Journal of the Optical Society of America A, 8*(4), 661-672.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivaiton* (Vol. 2). New York: Academic Press.

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Advances in learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Bahrick, H. P. (2000). Long-term maintenance of knowledge. In E. Tulvin & R. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 347-362). New York: Oxford University Press.

Bartleson, C. J. (1960). Memory colors of familiar objects. *Journal of the Optical Society of America., 50*(1), 73-77.

Bäuml, K.-H. (1999). Color constancy: the role of image surfaces in illuminant adjustment. *Journal of the Optical Society of America A, 16*(7), 1521-1530.

Berlin, B., & Kay, P. (1969). *Basic color terms. Their universality and evolution*. Berkely and Los Angeles, California: University of California Press.

Berns, R. S. (1996). Methods for characterizing CRT displays. *Displays, 16*(4), 173-182.

Berns, R. S., & Choh, H.-K. (1995). Cathode-ray-tube to reflection-print matching under mixed chromatic adaptation using RLAB. *Journal of Electronic Imaging, 04*(04), 347-359.

Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature, 402*, 877-879.

Bodrogi, P., & Tarczali, T. (2001). Colour memory for various sky, skin, and plant colours: effect of the image context. *Color Research and Application, 26*(4), 278-289.

Bonnardel, V. (2006). Color naming and categorization in inherited color vision deficiencies. *Visual Neuroscience, 23*(3-4), 637-643.

Bornstein, M. H. (1976). Name codes and color memory. *American Journal of Psychology, 89*(2), 269-279.

Bowmaker, J. K., & Dartnall, H. J. (1980). Visual pigments of rods and cones in a human retina. *Journal of Physiology, 298*, 501-511.

Boyaci, H., Doerschner, K., & Maloney, L. T. (2004). Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *Journal of Vision, 4*(9), 664-679.

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision, 3*(8), 541-553.

Boynton, R. M., Fargo, L., Olson, C. X., & Smallman, H. S. (1989). Category effects in color memory. *Color Research & Application, 14*(5), 229-234.

Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Colour Research and Applicaiton, 12*(2), 94-105.

Boynton, R. M., & Olson, C. X. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research, 30*(9), 1311-1317.

Brainard, D. H. (1998). Color constancy in the nearly natural image.2. Achromatic loci. *Journal of Optical Society Of America A, 15*(2), 307-325.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. I. Asymmetric matches. *Journal of Optical Society Of America A, 14*(9), 2091-2110.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A, 14*(7), 1393-1411.

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color matching: how color appearance depends on the illuminant. *Journal of the Optical Society of America A, 9*, 1433-1448.

Braun, K. M., & Fairchild, M. D. (1996). *Psychophysical generation of matching images for cross-media color reproduction.* Paper presented at the 4th Color Imaging Conference: Color science, systems and applications, Scottsdale.

Braun, K. M., Fairchild, M. D., & Alessi, P. J. (1996). Viewing techniques for cross-media image comparisons. *Color Research and Application, 21*(1), 6-17.

Brunswik, E. (1928). Zur Entwicklung der Albedowahrnehmung. *Zeitschrift für Psychologie, 109*, 40-115.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience, 21*, 47-74.

Cambridge Research Systems. (2008). Retrieved 06 August 2008, from http://www.crsltd.com/catalog/colorcal/specifications.html

Chichilnisky, E. J., & Wandell, B. A. (1995). Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching. *Vision Research, 35*(2), 239-254.

Chichilnisky, E. J., & Wandell, B. A. (1999). Trichromatic opponent color classification. *Vision Research, 39*(20), 3444-3458.

Cole, B. L., Lakkis, C., Lian, K.-Y., & Sharpe, K. (2006). Categorical color naming of surface color codes by people with abnormal color vision. *Optometry and Vision Science, 83*(12), 879-886.

Collins, M. (1932). Some observations on immediate colour memory. *British Journal of Psychology, 22*, 344-352.

Cornelissen, F. W., & Brenner, E. (1995). Simultaneous color constancy revisited - an analysis of viewing strategies. *Vision Research, 35*(17), 2431-2448.

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research, 32*(7), 1359-1366.

Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E., & Kalina, R. E. (1987). Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science, 236*(4801), 579-582.

D'Ath, P. J., Thomson, W. D., & Wilkins, A. J. (2007). Memory for the color of non-monochromatic lights. *Color Research and Application, 32*(1), 11-15.

D'Zmura, M. (1992). Color constancy: Surface color from changing illumination. *Journal of the Optical Society of America, 9*(3), 490-493.

D'Zmura, M., & Iverson, G. (1993a). Color constancy: I. Basic theory of two-stage linear recovery of spectral desriptions for lights and surfaces. *Journal of the Optical Society of America A, 10*, 2148-2165.

D'Zmura, M., & Iverson, G. (1993b). Colour Constancy II Results for two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of Optical Society of America A, 10*(10), 2166-2180.

D'Zmura, M., & Lennie, P. (1986). Mechanisms of color constancy. *Journal of the Optical Society of America A, 3*(10), 1662-1672.

Daugirdiene, A., Murray, I. J., Vaitkevicius, H., & Kulikowski, J. (2006). Cone contrast computations: Physical versus perceived background and colour constancy. *Spatial Vision, 19*(2-4), 173-192.

Davidoff, J., Goldstein, J., & Roberson, D. (2009). Nature versus nurture: The simple contrast. *Journal of Experimental Child Psychology, 102*(2), 246-250.

de Almeida, V. M. N., Fiadeiro, P. T., & Nascimento, S. M. C. (2004). Color constancy by asymmetric color matching with real objects in three-dimensional scenes. *Visual Neuroscience, 21*(3), 341-345.

de Almeida, V. M. N., Fiadeiro, P. T., Nascimento, S. M. C., & Foster, D. H. (2002). Colour constancy under illuminant changes with 3-D and 2-D views of real scenes. *Perception, 31*(Suppl), 135.

de Fez, M. D., Capilla, P., Luque, M. J., & Perez-Carpinell, J. (2001). Asymmetric Colour Matching: Memory Matching Versus Simultaneous Matching. *Color Research and Application, 26*(6), 458-468.

De Valois, R. L., Abramov, I., & Jacobs, G. H. (1966). Analysis of response patterns of LGN cells. *Journal of the Optical Society of America, 56*(7), 966-977.

De Valois, R. L., & De Valois, K. K. (1993). A multi-stage color model. *Vision Research, 33*(8), 1053-1065.

De Valois, R. L., De Valois, K. K., Switkes, E., & Mahon, L. (1997). Hue scaling of isoluminant and cone-specific lights. *Vision Research, 37*(7), 885-897.

Delahunt, P. B., & Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision, 4*(2), 57-81.

Derrington, A. (2001). The lateral geniculate nucleus. *Current Biology, 11*(16), R635-R637.

Derrington, A., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nubleus of macaque. *Journal of Physiology, 357*, 241-265.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2004). Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *Journal of Vision, 4*(2), 92-105.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2007). Testing limits on matte surface color perception in three-dimensional scenes with complex light fields. *Vision Research, 47*(28), 3409-3423.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the U S A, 104*(3), 1097-1102.

Epps, H. H., & Kaya, N. (2004). *Color matching from memory.* Paper presented at the AIC 2004 Color and Paints, Interim Meeting of the International Color Association, Porto Alegre, Brazil.

Erens, R. G. F., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics, 54*(2), 145-156.

Fagot, J., Goldstein, J., Davidoff, J., & Pickering, A. (2006). Cross-species differences in color categorization. *Psychonomic Bulletin & Review, 13*(2), 275-280.

Fairchild, M. D. (2005). *Color Appearance Models* (2nd ed.). Chichester, West Sussex, England: John Wiley & Sons, Ltd.

Fairchild, M. D., & Lennie, P. (1992). Chromatic adaptation to natural and incandescent illuminants. *Vision Research, 32*(11), 2077-2085.

Fairchild, M. D., & Reniff, L. (1995). Time course of chromatic adaptation for color-appearance judgements. *Journal of Optical Society of America A, 12*(5), 824-833.

Farnsworth, D. (1957). *The Farnsworth-Munsell 100-Hue Test for the examination of color discrimination* (revised ed.). Baltimore: Munsell Color Company.

Fonteneau, E., & Davidoff, J. (2007). Neural correlates of colour categories. *Neuroreport, 18*(13), 1323-1327.

Foster, D. H. (2003). Does colour constancy exist? *Trends in Cognitive Sciences, 7*(10), 439-443.

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2001). Colour constancy from temporal cues: better matches with less variability under fast illuminant changes. *Vision Research, 41*, 285-293.

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2006). Color constancy in natural scenes explained by global image statistics. *Visual Neuroscience, 23*(3-4), 341-349.

Foster, D. H., & Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings. Biological Sciences/The Royal Society, 257*(1349), 115-121.

Francis, M. A., & Irwin, R. J. (1998). Stability of memory for colour in context. *Memory, 6*(6), 609-621.

Franklin, A., & Davies, I. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology, 22*, 349-377.

Franklin, A., Wright, O., & Davies, I. (2009). What can we learn from toddlers about categorical perception of color? *Journal of Experimental Child Psychology, 102*(2), 239-245.

Funt, B. V., & Drew, M. S. (1993). Color space analysis of mutual illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*, 1319-1326.

Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience, 4*(7), 563-572.

Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science, 195*(4274), 185-187.

Gilchrist, A. L. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception & Psychophysics, 28*(6), 527-538.

Granville, W. C., & Jacobson, E. (1944). Colorimetric specification ot the Color Harmony Manual from spectrophotometric measurements. *Journal of the Optical Society of America, 34*(7), 382-393.

Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009a). Do people match surface reflectance fundamentally differently than they match emitted light? *Vision Research,49*(7), 702-707.

Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009b). Reliable identification by color under natural conditions. *Journal of Vision, 9*(1), 1-8.

Granzier, J. J. M., Smeets, J. B. J., & Brenner, E. (2006). *A direct test of the "Grey World Hypothesis".* Paper presented at the CGIV Leeds, UK.

Hamwi, V., & Landis, C. (1955). Memory for color. *Journal of Psychology: Interdisciplinary and Applied, 39*, 183-194.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367-1368.

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4), 1-15.

Hård, A., Sivik, L., & Tonnquist, G. (1996a). NCS, Natural Color System - from concept to research and application. Part 1. *Color Research & Application, 21*(3), 180-205.

Hård, A., Sivik, L., & Tonnquist, G. (1996b). NCS, Natural Color System - from concept to research and application. Part 2. *Color Research & Application, 21*(3), 206-220.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology, 93*(1), 10-20.

Hendry, S. H., & Reid, R. C. (2000). The koniocellular pathway in primate vision. *Annual Review of Neuroscience, 23*, 127-153.

Holmes, A., Franklin, A., Clifford, A., & Davies, I. (2009). Neurophysiological evidence for categorical perception of color. *Brain and Cognition, 69*(2), 426-434.

Hseue, T.-C., Shen, Y.-C., Chen, P.-C., Hsu, W.-H., & Liu, Y.-T. (1998). Cross-media performance evaluation of color models for unequal luminance levels and dim surround. *Color Research and Application, 23*(3), 169-177.

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual-cortex. *Proceedings of the Royal Society of London Series B-Biological Sciences, 198*(1130), 1-59.

Hunt, R. W. G. (1950). The effects of daylight and tungsten light-adaptation on color perception. *Journal of the Optical Society of America, 40*(6), 362-371.

Hurlbert, A. (1998). Computational models of color constancy. In V. Walsh & J. Kulikowski (Eds.), *Perceptual Constancy: Why things look as they do* (pp. 283-321). Cambridge: Cambridge University Press.

Hurlbert, A., & Ling, Y. (2006). Contextual effects of familiar object colours on colour perception. *Perception, 35*(Suppl), 23.

Hurlbert, A., & Wolf, K. (2004). Color contrast: a contributory mechanism to color constancy. *Progress in Brain Research, 144*, 147-160.

Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review, 64*(6), 384-404.

Ikeda, T., & Osaka, N. (2007). How are colors memorized in working memory? A functional magnetic resonance imaging study. *Neuroreport, 18*(2), 111-114.

International Color Consortium. (1996). Retrieved 19.01.2009, from http://www.color.org/sRGB.xalter

Jameson, D., & Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology, 40*, 1-22.

Jameson, D., Hurvich, L. M., & Varner, F. D. (1979). Receptoral and postreceptoral visual prosesses in recovery from chromatic adaptation. *Proceeding of the National Academy of Sciences of the USA, 76*(6), 3034-3038.

Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A, 13*(10), 1981-1991.

Johnston, A., & Curran, W. (1996). Investigating shape-from-shading illusions using solid objects. *Vision Research, 36*(18), 2827-2835.

Kaiser, P. K., & Boynton, R. M. (1996). *Human Color Vision* (2 ed.). Washington, DC: Optical Society of America.

Katoh, N. (1994). *Practical method for appearance match between soft copy and hard copy.* Paper presented at the SPIE, San Jose, CA, USA.

Kaufmann, R., & O'Neill, M. C. (1993). Colour names and focal colours on electronic displays. *Ergonomics, 36*(8), 881-890.

Kay, P., Berlin, B., & Merrifield, W. (1991). Bi-cultural implications of systems of color naming. *Journal of Linguistic Anthropology, 1*(1), 12-25.

Kay, P., & Kuehni, R. G. (2008). Why colour words are really ... colour words. *Journal of the Royal Anthropological Institute, 14*(4), 886-887.

Kay, P., & Regier, T. (2006). Language, thought and color: Recent developments. *Trends in Cognitive Sciences, 10*(2), 51-54.

Kennard, C., Lawden, M., Morland, A. B., & Ruddock, K. H. (1995). Colour identification and colour constancy are impaired in a patient with incomplete achromatopsia associated with prestriate cortical lesions. *Proceedings of the Royal Society of London Series B - Biological Sciences, 260*(1358), 169-175.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences USA, 96*, 307-312.

Kraft, J. M., Maloney, S. I., & Brainard, D. H. (2002). Surface-illuminant ambiguity and color constancy: Effects of scene complexity and depth cues. *Perception, 31*(2), 247-263.

Kuehni, R. G. (2002). *Uniform color space is not homogeneous.* Paper presented at the AIC: 9th Congress of the International Colour Assosiation.

Kuehni, R. G. (2003). *Color space and its divisions*. Hoboken, New Jersey: John Wiley & Sons.

Kulikowski, J. J., & Vaitkevicius, H. (1997). Colour constancy as a function of hue. *Acta Psychologica, 97*, 25-35.

Lee Filters. (2008). Retrieved April 2008, from http://www.leefilters.com/

Lee, H. C. (1986). Method for computing the scene-illuminant chromaticity from specular highlights. *Journal of the Optical Society of America 3*(10), 1694-1699.

Lindbloom, B. (1989). *Accurate color reproduction for computer craphics applications.* Paper presented at the SIGGGRAPH, Boston.

Ling, Y., & Hurlbert, A. (2006). *Colour-memory-dependent colour constancy: 2D vs 3D real surfaces.* Paper presented at the CGIV 2006, Leeds, UK.

Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of Optical Society Of America A, 25*(6), 1215-1226.

Livens, S., Anthonis, A., Mahy, M., & Scheunders, P. (2003). A cross media tonal mapping model obtained from psychometric experiments. *Color Imaging VIII: Processing, Hardcopy, and Applications, 5008*, 14-23.

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience, 7*(11), 3416-3468.

Lo, M.-C., Luo, M. R., & Rhodes, P. A. (1996). Evaluating colour models' performance between monitor and print images. *Color Research & Application, 21*(4), 277-291.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition, 5*(6), 696-699.

Lucassen, M. P., & Walraven, J. (1996). Color constancy under natural and artificial illumination. *Vision Research, 36*(17), 2699-2711.

Lucy, J. A., & Shweder, R. A. (1979). Whorf an his critics: Linguistic and nonlinguistic influences on color memory. *American Anthropologist, 81*(3), 581-615.

MacAdam, D. L. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America, 32*(5), 247-274.

MacAdam, D. L. (1974). Uniform color scales. *Journal of the Optical Society of America, 64*(12), 1691-1702.

MacAdam, D. L. (1978). Colorimetric data for samples of OSA uniform color scales. *Journal of the Optical Society of America A, 68*(1), 121-130.

MacLeod, D. I. A., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America, 69*(8), 1183-1186.

Mahy, M., Eycken, L. V., & Oosterlinck, A. (1994). Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV. *Color Research & Application, 19*(2), 105-121.

Maloney, L. T. (1999). Physics-based approaches to modeling color perception. In B. Boycott, Gegenfurtner KR and Sharpe LT (Ed.), *Color vision, from genes to perception.* (pp. 387-422): Cambridge University Press.

Maloney, L. T. (2002). Illuminant estimation as cue combination. *Journal of Vision, 2*(6), 493-504.

Mandic, L., Delac, K., & Grgic, M. (2007). *Corresponding colors: Visual and predictive data.* Paper presented at the ELMAR 2007, Zadar.

Mandic, L., Grgic, S., & Grgic, M. (2006). *Comparison of color difference equations.* Paper presented at the Proceedings ELMAR-2006.

Marks, W. B., Dobelle, W. H., & MacNichol, E. F. Jr. (1964). Visual pigments of single primate cones. *Science, 143*, 1181-1183.

McCann, J. J., McKee, S. P., & Taylor, T. H. (1976). Quantitative studies in retinex theory a comparison between theoretical predictions and observer responses to the "color mondrian" experiments. *Vision Research, 16*(5), 445-458.

Montero, V. M. (1991). A quantitative study of synaptic contacts on interneurons and relay cells of the cat lateral geniculate nucleus. *Experimental Brain Research, 86*(2), 257-270.

Moroney, N., Fairchild, M. D., Hunt, R. W. G., Li, C., Luo, M. R., & Newman, T. (2002). The CIECAM02 colour appearance model. *Proceedings of IS&T/SID 10th Color Imaging Conference, Scottsdale, Arizona*, 298-204.

Murray, I. J., Daugirdiene, A., Stanikunas, R., Vaitkevicius, H., & Kulikowski, J. J. (2006a). Cone contrasts do not predict color constancy. *Visual Neuroscience, 23*(3-4), 543-547.

Murray, I. J., Daugirdiene, A., Vaitkevicius, H., Kulikowski, J. J., & Stanikunas, R. (2006b). Almost complete colour constancy achieved with full-field adaptation. *Vision Research, 46*(19), 3067-3078.

Naka, K. I., & Rushton, W. A. (1966). S-potentials from colour units in the retina of fish (Cyprinidae). *Journal of Physiology, 185*(3), 536-555.

Nascimento, S. M. C., de Almeida, V. M., Fiadeiro, P. T., & Foster, D. H. (2005). Effect of scene complexity on colour constancy with real three-dimensional scenes and objects. *Perception, 34*, 947-950.

Nascimento, S. M. C., Ferreira, F. P., & Foster, D. H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A, 19*(8), 1484-1490.

Nascimento, S. M. C., & Foster, D. H. (1997). Detecting natural changes of cone-excitation ratios in simple and complex coloured images. *Proceedings of the Royal Society London B, 264*, 1395-1402.

Nathans, J., Darcy, T., & Hogness, D. S. (1986). Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science, 232*(4747), 193-202.

Nayar, S. K., Ikeuchi, K., & Kanade, T. (1991). Shape from interreflections. *International Journal of Computer Vision, 6*, 173-195.

Nerger, J. L., Volbrecht, V. J., & Ayde, C. J. (1995). Unique hue judgments as a function of test size in the fovea and at 20-deg temporal eccentricity. *Journal of the Optical Society of America, 12*(6), 1225-1232.

Newhall, S. M., Burnham, R. W., & Clark, J. R. (1957). Comparison of successive with simultaneous color matching. *Journal of the Optical Society of America, 47*(1), 43-56.

Nickerson, D. (1981). OSA uniform color scale samples: A unique set. *Color Research & Application, 6*(1), 7-33.

Nilsson, T. H., & Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 7*(1), 141-150.

Oicherman, B., Luo, M. R., Rigg, B., & Robertson, A. R. (2008). Effect of observer metamerism on colour matching of display and surface colours. *Color Research & Application, 33*(5), 346-359.

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision, 8*(5), 1-16.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica, Suppl. 6*(1), 11-97.

Perez-Carpinell, J., Baldovi, R., De Fez, M. D., & Castro, J. (1998a). Color memory matching: time effect and other factors. *Colour Research and Application, 23*(4), 234-247.

Perez-Carpinell, J., De Fez, M. D., Baldovi, R., & Soriano, J. C. (1998b). Familiar objects and memory color. *Color Research & Application, 23*(6), 416-427.

Perry, V. H., Oehler, R., & Cowey, A. (1984). Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. *Neuroscience, 12*(4), 1101-1123.

Photo Research, Inc. (2008). Retrieved 20 August 2008, from http://www.photoresearch.com/

Pobboravsky, I. (1988). *Effect of small color differences in color vision on the matching of soft and hard proofs.* Paper presented at the TAGA.

Pointer, M. R. (1981). A comparison ot the CIE 1976 color spaces. *Color Research & Application, 6*(2), 108-118.

Pointer, M. R., & Attridge, G. G. (1998). The number of discernible colors. *Colour Research and Application, 23*(1), 52-54.

Pointer, M. R., Attridge, G. G., & Jacobson, R. E. (2002). Perceived colour differences in displayed colours Part 1: hard copy to soft copy matching. *Imaging Science Journal, 50*(1), 1-9.

Pokorny, J., Lutze, M., Cao, D., & Zele, A. J. (2006). The color of night: Surface color perception under dim illuminations. *Visual Neuroscience, 23*(3-4), 525-530.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*, 163-166.

Ratner, C., & McCarthy, J. (1990). Ecologically relevant stimuli and color memory. *Journal of General Psychology, 117*(4), 369-377.

Reeves, A. J., Amano, K., & Foster, D. H. (2008). Color constancy: Phenomenal or projective? *Perception & Psychophysics, 70*(2), 219-228.

Regier, T., Cook, R. S., & Kay, P. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the U S A, 102*(23), 8386-8391.

Regier, T., Kay, P., & Cook, R. S. (2005). *Universal foci and varying boundaries in lumguistic color categories.* Paper presented at the 27th Annual Conference of the Cognitive Science Society, In B. G. Bara, L. Barsalou and M. Bucciarelli (Eds.)

Rich, D. C., & Jalijali, J. (1995). Effects of observer metamerism in the determination of human color-matching functions. *Color Research & Application, 20*(1), 29-35.

Rinner, O., & Gegenfurtner, K. R. (2000). Time course of chromatic adaptation for color appearance and discrimination. *Vision Research, 40*(14), 1813-1826.

Roberson, D., Davidoff, J., Davies, I., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50*(4), 378-411.

Roberson, D., Davies, I. R. L., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology, 129*(3), 369-398.

Roberson, D., & Hanley, J. R. (2007). Color vision: Color categories vary with language after all. *Current Biology, 17*(15), R605-607.

Roberson, D., Shapiro, L. R., Davidoff, J., & Davies, I. R. L. (2004). The development of color categories in two languages: a longitudinal study. *Journal of Experimental Psychology: General, 133*(4), 554-571.

Romero, J., Hita, E., & del Barco, J. (1986). A comparative study of successive and simultaneous methods in colour discrimination. *Vision Research, 26*(3), 471-476.

Ruppertsberg, A. I., & Bloj, M. (2008). Creating physically accurate visual stimuli for free: Spectral rendering with RADIANCE. *Behavior Research Methods, 40*(1), 304-308.

Ruttiger, L., Mayser, H., Serey, L., & Sharpe, L. T. (2001). The color constancy of the red-green color blind. *Color Research and Application, 26*(Suppl), 209-213.

Sachtler, W. L., & Zaidi, Q. (1992). Chromatic and luminance signals in visual memory. *Journal of the Optical Society of America A, 9*(6), 877-894.

Scandinavian Colour Institute: Natural Colour System. (2004). Retrieved April 2008, from http://www.ncscolour.com/webbizz/mainPage/main.asp

Schiller, P., Sandell, J., & Maunsell, J. (1986). Functions of the ON and OFF channels of the visual system. *Nature, 322*, 824-825.

Schirillo, J., Reeves, A., & Arend, L. (1990). Perceived lightness, but not brightness, of achromatic surfaces depends on perceived depth information. *Perception & Psychophysics, 48*(1), 82-90.

Schultz, S., Doerschner, K., & Maloney, L. T. (2006). Color constancy and hue scaling. *Journal of Vision, 6*(10), 1102-1116.

Seliger, H. H. (2001). Measurement of memory of color. *Color Research and Application, 27*(4), 233-242.

Shepard, R. N. (1992). The perceptual organization of colors: An adaptation to regularities of the terrestrial world? In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 495-532). New York: Oxford University Press.

Shevell, S. K., & Miller, P. R. (1996). Color perception with test and adapting lights perceived in different depth planes. *Vision Research, 36*(7), 949-954.

Shiraiwa, Y., Hidaka, Y., Mizuno, T., Sasaki, T., Ohta, K., & Usami, A. (1998). *Color appearance matching in hard-copy and soft-copy images in different office environments.* Paper presented at the SPIE, San Jose, CA, USA.

Siegel, S., & Castellan Jr, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Siple, P., & Springer, R. M. (1983). Memory and preference for the colors of objects. *Perception & Psychophysics, 34*(4), 363-370.

Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision, 4*(9), 693-710.

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society London B, 360*(1458), 1329-1346.

Smithson, H. E., Khan, S. S., Sharpe, L. T., & Stockman, A. (2006). Transitions between color categories mapped with a reverse Stroop task. *Visual Neuroscience, 23*(3-4), 453-460.

Song, T., & Luo, M. R. (2002). *Colour difference thresholds for cross-media colour image reproductions.* Paper presented at the AIC: 9th Congress of the International Colour Association, Rochester.

Speigle, J. M., & Brainard, D. H. (1996). Is color constancy task independent? *Proceedings of The Fourth Color Imaging Conference: Color Science, Systems and Applications, Soc. Imaging Science & Technology, Springfield*, 167-172.

Speigle, J. M., & Brainard, D. H. (1999). Predicting color from gray: the relationship between achromatic adjustment and asymmetric matching. *Journal of the Optical Society of America, 16*(10), 2370-2376.

Stockman, A., & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research, 40*(13), 1711-1737.

Tarczali, T., Park, D.-S., Bodrogi, P., & Kim, C. Y. (2006). Long-term memory colors of Korean and Hungarian observers. *Color Research and Application, 31*(3), 176-183.

Tate, J. D., & Springer, R. M. (1971). Effects of memory time on successive judgments. *Psychological Bulletin, 76*(6), 394-408.

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology A, 58*(5), 931-960.

Troost, J. M., & de Weert, C. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics, 50*(6), 591-602.

Uchikawa, H., Uchikawa, K., & Boynton, R. M. (1989). Influence of achromatic surrounds on categorical perception of surface colors. *Vision Research, 29*(7), 881-890.

Uchikawa, K. (1983). Purity discrimination: successive vs simultaneous comparison method. *Vision Research, 23*, 53-58.

Uchikawa, K., & Shinoda, H. (1996). Influence of basic color categories on color memory discrimination. *Color Research and Application, 21*(6), 430-439.

Uchikawa, K., Uchikawa, H., & Boynton, R. M. (1989). Partial color constancy of isolated surface colors examined by a color-naming method. *Perception, 18*(1), 83-91.

Wandell, B. A. (1995). *Foundations of Vision*. Sunderland, Massachusetts: Sinauer Associates, Inc. Publishers.

Ward Larson, G., & Shakespeare, R. (1998). *Rendering with Radiance: The Art and Science of Lighting Visualization*. San Francisco: Morgan Kaufmann Publishers.

Webster, M. A. (1996). Human colour perception and its adaptation. *Network-Computation in Neural Systems, 7*(4), 587-634.

Webster, M. A., & Mollon, J. D. (1995). Colour constancy influenced by contrast adaptation. *Nature, 373*(6516), 694-698.

Webster, M. A., Webster, S. M., Malkoc, G., & Bilson, A. C. (2002). Color contrast and contextual influences on color appearance. *Journal of Vision, 2*(6), 505-519.

Werner, A. (2006). The influence of depth segmentation on colour constancy. *Perception, 35*(9), 1171-1184.

Werner, A., Sharpe, L. T., & Zrenner, E. (2000). Asymmetries in the time-course of chromatic adaptation and the significance of contrast. *Vision Research, 40*(9), 1101-1113.

Westland, S., & Ripamonti, C. (2004). *Computational Colour Science using MATLAB* (1st ed.). Chichester: John Wiley & Sons.

Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 509-520.

Wright, W. D. (1946). *Researches on normal and defective colour vision*. St. Louis: C. V. Mosby.

Wu, R. C., Wardman, R. H., & Luo, M. R. (2005). A comparison of lightness contrast effects in CRT and surface colours. *Colour Research and Application, 30*(1), 13-20.

Wyszecki, G. (1986). Color Appearance. In K. R. Boff, L. Kaufmann & J. P. Thomas (Eds.), *Handbook of Perception* (Vol. 1, pp. 9-1 - 9-57). New York: John Wiley & Sons.

Wyszecki, G., & Stiles, W. S. (2000). *Color Science* (2 ed.). New York: John Wiley & Sons.

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research, 41*, 2581-2600.

Yang, J. N., & Shevell, S. K. (2002). Stereo disparity improves color constancy. *Vision Research, 42*, 1979-1989.

Yang, J. N., & Shevell, S. K. (2003). Surface color perception under two illuminants: The second illuminant reduces color constancy. *Journal of Vision, 3*(5), 369-379.

Yendrikhovskij, S. N., Blommaert, F. J. J., & de Ridder, H. (1999). Representation of memory prototype for an object colour. *Color Research & Application, 24*(6), 393-410.

Zaidi, Q., & Bostic, M. (2008). Color strategies for object identification. *Vision Research, 48*(26), 2673-2681.

Zaidi, Q., Spehar, B., & DeBonet, J. (1997). Color constancy in variegated scenes: Role of low-level mechanisms in discounting illumination changes. *Journal of the Optical Society of America A, 14*, 2608-2621.

Zeki, S. (1983a). Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience, 9*(4), 741-765.

Zeki, S. (1983b). Colour coding in the cerebral cortex: The responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience, 9*(4), 767-781.

Appendix

Tables

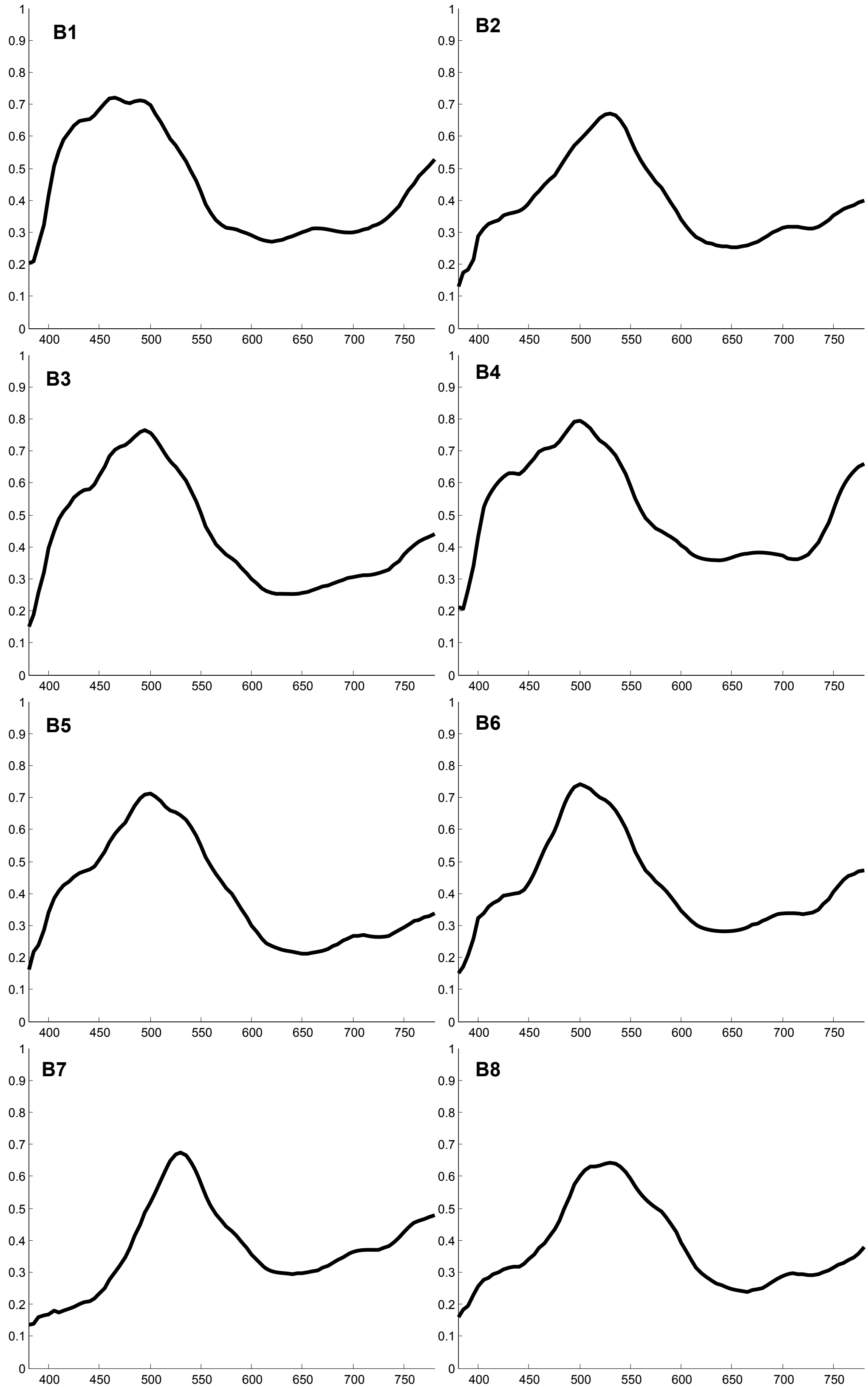


Table AP-1. Surface reflectance functions of the swatches B1 – B8; x-axis: wavelength [nm] and y-axis: relative radiance.

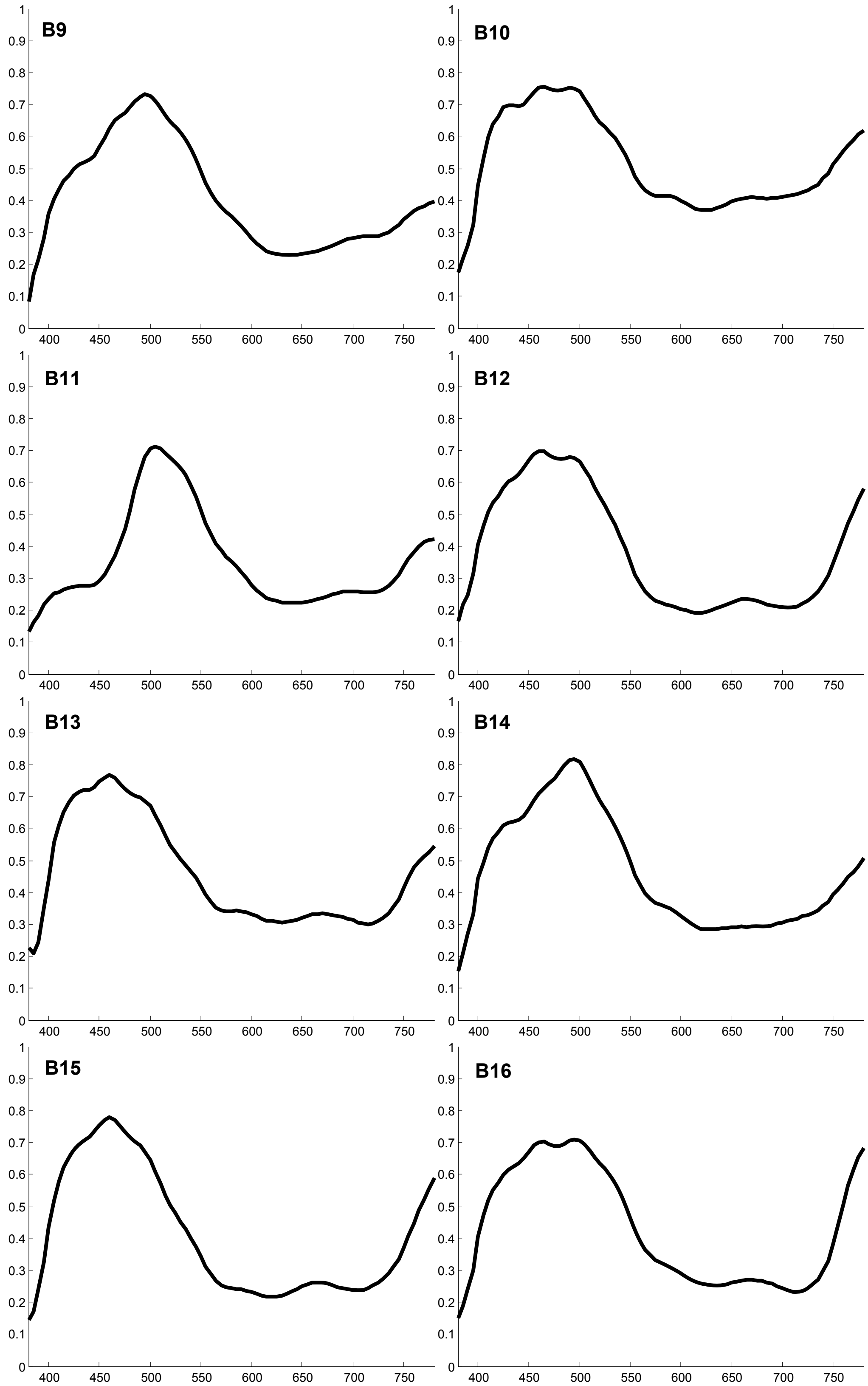


Table AP-1 – continued. Surface reflectance functions of the swatches B9 – B16.

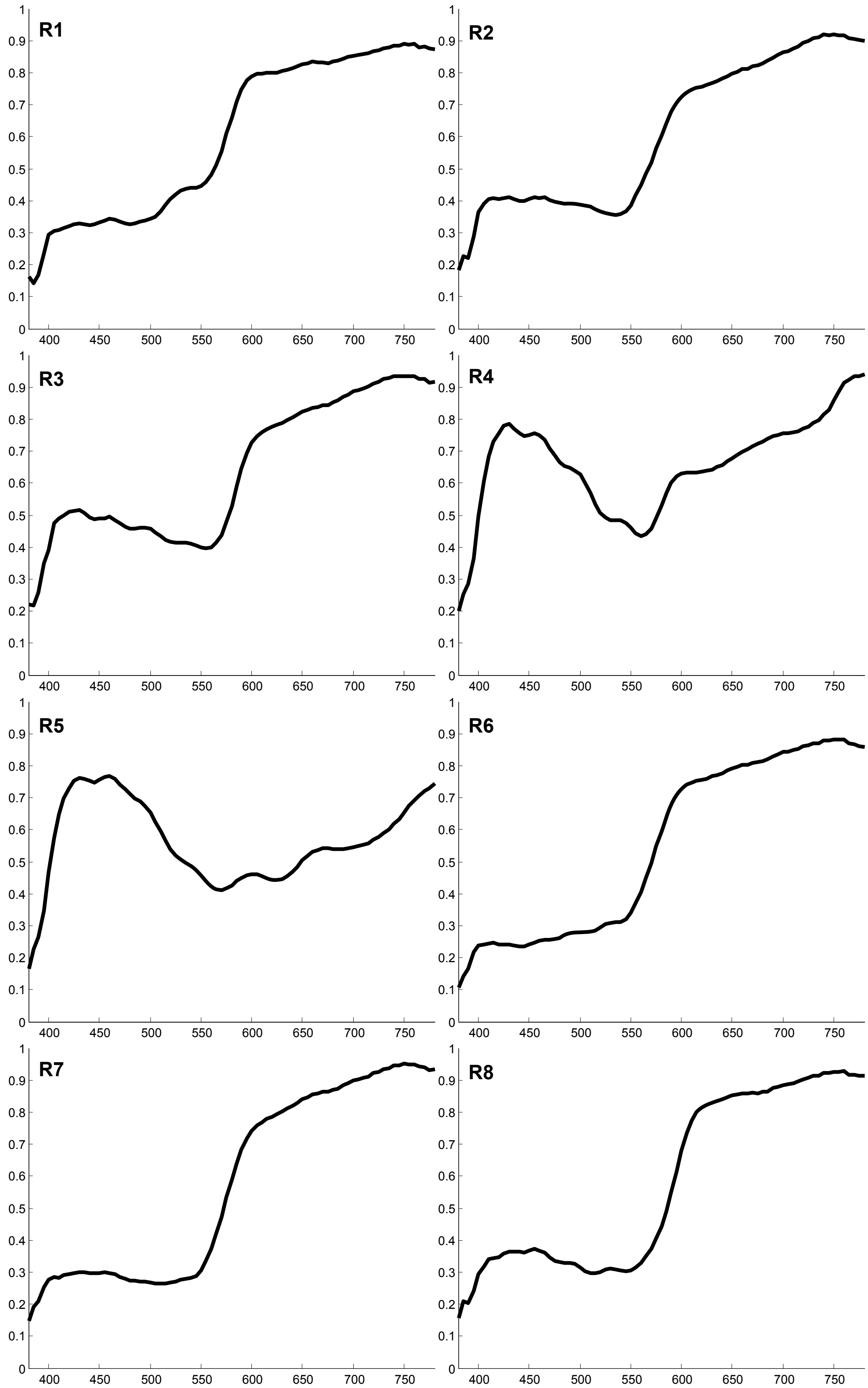


Table AP-1 – continued. Surface reflectance functions of the swatches R1 – R8.

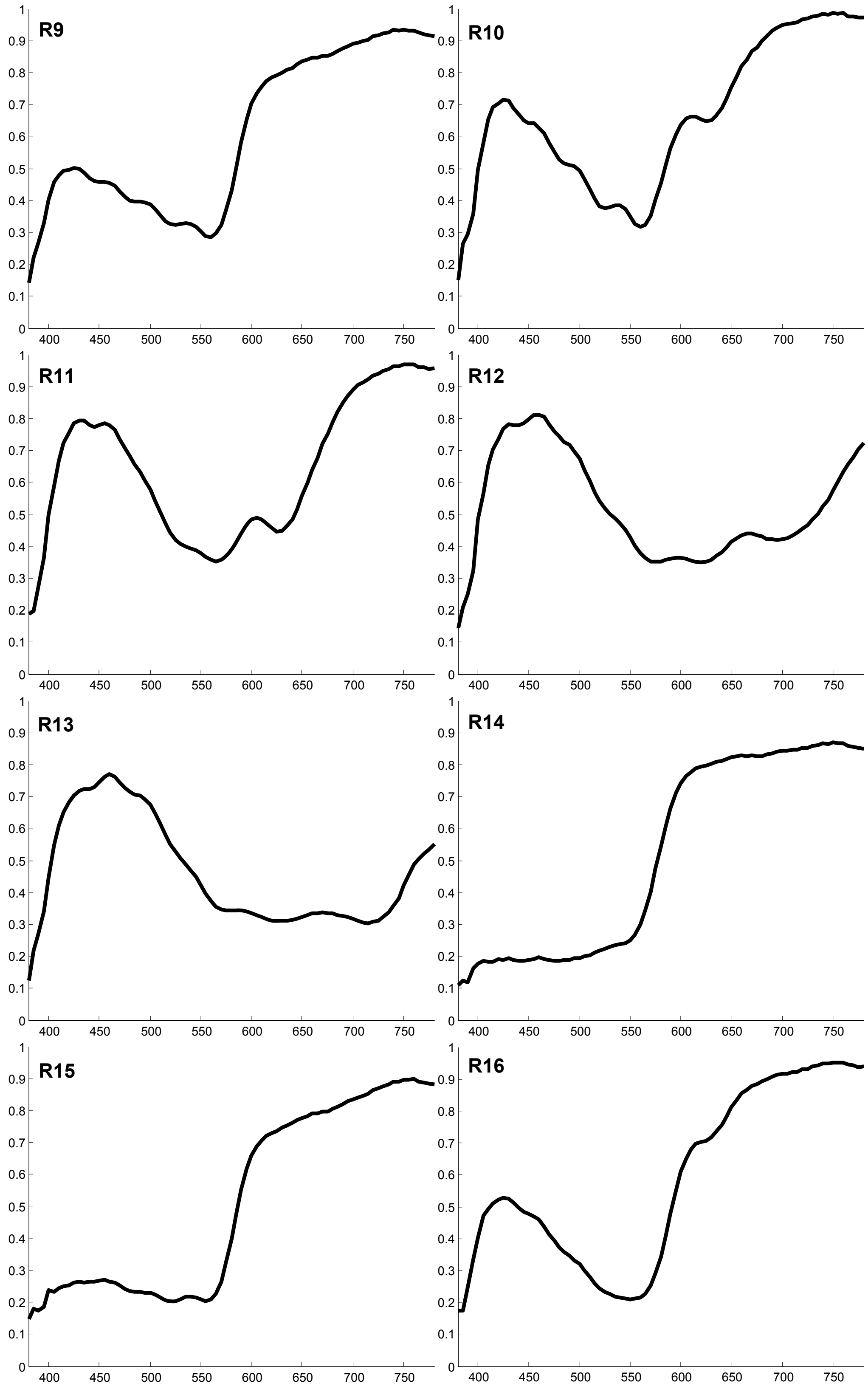


Table AP-1 – continued. Surface reflectance functions of the swatches R9 – R16.

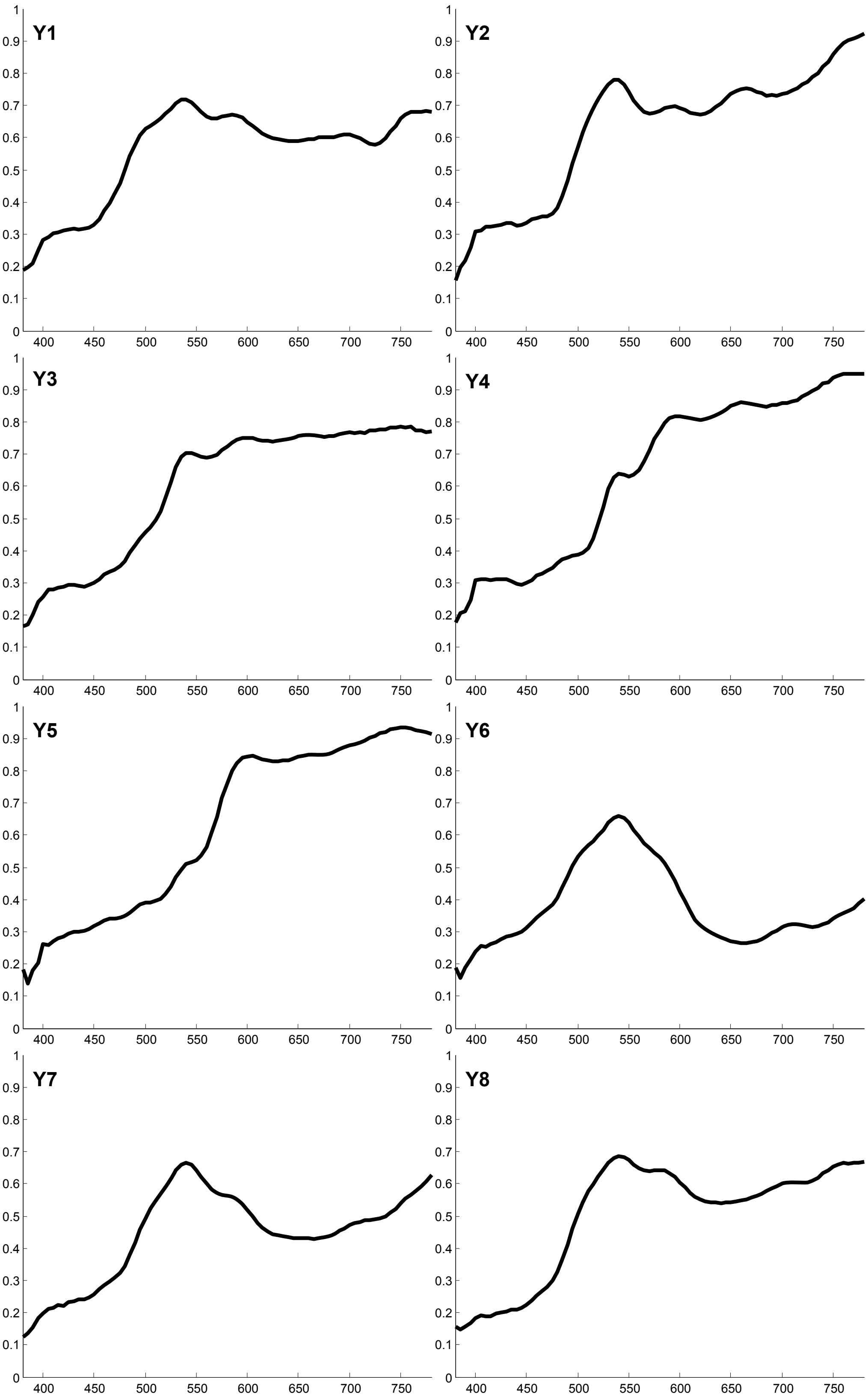


Table AP-1 – continued. Surface reflectance functions of the swatches Y1 – Y8.

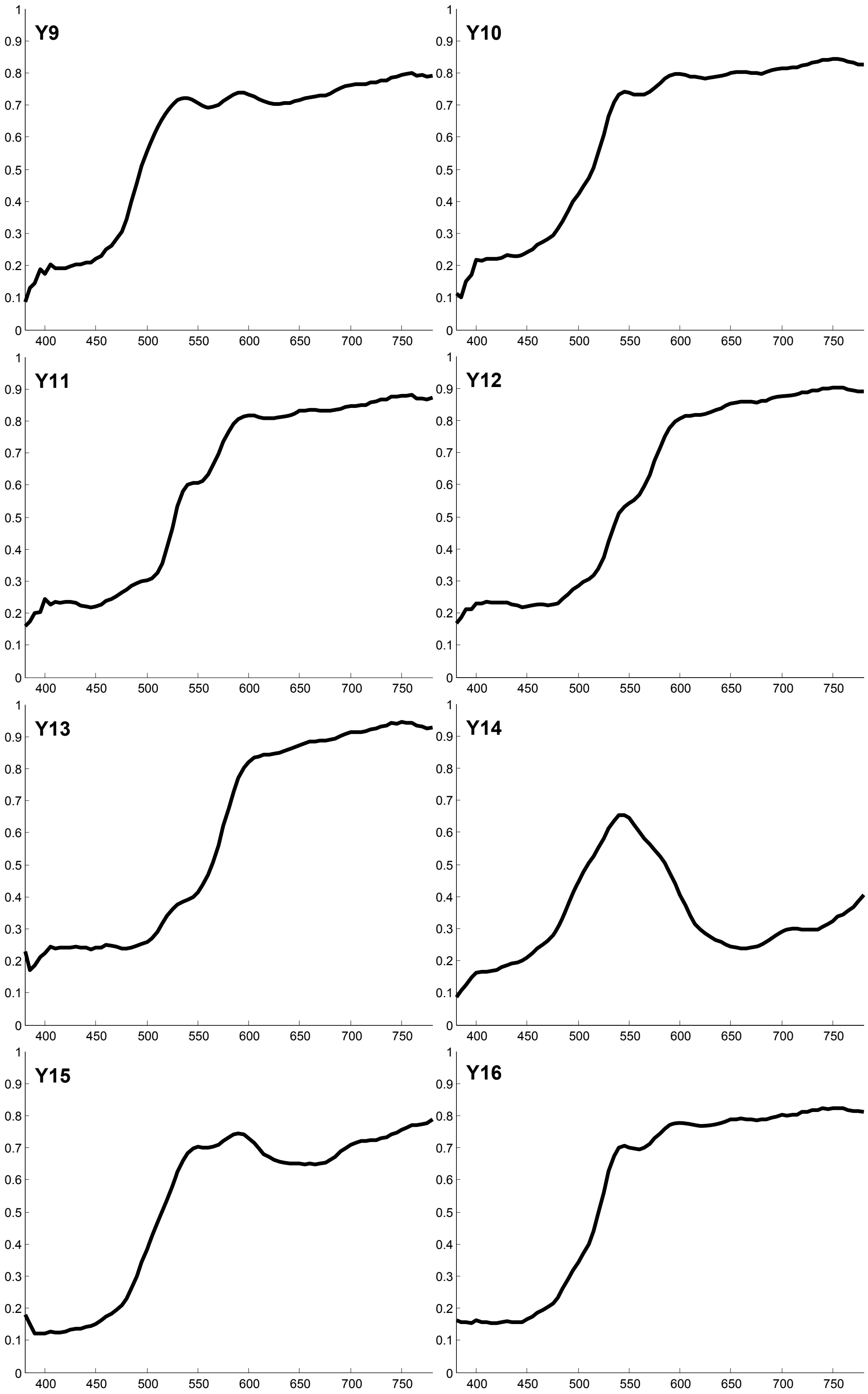


Table AP-1 – continued. Surface reflectance functions of the swatches Y9 – Y16.

Learning palettes		
1	**2**	**3**
B9	B9	B9
B16	B16	B16
R8	R8	R8
R10	R10	R10
Y8	Y8	Y8
Y10	Y10	Y10
R5	R6	R4
B2	B10	R12
B10	Y14	R16
Y4	R5	R14
R12	Y2	B2
Y11	Y11	Y4
Y15	Y4	Y14
R14	Y6	R6
R16	R16	Y15
Y2	B2	B10

Table AP-2. Composition of the learning palettes. Each learning palette consisted of the six target colours (first six rows) and ten other colours selected from all three colour groups.

AP-3	blue

AP-3 red

AP-3	yellow

D1		Tun		D2		Lily	
Label	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**
9	0.00	9	0.00	9	0.00	9	0.00
3	3.15	3	2.33	3	2.79	3	2.94
14	6.34	14	6.12	14	6.20	14	6.12
5	8.41	16	6.54	16	8.03	16	7.90
16	8.57	5	8.34	5	8.55	5	7.96
4	10.39	1	12.42	4	11.01	4	12.16
1	15.14	4	12.55	1	14.04	1	14.07
6	15.80	6	15.39	6	15.91	6	16.46
10	16.87	10	15.77	10	16.13	10	16.77
12	18.52	13	17.24	12	18.16	12	17.47
2	22.29	12	17.68	13	20.64	13	19.71
13	22.81	11	19.42	2	22.88	2	22.39
11	24.87	15	22.19	11	23.37	11	23.52
15	27.12	2	23.19	15	25.29	15	24.12
8	29.04	8	29.77	8	29.70	8	29.31
7	38.76	7	37.90	7	38.83	7	40.10

D1		Tun		D2		Lily	
Label	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**
16	0.00	16	0.00	16	0.00	16	0.00
14	3.61	14	2.73	14	2.92	14	3.49
3	5.79	3	4.47	3	5.55	3	5.32
1	6.74	1	6.14	1	6.46	1	6.68
9	8.57	9	6.54	9	8.03	9	7.90
12	11.52	13	11.56	12	11.76	12	11.23
10	11.57	12	12.80	10	11.95	13	12.43
4	12.53	10	13.10	4	13.18	10	12.55
13	14.75	4	14.01	13	13.48	4	13.48
5	16.69	5	14.62	5	16.28	5	15.57
15	18.79	15	16.15	15	17.63	15	16.63
6	23.20	6	20.14	6	22.62	6	22.39
2	29.36	11	25.05	2	29.56	2	28.41
11	33.02	2	28.22	11	30.93	11	30.48
8	35.56	8	34.37	8	35.92	8	34.99
7	45.51	7	42.28	7	45.10	7	45.38

Table AP-4. Top - ΔE values for target colour B9. Bottom - ΔE values for target colour B16. Colour samples are sorted by their ΔE values for illuminant D1, Tun, D2 and Lily.

D1		Tun		D2		Lily	
Label	**∆ E**	**Label**	**∆ E**	**Label**	**∆ E**	**Label**	**∆ E**
8	0.00	8	0.00	8	0.00	8	0.00
2	10.05	15	8.93	15	9.56	15	8.26
15	10.12	2	10.24	2	10.10	2	9.35
9	10.49	9	11.02	9	10.77	7	10.49
7	11.63	7	11.36	7	11.40	9	10.96
3	13.19	3	14.06	3	13.34	3	13.66
6	19.09	1	16.69	1	18.29	1	15.55
1	19.28	16	19.04	6	19.89	6	18.83
16	21.39	6	19.90	16	20.73	16	19.25
14	24.16	14	23.07	14	22.32	14	22.98
10	27.02	10	29.37	10	27.74	10	29.43
4	35.20	4	38.31	4	36.21	4	38.35
11	45.94	11	51.97	11	47.91	11	50.94
5	49.29	5	54.69	5	51.05	5	53.97
12	59.11	12	66.90	12	61.83	12	64.80
13	61.97	13	70.75	13	64.93	13	68.56

D1		Tun		D2		Lily	
Label	**∆ E**	**Label**	**∆ E**	**Label**	**∆ E**	**Label**	**∆ E**
10	0.00	10	0.00	10	0.00	10	0.00
4	12.98	4	12.12	4	12.59	4	12.06
9	17.22	3	17.95	9	17.70	16	16.24
16	18.21	16	18.10	16	17.85	3	18.33
3	18.32	9	19.20	3	18.13	9	18.81
11	20.06	11	23.48	11	21.19	11	22.24
2	27.00	2	26.75	2	26.81	2	26.61
8	27.02	5	29.06	8	27.74	5	27.82
5	27.25	8	29.37	5	28.15	8	29.43
15	35.07	15	36.97	15	35.33	15	36.52
12	36.17	1	38.83	12	37.46	12	38.27
7	38.09	7	39.45	7	38.28	1	39.06
1	39.87	12	40.15	1	39.25	7	39.12
13	40.51	13	44.85	13	41.72	13	43.15
6	43.79	6	46.89	6	45.26	6	46.09
14	51.04	14	52.07	14	49.76	14	52.13

Table AP-4 – continued. Top - R8. Bottom - R10.

D1		Tun		D2		Lily	
Label	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**
8	0.00	8	0.00	8	0.00	8	0.00
2	8.17	2	7.02	2	7.67	2	7.85
9	9.20	9	10.05	9	9.32	9	10.44
7	10.50	7	10.65	7	10.69	7	12.02
1	12.93	1	11.35	1	12.56	1	12.43
3	14.61	3	12.59	3	13.43	3	13.25
10	17.52	10	18.79	10	17.30	10	18.19
14	19.03	4	20.45	15	19.63	15	20.86
15	19.90	15	20.60	14	19.97	4	21.19
4	23.84	14	22.50	4	22.03	14	23.90
16	24.44	5	25.81	16	24.39	5	25.94
6	24.93	16	25.83	6	25.66	11	26.06
11	27.83	11	26.90	11	26.35	16	26.40
5	31.76	6	27.56	5	29.06	6	29.70
12	33.50	12	31.03	12	31.43	12	30.66
13	40.15	13	35.35	13	37.14	13	34.20

D1		Tun		D2		Lily	
Label	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**	**Label**	**Δ E**
10	0.00	10	0.00	10	0.00	10	0.00
9	9.25	15	8.61	9	8.34	9	7.96
16	10.10	16	8.83	16	9.89	11	8.95
3	10.28	9	8.88	3	9.92	4	10.15
15	11.21	4	9.04	15	10.22	15	10.28
11	11.87	11	9.57	11	10.57	16	10.44
4	12.80	3	10.01	4	11.23	3	10.60
8	17.52	12	14.81	12	16.58	12	14.25
2	17.62	5	14.92	8	17.30	5	16.10
12	18.58	2	18.13	2	17.43	2	17.72
5	21.08	8	18.79	5	19.24	8	18.19
1	24.82	13	21.15	13	24.66	13	20.47
7	26.85	1	26.56	1	24.94	1	26.72
13	26.87	7	29.29	7	27.25	7	29.59
14	36.49	14	40.87	14	37.20	14	41.93
6	41.45	6	46.25	6	42.37	6	47.51

Table AP-4 – continued. Top - Y8. Bottom - Y10.

D2

Measuring point		R10			Y8		
		CIE x	**CIE y**	**CIE Y**	**CIE x**	**CIE y**	**CIE Y**
1	RW	0.447	0.361	8.52	0.446	0.458	10.63
	CS	0.445	0.358	7.73	0.443	0.460	9.58
2	RW	0.388	0.375	11.16	0.389	0.379	11.38
	CS	0.391	0.372	10.43	0.392	0.375	10.40
3	RW	0.459	0.359	10.79	0.448	0.462	14.05
	CS	0.462	0.356	9.88	0.451	0.465	13.02
4	RW	0.418	0.402	4.72	0.418	0.405	4.80
	CS	0.415	0.403	4.26	0.414	0.407	4.31

Lily

Measuring point		R10			Y8		
		CIE x	**CIE y**	**CIE Y**	**CIE x**	**CIE y**	**CIE Y**
1	RW	0.497	0.325	7.60	0.499	0.408	8.43
	CS	0.495	0.322	6.92	0.501	0.412	7.60
2	RW	0.434	0.334	9.24	0.435	0.337	9.25
	CS	0.433	0.335	8.45	0.431	0.335	8.51
3	RW	0.501	0.324	9.70	0.501	0.411	11.09
	CS	0.501	0.327	8.79	0.504	0.408	9.98
4	RW	0.467	0.357	3.83	0.467	0.359	3.83
	CS	0.467	0.359	3.51	0.466	0.361	3.45

Table AP-5. Chromaticity and luminance measurements were taken at four different positions in the RW and CS scene for target colour R10 and Y8 under the illuminants D2 (top) and Lily (bottom).

AP-6 - 1

AP-6 – 2

AP-6 - 3

Obs. 1	RW				CS		
Label	D1	D2	Tun	Lily	D1	D2	Lily
B1	2	2	9	1	2	9	1
B2	3	3	3	3	3	3	8
B3	2	2	2	2	2	3	9
B4	2	2	3	9	2	3	1
B5	3	2	3	2	2	3	9
B6	3	3	3	3	3	3	9
B7	3	3	3	3	3	3	8
B8	3	3	3	3	3	3	8
B9	2	2	3	2	2	3	9
B10	1	1	9	1	9	9	1
B11	3	3	3	3	3	3	3
B12	2	2	2	2	2	2	1
B13	1	1	9	1	1	9	1
B14	2	2	2	2	2	2	9
B15	2	2	1	1	2	2	1
B16	2	2	2	2	2	2	1
R1	5	5	5	5	5	5	5
R2	5	5	5	5	5	5	5
R3	7	7	7	7	7	5	7
R4	1	1	7	1	1	1	7
R5	1	1	1	1	2	9	1
R6	5	5	5	5	5	5	5
R7	5	5	5	5	5	5	5
R8	7	5	5	7	5	5	5
R9	7	7	7	7	7	7	7
R10	1	1	7	7	1	7	7
R11	1	1	1	1	1	1	1
R12	2	2	1	1	2	1	1
R13	2	2	2	2	2	2	1
R14	5	5	5	5	5	5	5
R15	5	5	5	6	5	5	5
R16	7	7	7	6	7	7	7
Y1	3	3	4	9	9	4	8
Y2	3	3	4	3	9	4	5
Y3	4	4	4	5	5	5	5
Y4	5	5	5	5	5	5	5
Y5	5	5	5	5	2	5	5
Y6	3	3	3	3	2	3	9
Y7	3	3	3	3	3	3	8
Y8	3	3	3	3	3	3	8
Y9	5	3	4	4	4	4	4
Y10	4	4	4	5	4	4	5
Y11	5	5	5	5	5	5	5
Y12	5	5	5	5	5	5	5
Y13	5	5	5	5	5	5	5
Y14	3	3	3	3	3	3	3
Y15	3	4	4	4	4	4	4
Y16	4	4	4	4	4	4	5

Table AP-7. Summary of predominant colour terms for all experimental colours under all illuminants and presentation media. Selection is coded after the following scheme: 1 – purple, 2 – blue, 3 – green, 4 – yellow, 5 – orange, 6 – red, 7 – pink, 8 – brown and 9 – grey.

Obs. 2	RW				CS		
Label	D1	D2	Tun	Lily	D1	D2	Lily
B1	2	2	2	1	2	9	1
B2	3	3	3	3	3	3	3
B3	2	2	3	2	2	3	9
B4	2	3	3	9	2	3	8
B5	3	3	3	2	3	3	3
B6	3	3	3	3	3	3	3
B7	3	3	3	3	3	3	3
B8	3	3	3	3	3	3	3
B9	2	3	3	2	2	3	9
B10	2	2	9	1	1	9	1
B11	3	3	3	3	3	3	3
B12	2	2	2	2	2	2	1
B13	1	2	9	3	1	9	1
B14	2	2	3	2	2	3	9
B15	2	2	2	2	2	2	1
B16	2	2	2	2	2	2	9
R1	5	5	5	5	5	5	5
R2	7	7	5	7	5	5	7
R3	7	7	7	7	7	5	7
R4	1	1	7	7	1	7	7
R5	1	2	1	1	1	9	1
R6	5	5	5	5	5	5	5
R7	5	5	5	5	5	5	5
R8	7	7	7	7	7	5	7
R9	7	7	7	7	7	7	7
R10	7	7	7	7	7	7	7
R11	1	1	7	1	1	1	1
R12	2	2	2	1	1	1	1
R13	2	2	2	1	2	2	1
R14	2	5	5	5	5	5	5
R15	7	7	7	5	7	5	5
R16	7	7	7	7	7	7	7
Y1	3	3	3	3	3	3	5
Y2	3	3	4	5	3	4	5
Y3	5	5	5	5	5	5	5
Y4	5	5	5	5	5	5	5
Y5	5	5	5	5	5	5	5
Y6	3	3	3	3	3	3	3
Y7	3	3	3	3	3	3	3
Y8	3	3	3	4	3	3	4
Y9	3	3	4	4	3	4	4
Y10	4	4	5	5	5	4	5
Y11	5	5	5	5	5	5	5
Y12	5	5	5	5	5	5	5
Y13	5	5	5	5	5	5	5
Y14	3	3	3	3	3	3	3
Y15	4	4	4	4	4	4	4
Y16	4	4	5	5	4	4	4

Table AP-7 – continued.

Obs. 3	RW				CS		
Label	D1	D2	Tun	Lily	D1	D2	Lily
B1	2	2	9	1	2	2	1
B2	3	3	3	3	3	3	3
B3	2	2	3	2	3	3	9
B4	2	3	3	9	3	3	9
B5	3	3	3	3	3	3	3
B6	3	3	3	3	3	3	3
B7	3	3	3	3	3	3	3
B8	3	3	3	3	3	3	3
B9	3	3	3	2	3	3	9
B10	1	9	9	1	2	9	1
B11	3	3	3	3	3	3	3
B12	2	2	3	2	2	2	2
B13	1	1	9	1	1	9	1
B14	2	2	9	2	3	3	9
B15	2	2	2	2	2	2	1
B16	2	2	2	2	2	2	2
R1	5	5	5	5	5	5	5
R2	5	5	5	5	5	5	5
R3	7	5	7	7	7	5	7
R4	1	1	7	1	1	1	7
R5	1	2	1	1	2	2	1
R6	5	5	5	5	5	5	5
R7	5	5	5	5	5	5	5
R8	5	5	5	5	5	7	5
R9	7	7	7	7	7	7	7
R10	1	7	7	7	7	7	7
R11	1	1	1	1	1	1	1
R12	1	1	2	2	2	2	1
R13	2	2	2	2	2	2	2
R14	5	5	5	5	5	5	5
R15	5	5	5	5	5	5	5
R16	7	7	7	7	7	7	7
Y1	3	3	3	3	3	3	4
Y2	3	3	4	3	3	3	4
Y3	4	4	4	5	4	4	5
Y4	5	5	5	5	4	5	5
Y5	5	5	5	5	5	5	5
Y6	3	3	3	4	3	3	3
Y7	3	3	3	3	3	3	3
Y8	3	3	3	3	3	3	3
Y9	3	3	4	3	3	3	4
Y10	4	4	4	4	4	4	5
Y11	5	5	5	5	5	5	5
Y12	5	5	5	5	5	5	5
Y13	5	5	5	5	5	5	5
Y14	3	3	3	3	3	3	3
Y15	3	4	4	4	4	4	4
Y16	4	4	4	4	4	4	4

Table AP-7 – continued.

Obs. 4	RW				CS		
Label	**D1**	**D2**	**Tun**	**Lily**	**D1**	**D2**	**Lily**
B1	2	2	2	1	2	9	1
B2	3	3	3	3	3	3	8
B3	2	2	3	3	2	3	9
B4	2	2	9	9	2	3	7
B5	3	2	3	2	3	3	9
B6	3	3	3	3	3	3	8
B7	3	3	3	3	3	3	8
B8	3	3	3	3	3	3	8
B9	2	2	3	2	2	3	9
B10	2	9	9	1	9	9	7
B11	3	3	3	3	3	3	3
B12	2	3	2	2	2	2	3
B13	1	1	9	1	1	9	7
B14	2	2	3	2	2	3	7
B15	2	2	3	1	2	2	1
B16	2	2	3	2	2	2	1
R1	5	5	5	5	5	5	5
R2	7	7	5	7	5	5	7
R3	7	7	5	7	7	7	7
R4	1	1	7	7	1	7	7
R5	2	2	9	1	1	9	1
R6	5	5	5	5	5	5	5
R7	5	5	5	5	5	5	6
R8	7	5	5	7	7	5	7
R9	7	7	7	7	7	7	7
R10	7	7	7	7	7	7	7
R11	1	1	1	1	1	2	7
R12	2	2	2	2	2	3	1
R13	2	4	2	2	2	3	2
R14	5	5	6	6	5	5	6
R15	6	6	6	6	6	6	6
R16	7	7	7	7	7	7	7
Y1	3	3	4	9	3	4	8
Y2	3	4	4	8	8	4	5
Y3	8	4	4	8	5	4	5
Y4	8	5	5	7	5	5	7
Y5	5	7	5	7	5	5	7
Y6	3	3	3	3	3	3	3
Y7	3	3	3	9	3	3	8
Y8	3	3	4	8	3	3	8
Y9	3	4	4	4	4	4	5
Y10	4	3	4	4	4	4	5
Y11	5	5	5	5	5	5	5
Y12	5	5	5	5	5	5	5
Y13	5	5	5	7	5	5	6
Y14	3	3	3	3	3	3	3
Y15	4	4	4	4	4	4	5
Y16	4	4	4	4	4	4	5

Table AP-7 – continued.

Obs. 5	RW				CS		
Label	**D1**	**D2**	**Tun**	**Lily**	**D1**	**D2**	**Lily**
B1	2	2	2	2	2	9	1
B2	3	3	3	3	3	3	8
B3	2	2	3	2	3	3	9
B4	2	3	3	9	3	3	9
B5	3	3	3	3	3	3	3
B6	3	3	3	3	3	3	3
B7	3	3	3	3	3	3	3
B8	3	3	3	3	3	3	3
B9	3	3	3	2	3	3	9
B10	2	2	9	1	9	9	7
B11	3	3	3	3	3	3	3
B12	2	2	2	2	2	2	2
B13	2	2	2	1	2	9	1
B14	2	2	3	2	3	3	9
B15	2	2	5	2	2	2	1
B16	2	2	3	2	2	2	9
R1	5	5	5	5	5	5	5
R2	6	6	5	7	5	5	5
R3	7	7	5	7	6	5	7
R4	1	1	1	1	9	1	1
R5	2	2	2	1	2	9	1
R6	5	5	5	5	5	5	5
R7	5	5	5	5	5	5	5
R8	6	6	5	5	6	5	5
R9	7	7	7	6	7	6	7
R10	1	1	7	6	1	7	7
R11	1	1	1	1	1	1	1
R12	2	2	2	2	2	2	1
R13	2	2	2	2	2	2	2
R14	5	5	5	5	5	5	5
R15	6	6	5	5	6	5	6
R16	7	7	7	7	7	7	7
Y1	3	3	3	8	3	3	8
Y2	3	3	4	5	3	3	5
Y3	8	8	4	5	5	4	5
Y4	5	5	5	5	5	5	5
Y5	5	5	5	5	5	5	5
Y6	3	3	3	3	3	3	3
Y7	3	3	3	3	3	3	8
Y8	3	3	3	8	3	3	8
Y9	3	3	4	8	3	3	5
Y10	4	4	4	5	4	4	5
Y11	5	5	5	5	5	4	5
Y12	5	5	5	5	5	5	5
Y13	5	5	5	5	5	5	5
Y14	3	3	3	3	3	3	3
Y15	3	3	4	8	4	4	5
Y16	4	4	4	5	4	4	5

Table AP-7 – continued.

Obs. 6	RW				CS		
Label	**D1**	**D2**	**Tun**	**Lily**	**D1**	**D2**	**Lily**
B1	2	2	9	1	2	9	1
B2	3	3	3	3	3	3	3
B3	3	3	3	2	3	3	9
B4	3	3	3	2	3	3	9
B5	3	3	3	3	3	3	3
B6	3	3	3	3	3	3	3
B7	3	3	3	3	3	3	3
B8	3	3	3	3	3	3	3
B9	3	3	3	2	3	3	9
B10	2	2	9	1	2	9	1
B11	3	3	3	4	3	3	3
B12	2	2	2	2	2	2	1
B13	2	2	1	1	2	9	1
B14	1	3	3	2	2	3	1
B15	2	2	2	1	2	2	1
B16	2	2	2	2	2	3	1
R1	5	5	5	5	5	5	5
R2	5	5	5	5	7	5	7
R3	7	7	5	7	7	5	7
R4	1	1	7	7	1	7	7
R5	2	1	9	1	2	9	1
R6	5	5	5	5	5	5	5
R7	5	5	5	5	5	5	5
R8	7	5	5	7	7	5	7
R9	7	7	7	7	7	7	7
R10	7	7	7	7	7	7	7
R11	1	1	1	1	1	1	7
R12	2	2	2	1	2	2	1
R13	2	2	2	1	2	2	1
R14	5	5	5	5	5	5	5
R15	5	5	5	4	5	5	6
R16	7	7	7	7	7	7	7
Y1	3	3	4	4	4	4	5
Y2	3	3	4	4	4	4	5
Y3	4	4	4	5	4	4	5
Y4	5	5	5	5	5	5	5
Y5	4	4	5	5	5	5	5
Y6	3	3	3	3	3	3	3
Y7	3	3	3	4	3	3	5
Y8	3	3	4	4	3	4	5
Y9	4	4	4	4	4	4	5
Y10	4	3	4	5	5	4	5
Y11	5	5	5	5	5	5	5
Y12	5	5	5	5	5	5	5
Y13	5	5	5	5	5	5	3
Y14	3	3	3	3	3	3	3
Y15	4	4	4	4	4	4	5
Y16	4	4	4	5	4	4	5

Table AP-7 – continued.

Publications

Hedrich, M., Bloj, M, & Ruppertsberg, A. I. (2009). Color constancy improves for real 3D objects. *Journal of Vision* 9(4):16, 1-16.

Bloj, M., Hedrich, M., & Booth, D. (2008). Short-term colour memory for two-dimensional swatches under different illuminants: paper samples and computer displays. *Proceedings of the 2nd Material & Sensation Meeting*, Pau (France).

Hedrich, M., Ruppertsberg, A. I., & Bloj, M. (2008). Colour constancy for real 3D and 2D scenes under typical and atypical illuminant changes [Abstract]. *Journal of Vision*, *8*(6):573, 573a.

Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2007). "Colour perception in 2-D and 3-D" *Perception,* 36,193, Suppl.