PERCEPTION OF LIGHTING AND REFLECTANCE IN REAL AND SYNTHETIC STIMULI

KHUSHBU Y PATEL

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

SEPTEMBER 2018

ABSTRACT

The human visual system estimates the proportion of light reflected off of a surface despite variable lighting in a scene, a phenomenon known as *lightness constancy*. Classically, lightness constancy has been explained as a 'discounting' of the lighting intensity (Helmholtz, 1866), and this continues to be a common view today (e.g., Brainard & Maloney, 2011). However, Logvinenko and Maloney (2006) have made a radically different claim that the human visual system does not have any perceptual access to an estimation of lightness. The experiments described in Chapter 2 use a novel experimental paradigm to test this new theory proposed by Logvinenko and Maloney. We provide evidence against Logvinenko and Maloney's theory of lightness perception while adding to existing evidence that the visual system has good lightness constancy. In Chapter 3, we manipulate screen colour and texture cues to test the realism of computer-generated stimuli. We find that by matching the chromaticity of an LCD screen to the surrounding lighting and using a realistic texture, LCD screens can be made to appear similar to physical paper. Finally, Chapter 4 is an extension of the ideas from Chapter 3, in which the knowledge about how to adjust color and texture cues on an LCD monitor is applied to a lightness matching task. Here, the LCD screen is a small part of a larger physical setup. Additionally, levels of lightness constancy are compared across physical and simulated surfaces in the same novel experimental paradigm in Chapters 2 and 4. We find that physical and simulated surfaced elicit different levels of lightness constancy on the same task.

ACKNOWLEDGMENTS

I would like to present my absolute gratitude to my supervisor and mentor Dr. Richard Murray for his continued support, encouragement, guidance and patience throughout my undergraduate and now master's thesis. Your brilliance, meticulousness, equanimity and appreciativeness has made these past two years some of the most productive, creative, enjoyable and fulfilling of my life. I am thrilled to get to continue in the lab!

I also would like to express my gratitude to Dr. Laurie Wilcox for her thoughtful, encouraging, and indispensable advice - both scientific and career related. You have been an inspiration to me and I look forward to future collaborations. I would also like to thank Dr. Jacob Beck and Dr. Francis Wilkinson, for being a part of my committee and showing abundant enthusiasm for my work. I would also like to extend my thanks to Dr. Krista Ehinger for introspective suggestions on the thesis. Our continued two-day marathons, your support, and help, made the final stage of the writing process enjoyable.

Furthermore, I cannot begin to express adequately how thankful I am to my family for giving me the opportunity to continue my education here in Canada, and for being my pillar of strength throughout my accomplishments and struggles. I would also like to extend this thank you to all my friends. I have received a steady flow of encouragement and compliments from you, and it would not have been possible without them.

An NSERC Alexander Graham Bell Scholarship, a Queen Elizabeth II Graduate Scholarship in Science and Technology, and a CVR CREATE stipend funded this research. I would like to thank the funding agencies responsible for making my research possible.

Table of Contents

Chapter 3: Texture and Colour Cues for Emissive Surface Detection 43

List of Tables

List of Figures

Chapter 1

INTRODUCTION

An image on the retina can be described as a pattern of light intensity that varies as a function of physical parameters such as retinal position, wavelength and time. Much of this low-level information is lost in the translation from the retinal image to perception. Nevertheless, the retained information is adequate to discriminate scene properties like distance, shape, surface colours, and materials. A combination of low-level primary sensory cues like luminance and mid-level feature-based cues like junctions and larger boundary segments are used to infer surface properties such as colour, texture, and materials under varying lighting conditions. Using a psychophysical approach, this thesis tests theories of lightness perception and compares how surfaces in the physical world and on computer screens are perceived by the human visual system. The specific goals are laid out at the end of this chapter, after an introduction to the issues in the field and a summary of current models and theories.

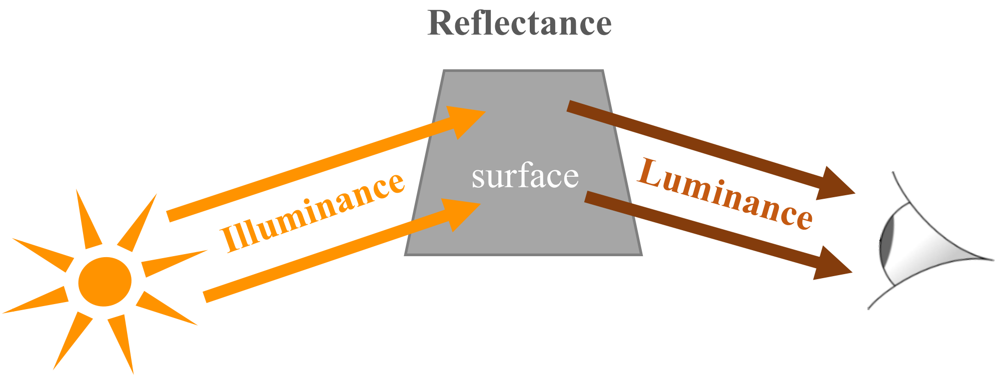
1.1 THE PROBLEM OF LIGHTNESS PERCEPTION

A white piece of paper appears white under both bright and dim light, even though the physical amount of light reflected from the surface is vastly different in these two cases (Figure 1.1). Our ability to perceive both pieces of paper in Figure 1 as white is surprising, because our eyes receive much more light from the left-hand piece of paper than from the right-hand piece of paper. Lightness constancy is this ability to maintain the perceived whiteness or blackness of a surface, or *lightness*, despite substantial changes in lighting intensity (Helmholtz, 1866). Though the human visual system accomplishes lightness constancy with little effort, it is an impressive ability that even the state-of-the-art computer vision systems cannot entirely mimic.

Figure 1.1 The photo shows two white pieces of paper. On the left side sunlight is directly incident on the paper while on the right side the paper is under the shadow of the wall. Nevertheless, we are easily able to see that the two pieces of paper are approximately the same shade of white. This is an example of lightness constancy.

1.2 KEY TERMINOLOGY AND CONCEPTS

To better explain how lightness constancy works, I first define the factors that play a role in the perception of lightness. *Luminous flux* is the power emitted per second in all directions, weighted by the spectral sensitivity of human eye at different wavelengths (i.e., 400 - 700 nm). Luminous flux is measured in lumens (lm). Luminous flux per steradian in a particular direction is called *luminous intensity*. The SI unit for luminous intensity is the candela (cd), which is define as one lumen per steradian. As demonstrated in Figure 1.2, *illuminance* is the the total luminous flux on a surface per unit area. The unit SI of illuminance is the lux (lx), defined as one lumen/m^2. Finally, *luminance* is flux per steradian per unit area, measured in cd/m^2. *Reflectance* is the proportion of light reflected from a surface and is measured as a ratio or percentage. Reflectance is an inherent property of the surface – it is always the same under all lighting conditions.



<u>Figure 1.2</u> Visual depiction of illumination, reflectance and luminance.

The luminance of a matte surface is the product of illuminance (I) and reflectance (R):

$$L = \frac{R \times I}{\pi} \tag{1.1}$$

The factor of π arises from the units used to define luminance (cd/sr m^2) and illuminance (cd/ m^2), but the key point here is that luminance is proportional to both illuminance and reflectance. Equation 1.1 shows that many combinations of gray colors (R) and lighting (I) can produce the same value of luminance (L). For example, a low-reflectance (dark gray) surface under a bright light and a high-reflectance (light gray) surface under a dim light can generate the same luminance. Explaining how our visual system achieves lightness constancy, despite the fact that numerous possible pairs of illumination and surface reflectance yield the same value of luminance, is a key challenge for vision science.

Brightness is a psychological phenomenon; it is the perceived luminance of a patch and is subjective. Judgments such as how much light a surface seems to emit are brightness judgments; the brightness of a surface ranges from ‘dim’ to ‘bright’. The perceived reflectance, or as mentioned above, the perceived whiteness or the blackness of a surface is its *lightness*; this value ranges from ‘dark’ to ‘light’. But if and how people compute these values separately is not well understood. Despite many years of lightness research (Helmholtz, 1866), we still do not completely understand how lightness works and no single theoretical framework is agreed upon. The following two sections summarize some low-level and mid-level models and theories that attempt to explain the perception of lightness and brightness.

1.3 LOW-LEVEL THEORIES AND MODELS

Low-level models are typically based on evidence from sensory physiology and perform computations over pixels in an image without taking into account how they compose objects and surfaces. Low-level theories often start with a concrete model that works for simple images, and then try to extend the reach of the model. The models discussed in the section follow this approach.

One low-level operation that is used to explain the perception of lightness is the *luminance ratio* (Wallach, 1948; Heinemann, 1955). In this mechanism, the brightness of a patch is determined by estimating its contrast; contrast quantifies the relative difference between the luminance of a test patch and its background (Pelli & Bex, 2013). This approach works in some cases because luminance ratios between surfaces remain invariant with changes in illumination (Jacobsen & Gilchrist, 1988; Wallach, 1976). The luminance ratio theory gives an account of relative lightness percepts (e.g., one surface is lighter than that other) but is less clear on absolute lightness percepts (e.g., this surface is light-gray, while that one is white). The luminance ratio mechanism is not sufficient to explain perceived lightness in complex scenes where three-dimensional and volumetric shapes are involved. This is because the luminance ratio along the curved surface of an object changes with respect to the background, even though the luminance of the surround stays constant (Tse, 2005). The luminance ratio mechanism has no account of local illumination as it relies solely on contrast information, so it breaks down when differently-lit regions are compared within a scene (Gilchrist et. al., 1999).

A model that builds on luminance ratios is the *retinex model* (Land, 1977; Land & McCann, 1971). In the retinex model, first the luminance ratios in an image are calculated by taking the ratios between areas where there are sudden changes in luminance values. Then, sequential products are calculated as shown in the center column in Figure 1.3. The goal of calculating sequential products is to find the highest luminance in the image and then normalize all the luminance values with respect to it. Therefore, the reflectance of a patch is equal to the ratio of its luminance divided by the luminance of the whitest patch in the scene, as shown in the left column in Figure 1.3.

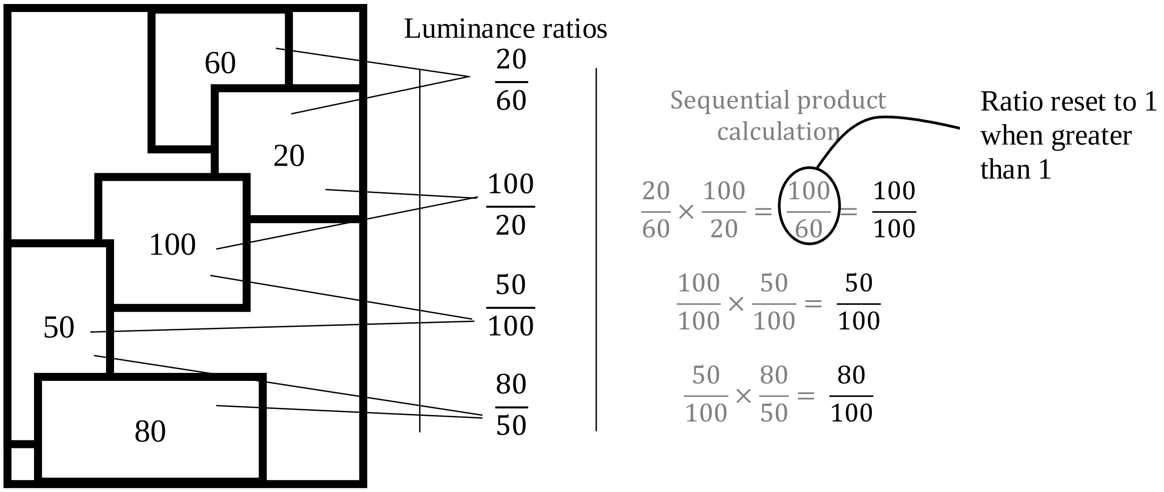


Figure 1.3 An example showing how the sequential products are calculated according to the Retinex theory (based on Land & McCann, 1971). The numbers within the rectangles on the left are the luminance values of each area. The ratios are taken across luminance edges; these ratios are shown in the center column. The ratios are finally multiplied to form sequential products that reset if the ratio in the center column is larger than 1 as shown in the right column.

The retinex model builds on the luminance ratio mechanism, as it also includes the normalization of the ratios to the whitest colour in the scene. The retinex model does not involve any segmentation or reasoning about surfaces or assumptions about lighting. Although, like the luminance ratio model, the retinex model also depends on contrast ratios, but it normalizes these ratios with respect to the whitest patch in the scene, which indirectly takes lighting into account (Land, 1977). This normalization also allows the retinex model to account for absolute lightness percepts, which is not possible when only calculating the luminance ratios (Land, 1977).

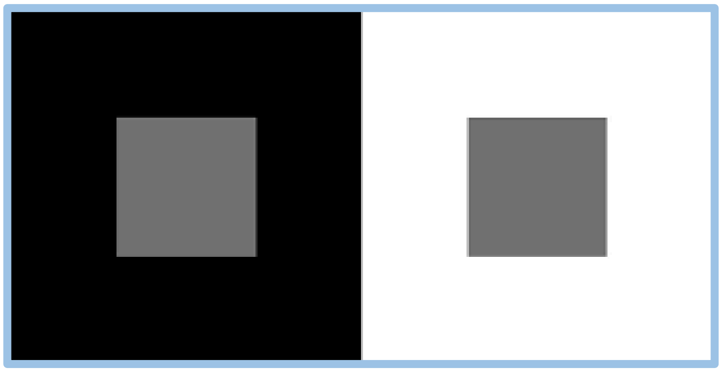


Figure 1.4 Simultaneous contrast illusion. Although the luminances of the central gray squares are the same, the gray square in the black surround appears brighter compared to the gray square in the white surround.

The next class of low-level mechanism that I will discuss is *lateral inhibition*. As shown in Figure 1.4, when a gray patch is placed in two different surroundings, black and white, it appears brighter in the black surround compared to the white surround, even though the luminance of the gray patch is actually the same in both. This demonstration is called simultaneous contrast, and lateral inhibition is a classic explanation of why simultaneous contrast occurs (Cornsweet, 1970). Lateral inhibition refers to the ability of a neuron, when excited, to reduce the activity of surrounding neurons (Cornsweet, 1970). Thus, when a luminance contrast is present, as between the gray patch and its background, that contrast is amplified through lateral inhibition; the difference in luminance is enhanced. As a result, the neural response to the gray patch on the white background is inhibited by its surround, but the response to the gray patch on the black background is not.

Lateral inhibition has inspired spatial filtering models of lightness perception. An example of this approach is the high pass filter model by Shapiro and Lu (2011). Retinal images can be expressed as a combination of various sinusoidal functions with different spatial frequencies. Lighting effects are often low-pass in nature, as illumination covers large areas of an image and the shadow boundaries tend to be blurry. A high pass filter allows the visual system to eliminate the low-frequency information from an image, which translates to discounting lighting effects to recover reflectance information. Therefore, the brightness of a patch can be computed as the low-pass content subtracted from the image. This high pass filter results in a positive peak surrounded by a region that is negative, like a center-surround receptive field. Additionally, a later version of the retinex model incorporated a version of spatial filtering by only including the strong edges and ignoring low frequency information in order to remove the effects of illuminations (Land, 1983).

The oriented difference-of-Gaussian (ODOG) model (Blakeslee & McCourt, 1999, 2003, 2012), is another low-level model that is a variation on the filter-based approach and a slightly more elaborate lateral inhibition model. This model uses a circular on-center and elliptical off-surround and several differently-oriented linear filters which are tuned to specific frequencies as shown in Figure 1.5.

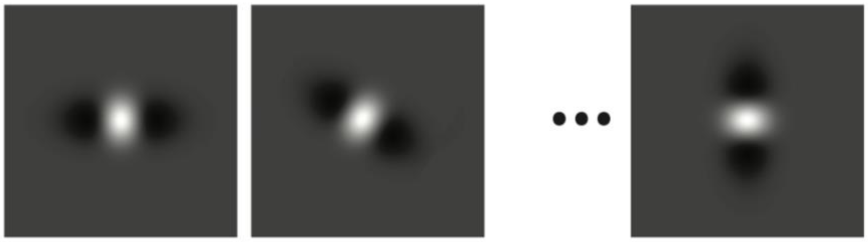


Figure 1.5 Examples of linear filters (taken from Kim, 2016).

The sum of outputs from all these filters gives a complete basic representation of the original image. Once the representation set of the input image is calculated, the model performs the following two steps to estimate brightness. First, the model attenuates low spatial frequencies. As discussed earlier in the high pass filter model, higher frequencies partly store the reflectance information, while lower frequencies changes are more likely to be caused by illumination changes (Horn, 1974; Land & McCann, 1971). Therefore, to estimate brightness it is important that the model eliminates the lowest spatial frequencies. Next, the model performs a contrast normalization because the output from all orientations is not of the same strength.

Figure 1.6 is used as example to demonstrate how the ODOG model uses this process to estimate brightness. A well-known brightness illusion called White's illusion (White, 1979) shows two rectangular patches placed between a vertical grating pattern in Figure 1.6. The ODOG model is supposed to predict how a human would perceive lightness of the image. When the image is passed through filters at various orientations, the vertically-oriented filters respond most strongly, and the horizontally-oriented filters respond more weakly. However, the horizontally-oriented filters convey more information about horizontal edges in the scene, such as the horizontal edges of the central gray rectangles. The model performs a contrast normalization weighted across all frequencies, at all orientations. In the case of Figure 1.6, the model predicts that the left patch is lighter, which is consistent with human perception. Following these two steps, the model is fairly good at accounting for many lightness illusions (Blakeslee & McCourt, 1999, 2003, 2008). However, the ODOG model cannot account for some brightness illusions (Blakeslee & McCourt, 2012), such as the argyle illusion (Figure 1.7; Kim, 2016). The authors claim that some higher-level processing is probably necessary to explain these illusions. All in all, the ODOG model is a low-level model that makes use of spatial filters that are normalized across orientations, which makes it a slightly more elaborate version of the lateral inhibition and spatial filtering models. It is surprising that it works as well as it does, given no assumptions are made about the lighting segments.

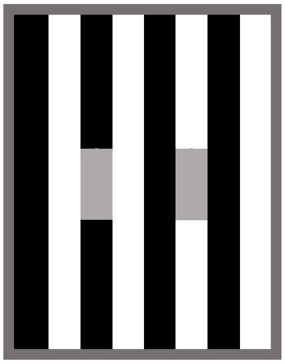


Figure 1.6 Two rectangular patches in between a vertical grating pattern (based on White, 1979). Although the luminance of the rectangular patches is the same, the left patch appears lighter than the right.

1.4 MID-LEVEL THEORIES AND MODELS

As we discussed in the previous section, low level theories do not explicitly take into account illumination changes in the image, which may be part of the reason why they fail to explain lightness and brightness perception in complex scenes (Gilchrist et. al., 1999). Segregating an image into sections of different illumination is called assigning ~~a~~ *lighting framework* to an image. As discussed in this section, mid-level theories try to combat the shortcomings of low-level theories by using simple properties of the image, like blurry boundaries or depth edges, to parse the image into separate lighting frameworks, and then compute lightness by applying a simple computation, or by using a different standard of white within each framework (Land, 1977; Gilchrist et. al., 1999). The *X-junction* is a mid-level feature of interest to describe lightness-related illusions (Metelli, 1970; Beck, Prazdny, & Ivry, 1984). It is defined as a region where four different luminance regions in an image meet at a point, and it often signals the presence of a lighting boundary. To correctly perceive the lightness of surfaces, the visual system might discount the influence of the illumination frameworks (different surrounding lighting conditions) by enhancing or attenuating the internal brightness response; the vertical dark and light strips in Figure 1.7 are the illumination frameworks. Adelson (1993) proposed that X-junctions were important cues to the appearance of illumination strips in an image which modulates the brightness of patches. For isoluminant patches, this gives rise to the percept of different shades of gray, based on the strip in which the patch appears. For example, as shown in Figure 1.7, even though the luminance of the gray patches A and B is the same, B appears brighter in the black surround compared to A in the white surround.

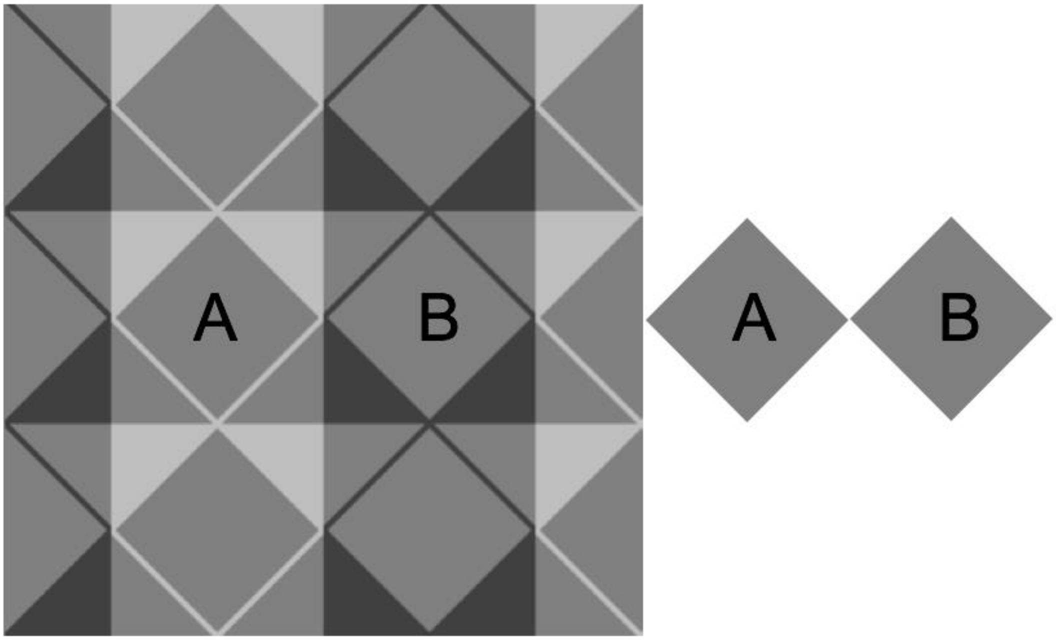


Figure 1.7 Patches A and B have the same surface colour (reflectance). Lightness constancy fails here because of the percept of different surrounding lighting conditions or illumination in the vertical strips. X junctions are important cues for perception of the vertical dark and light strips (based on Adelson, 1993).

Another mid-level process we will discuss is *anchoring.* According to the anchoring model, we estimate the perceived reflectances within a scene by first finding the whitest shade in the scene, the 'anchor', and relating its luminance to the others (Krawczyk, Myszkowski, & Seidel, 2005; Gilchrist et al., 1999). I will explain the idea further by dividing the simultaneous contrast illusion in Figure 1.4 into local and global perceptual frameworks. The global framework represents the entire image as a whole, while the local framework refers to smaller subsections of the images, here the two dark and white surrounds. At each local and global level, lightness is calculated as a weighted average of the two lightness values, derived from each local and global level. The idea is to find the whitest luminance value in both frameworks and find luminance ratios of the entire image with respect to the white. Thus, the anchoring model is an extension of the basic low-level luminance ratios, with an additional mid-level notion of separate surfaces. In the simultaneous contrast example (Figure 1.4), the global ratio of the luminance of the two gray patches would be different from that derived locally. However, an average between the global and local framework is taken, and the result is more accurate lightness and brightness predictions. Therefore, for simple low-level illusions like the simultaneous contrast illusion, the predictions are in line with what we observe, that is, the patch with the dark surround appears lighter than the white surround (Economou, Zdravkovic, & Gilchrist, 2007). While the theory accounts for many findings, in some cases it predicts no brightness differences where humans do perceive a difference (Blakeslee, Reetz, & McCourt, 2009; Bressan, 2006; Bressan & Actis-Grosso, 2001; Rudd & Zemach, 2005).

1.5 THE CURRENT RESEARCH

This thesis explores lightness constancy with physical and simulated surfaces to address three main goals: (1) test predictions of a new theory of lightness perception (chapter 2), (2) compare levels of lightness constancy for the same task on physical surfaces and LCD screen (chapters 2 and 4), (3) identify cues that allow people to distinguish between real and virtual surfaces (chapter 3).

The experiments described in chapter 2 use a novel experimental paradigm to test a the theory of lightness matching proposed by Logvinenko and Maloney (2006). Findings from these experiments will help set the direction of future research by providing evidence about the conditions under which humans can correctly estimate lighting and reflectance. Furthermore, to investigate the issue of whether real and computer-generated stimuli elicit similar visual responses, I compare the results of an experiment carried out with an apparatus made from physical paper patches and lights, to an experiment using computer-rendered stimuli. This will allow us to determine the extent to which lightness constancy for physical and computer-generated stimuli are analogous.

The experiments in chapter 3 further compare the perception of physical surfaces and surfaces on computer screens. I test if colour and texture cues are effective in making a computer-generated stimulus appear realistic, and difficult to discriminate from a physical paper patch. Most experiments in basic vision research use computer renderings rather than physical stimuli, as they provide more precise stimulus control. Findings from this chapter will help improve the realism of computer-generated stimuli.

Chapter 4 is an extension of the ideas from Chapter 3 where I apply the knowledge about adjusting color and texture cues on an LCD monitor to make it appear realistic. I compare physical paper patches and an LCD screen in a lightness matching task in conditions with and without color and texture cues. Findings from this chapter will determine which manipulation to an LCD screen can be used to produce lightness matches equivalent to the ones obtained with physical pieces of paper.

Chapter 2

LIGHTNESS MATCHING AND PERCEPTUAL SIMILARITY

A classic view is that lightness constancy is the result of a ‘discounting’ of lighting intensity, and this continues to be a prominent view today. Logvinenko and Maloney (2006) have proposed an alternative approach to understanding lightness constancy, in which observers do not make explicit estimates of reflectance, and lightness constancy is instead based on a perceptual similarity metric that depends on both the reflectance and the illuminance of surfaces viewed under different lighting conditions. Here we compare these two explanations using a novel, free-adjustment reflectance matching task. We test whether observers can match reflectance in a task where they are free to adjust both the illuminance and the reflectance of the match stimulus over a wide range[1].

---

[1] The body of this chapter has been published in *Journal of Vision* (Patel, Munasinghe, & Murray, 2018).

2.1 INTRODUCTION

People can reliably perceive the colours of surfaces in complex scenes, even though the spectrum and intensity of the light that a surface patch reflects to the eye depend not only on the reflectance properties of the patch itself, but also on illumination conditions and scene geometry. A special case of this ability is lightness perception, the ability to perceive the reflectance of black, white, and grey surfaces. Reflectance is the proportion of incident light reflected by a surface, as measured in photometric units, and lightness is perceived reflectance. Even the most basic principles of lightness perception are still contested. Some theories hold that lightness perception is based on low-level image properties such as bandpass energy (e.g., Blakeslee & McCourt, 1999; Shapiro & Lu, 2011), while others claim that mid-level features such as cues to surface boundaries and lighting conditions play a crucial role (e.g., Adelson, 2000; Bloj et al., 2004; Gilchrist, 2006; Murray, 2013). Our understanding of lightness perception is still so tentative that there is even room for basic questions about how to describe percepts of simple grey surfaces. Logvinenko and Maloney (2006) used difference scaling methods to investigate the perceptual dimensions of matte grey surfaces, and developed a quantitative model of how reflectance and illumination contribute to the perceptual similarity or dissimilarity of surface patches. Figure 2.1 illustrates their model. The *x* axis represents the reflectance of an achromatic surface patch, and the *y* axis represents illuminance (i.e., intensity of incoming illumination). Both axes are scaled logarithmically. The white dot represents the reflectance and illuminance of a reference patch. Any other surface patch has some degree of perceptual similarity to the reference patch, and the grey curves show iso-similarity contours according to Logvinenko and Maloney's model (their equation (5)).

Human observers judge all points on each iso-similarity contour to be equally similar to the reference patch.

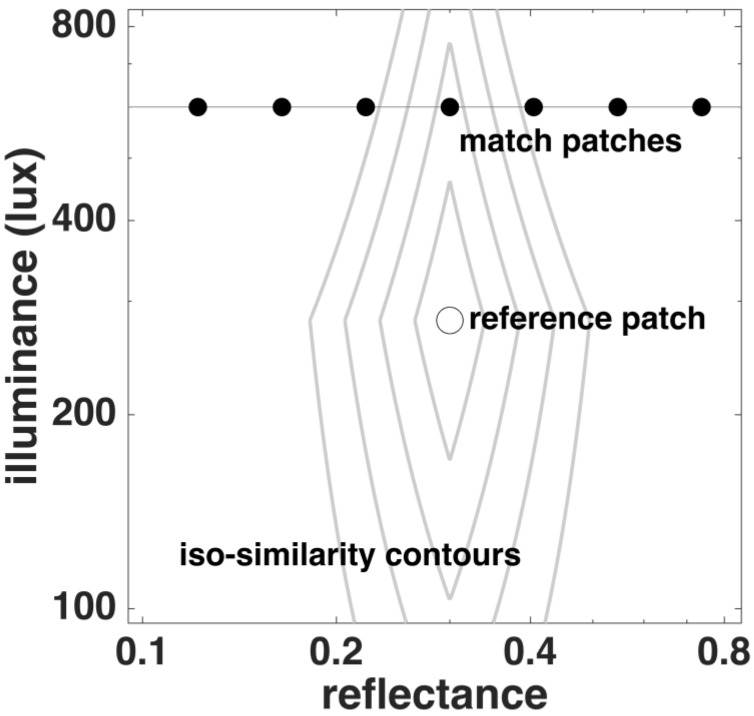


Figure 2.1. A perceptual similarity measure for achromatic surfaces. The white circle represents the reference patch in a lightness matching task, and the grey curves are iso-similarity contours relative to the reference patch, based on Logvinenko and Maloney's (2006) similarity metric. The black circles represent match patches with various reflectances at a fixed illumination level. For this illustration, we chose an arbitrary value for the parameter of the similarity metric that controls the slight pinching of the upper half of the iso-similarity contours and the bowing of the lower half.

The iso-similarity contours illustrate several of Logvinenko and Maloney's findings. First, when measured in logarithmic units, reflectance differences contribute more to perceptual dissimilarity than illumination differences do, so iso-similarity contours are closer to the reference patch in the *x* direction than in the *y* direction. Second, the diamond-like shape of the contours shows that similarity is given approximately by a city-block metric: to find the approximate similarity between the reference patch and another patch we take a weighted sum of the difference between them in log reflectance and in log illuminance. A third, subtler phenomenon is that two reflectances are seen as slightly less similar under high illuminance than under low illuminance. This is indicated by the inward bowing of the upper half of the diamonds and the outward bowing of the lower half.

Logvinenko and Maloney point out that in addition to establishing fundamental facts about the perceptual dimensions of achromatic surfaces, these results also suggest a new theory of lightness matching. In a typical asymmetric lightness matching task, the observer sees a reference patch under one illuminant, as well as a set of match patches under a different illuminant, and the observer's task is to choose the match patch that has the same reflectance as the reference patch. We can represent the reference patch as the white dot in Figure 2.1, and the match patches as the black dots along a horizontal iso-illuminance line. The usual view of lightness matching is that the observer estimates the reflectance of the reference and match patches (and investigating exactly how the observer computes reflectance is often the goal of such experiments), and then chooses the match patch with the reflectance estimate closest to that of the reference patch. Logvinenko and Maloney suggest that, in fact, observers may not estimate reflectance at all in such tasks. They point out that because typical lightness matching experiments show all match patches under a single illuminant, the observer does not have to compute reflectance in order to make a reflectance match. Instead, the observer can simply choose the match patch that is perceptually most similar to the reference patch according to the similarity measure illustrated in Figure 2.1, and this match will have approximately the same reflectance as the reference patch. That is, the observer solves a constrained optimization problem, matching reflectance by maximizing perceptual similarity along the iso-illuminance line, without computing a separate estimate of reflectance *per se*.

Expanding on this theory, Logvinenko and Maloney suggest that there are at least two dimensions to achromatic surface appearance: lightness, which depends mostly on surface reflectance, and surface brightness, which depends mostly on the intensity of incident illumination. Earlier researchers have also argued that lightness and brightness are distinct perceptual dimensions (e.g., Arend & Spehar, 1993); lightness is usually defined as perceived reflectance, and brightness as perceived luminance. One of Logvinenko and Maloney's contributions is to show that two such dimensions emerge spontaneously from a multidimensional scaling analysis that makes no prior assumptions about the perceptual dimensions of achromatic surfaces. Furthermore, it is sometimes ambiguous in the literature whether brightness is meant to be a property of an image or of a reflective surface depicted in an image; in Logvinenko and Maloney's account, brightness is unambiguously a property of the surface (hence "surface brightness"). Finally, brightness is usually defined as perceived luminance, but in Logvinenko and Maloney's account, surface brightness depends mostly on incident illuminance, not on the luminance of a surface, which depends on both illuminance and reflectance.

Logvinenko and Maloney also suggest that observers are unable to judge these two perceptual dimensions independently, e.g., they are unable to abstract the lightness values of two surfaces under different illuminants and compare the lightness values directly. When asked to choose a match patch whose reflectance is the same as that of a reference patch, observers simply choose the match patch that is most similar to the reference patch according to the measure illustrated in Figure 2.1. Logvinenko and Maloney note that this hypothesis explains why observers are sometimes dissatisfied in asymmetric matching tasks, and report that no available match patch perfectly matches the reference patch (e.g., Adelson, 1993, footnote 8). According to the similarity-maximizing theory, the observer chooses the match patch that is most similar to the reference patch (as in Figure 2.1), but because the reference and match patches are seen under different illuminants, even this optimal match has some nonzero level of dissimilarity to the reference patch. The match patch that the observer chooses may have the same lightness as the reference patch, but the observer is unable to judge this perceptual dimension alone. Thus the matching task is designed in such a way that the observer’s attempt to choose a perfectly similar match cannot succeed.

This theory of asymmetric lightness matching takes a fundamentally new approach to lightness perception, in that it suggests that observers do not actually estimate reflectance, or even a proxy of reflectance (e.g., Blakeslee & McCourt, 1999). It implies that observers’ ability to match reflectance depends strongly on the design of the experiment: when all match patches are on a single iso-illuminance line, an observer who does not perceive reflectance can nevertheless match reflectance by maximizing a measure of perceptual similarity that depends on both reflectance and illuminance.

To develop a test of this theory, we reasoned that if reflectance matches are not based on reflectance estimates, and if people's ability to match reflectance is a byproduct of how lightness matching experiments are typically designed, then it should be possible to find a new experimental design that makes it difficult or impossible to match reflectance by minimizing perceptual dissimilarity. Here we test whether observers can match reflectance in a task where they are free to adjust both the illuminance and the reflectance of the match stimulus over a wide range. In this task, the match stimuli do not all lie on an iso-illuminance line, and so there is no constraint that allows observers to make a correct reflectance match by choosing the stimulus that is least dissimilar to the reference stimulus. If observers have no estimate of reflectance, but only a similarity measure that depends on both reflectance and illuminance (as in Figure 2.1), then in a task where they have control over both reflectance and illuminance they should find it difficult or impossible to make a reflectance match. To anticipate, we find that observers can make reflectance matches in such a free-adjustment task, and this result supports the alternative hypothesis that observers use explicit estimates of reflectance to make lightness matches.

2.2 EXPERIMENT 1

Methods

Participants. There were six observers; one was author KP. The others were unaware of the purpose of the experiment and were paid for participation. The observers were 19 to 24 years old, and four were female. In both experiments reported here, all observers gave written informed consent and reported normal vision, and all procedures were approved by the Office of Research Ethics at York University.

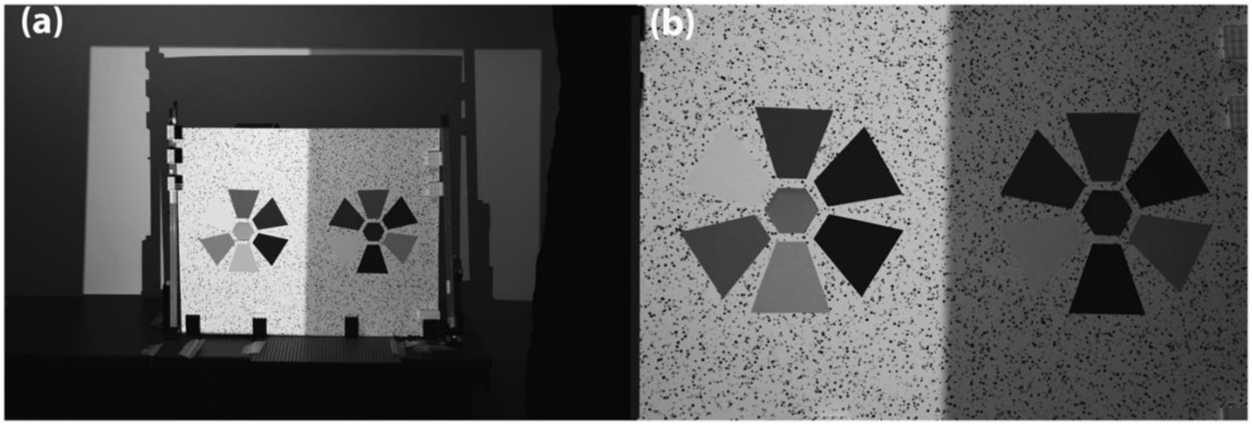


Figure 2.2 Apparatus for Experiment 1. (a) The observer's view. (b) Close-up of the central panel. The observer adjusted the illuminance on the left-hand side and the reflectance of the left-hand hexagon (the match hexagon). The illuminance on the right-hand side was fixed, and the reflectance of the right-hand hexagon (the reference hexagon) varied from trial to trial.

Stimuli. The stimuli were modelled after those in Logvinenko and Maloney's (2006) Experiment 1. Observers viewed a panel of white poster paper that measured 61 cm horizontally and 50 cm vertically, at a viewing distance of 170 cm, subtending 20.3° x 16.7° (Figure 2.2). The panel had a reflectance of 0.78 (i.e., 78%), and was printed with a field of small, randomly placed circles with diameters ranging from 0.05 to 0.30 cm and reflectances ranging from 0.08 to 0.75. The panel had two hexagonal apertures with maximal diameters of 5.0 cm (1.7º), through which papers of various reflectances could be seen (details below). The distance between the centers of the two apertures was 25.5 cm (8.2º). Six trapezoids with reflectances of 0.07, 0.10, 0.21, 0.29, 0.45 and 0.74 were placed around each aperture. The trapezoids were 6.9 cm long (2.3°), and 2.8 cm wide (0.9°) at the shorter end and 7.8 cm wide (2.6°) at the longer end. The gap between the apertures and the trapezoids was 1.0 cm (0.3°) wide.

We used computer-controlled motors to show papers of various reflectances through the two hexagonal apertures. Immediately behind the reference aperture (Figure 2, right-hand hexagon), flat against the back of the main textured panel described above, was a circle of poster paper with a diameter of 31.0 cm. The circle was divided into three equal sectors with reflectances of 0.20, 0.31, and 0.49. A computer-controlled motor was able to rotate the circle so that a uniform section of paper with one of the three reflectances was visible through the reference aperture. Immediately behind the match aperture (Figure 2, left-hand hexagon), flat against the back of the panel, was a conveyor-belt-like loop of poster paper, 152 cm long and 11 cm wide. The reflectance of the conveyor belt ranged continuously from 0.06 to 0.80. Reflectance increased nonlinearly along the belt, in such a way that reflectances were evenly spaced according to the Munsell scale (Schanda, 2007). A computer-controlled motor was able to advance the conveyor belt so that any section of the belt was visible through the match aperture. Just 5 cm of the 152 cm belt was visible through the match aperture at any time, so the visible segment of paper always had approximately uniform reflectance, and the reflectance gradient was not perceptible.

We created the main panel, three-sector circle, and conveyor belt by printing them on a poster printer that we had characterized by measuring the mapping from greyscale RGB values to reflectances. We measured reflectances using a Minolta LS-110 photometer and a Labsphere Spectralon 99% diffuse reflectance standard (model SRS-99-020). Under constant illumination, we measured the luminance l_S of the reflectance standard and the luminance l_T of the target surface whose reflectance was being measured, and assigned the target surface the reflectance $r_T = 0.99\ l_T / l_S$.

The apparatus was illuminated by overhead fluorescent lights, and by an LCD data projector (Epson PowerLite HC2040) placed 104 cm in front of the panel. We characterized the data projector using a photometer and reflectance standard (same models as above) to measure the mapping from greyscale RGB values to illuminances at the reference and match apertures. During the experiment, lighting on the right-hand side of the main panel was fixed, with an illuminance of 314 lux at the reference aperture, meaning that a surface with a reflectance of 1.0 would have a luminance of 100 cd/m^2. Lighting on the left-hand side of the panel was adjustable via a computer program that controlled the data projector, with illuminance at the match patch adjustable between 628 and 2512 lux, meaning that a Lambertian surface with a reflectance of 1.0 would have a luminance between 200 and 800 cd/m^2. There was very little light scatter: increasing the illuminance from minimum to maximum on the match side increased the illuminance on the reference side by just 4%. The lighting boundary between the left- and right-hand sides was created by the data projector, which projected a two-tone image that was brighter on the left than on the right. We defocused the projector slightly so that no pixellation was visible in the projected image, and so that the lighting boundary was penumbra-like and unlikely to be mistaken for a reflectance boundary.

Procedure. Each observer ran in one 30-minute session that included 120 trials. At the start of each trial, the printed circle behind the reference aperture was rotated to show one of the three reference reflectances, randomly chosen with the constraint that the same reflectance was not shown on consecutive trials. Additionally, at the start of each trial, the conveyor belt behind the match aperture was advanced to show a random initial reflectance, and the data projector set the illuminance at the match aperture to a randomly chosen value in the adjustable range given above. After these initial settings were made, a short beep indicated to the observer that the apparatus was ready for their response. The observer used two keys on a computer keyboard to increase or decrease the match reflectance via the conveyor belt, and two other keys to increase or decrease the match illuminance via the data projector. The adjustments were effectively continuous, with no artificially introduced steps in reflectance or illuminance. A brief tap of the reflectance keys moved the conveyor belt 1 to 2 cm of its 152 cm length, which adjusted reflectance by about 1%. Illuminance steps were limited only by the 256 grey levels of the projector. The observer was instructed to adjust the paper and lighting at the match aperture so that the paper appeared to be the same shade of grey paper as the paper visible through the reference aperture, but to be illuminated at a different lighting intensity than the reference aperture. On a few practice trials before the main experiment, we trained the observer to adjust both reflectance and illuminance on every trial. The observer had unlimited time to respond and pressed a fifth key to indicate that they had finished the trial. A short beep acknowledged their response, and then the next trial began. The median duration of the free-adjustment phase of the trial was 16 seconds, across all observers and trials. The relatively long response time was largely due to the slow speed of the apparatus.

As described above, the adjustable match illuminance range was above the fixed reference illuminance. We chose this design so that observers could not adjust the match illuminance to equal the reference illuminance, because then reflectance matching could be accomplished via luminance matching. In the present design, observers always had to match reflectance across two regions of very different illuminance, which provides a stronger test of lightness constancy. Our results support the theory that observers make lightness matches by estimating and comparing reflectance, so we expect that we would find similar results if the match illuminance range was lower than the reference illuminance, but this would need to be confirmed experimentally.

We discarded all trials where observers did not adjust both reflectance and illuminance, since on these trials observers were effectively choosing a stimulus along an iso-illuminance or iso-reflectance constraint line, instead of freely adjusting both reflectance and illuminance as required by the experimental design. On average, observers adjusted both dimensions on 70% of the trials.

Results and discussion

Figure 2.3 shows the reflectance and illuminance of the observers' match settings (small circular data points), along with the three reference patches (large circular data points). The yellow, blue, and green points show the conditions where the reference reflectance was 0.20, 0.31, and 0.49, respectively. The match settings can be compared with two lines on the graphs: the vertical dashed lines show iso-reflectance lines, where points have the same reflectance as the reference stimuli, and the diagonal dotted lines show iso-luminance lines, where points have the same luminance as the reference stimuli. If an observer has perfect lightness constancy, they will choose matches along the iso-reflectance lines. If an observer has no lightness constancy, and simply matches the luminance of the reference and match stimuli, then they will choose matches along the iso-luminance lines. Here observers' matches were close to the vertical iso-reflectance lines, but with some influence of illuminance, indicating imperfect discounting of illumination and hence partial lightness constancy, as is commonly found in studies of lightness constancy and colour constancy (e.g., Kraft & Brainard, 1999).

The Thouless ratio is a common measure of lightness constancy (Thouless, 1931), defined as:

$$\tau = \frac{\log(r_M) - \log(r_0)}{\log(r_R) - \log(r_0)} \tag{1}$$

Here r_R is the reference patch reflectance, r_M is the match reflectance chosen by the observer, and r_0 is the match reflectance that would be chosen by an observer who has no lightness constancy and simply matches image luminance. A Thouless ratio of one indicates perfect lightness constancy, and a value of zero indicates that the observer has no constancy and is matching luminance. In the appendix, we show that on log-log axes, the match points corresponding to a fixed Thouless ratio τ form a straight line that has slope $(\tau - 1)^{-1}$ and passes through the point (r_R, i_R) that represents the reflectance r_R and illuminance i_R of the reference patch. In Figure 2.3 we show a sum-of-squares fit of a straight line to each observer's match settings, constrained to pass through the reference patch; this shows the best fit of the model outlined in the appendix, where the Thouless ratio is assumed to be constant across illuminance levels. Figure 2.3 also shows the Thouless ratio τ corresponding to each fitted line. Estimates of τ ranged from 0.67 to 0.94, and bootstrapped 95% confidence intervals ranged from ±0.02 to ±0.06. These Thouless ratios indicate relatively good lightness constancy. For comparison, Gilchrist (2006) calculated the Thouless ratios in some of Katz's classic experiments on lightness constancy, and found values ranging from 0.35 to 0.70.

An alternative method of quantifying lightness constancy is to find the average of the Thouless ratios calculated from individual match points. This approach gave very similar results to the model fitting method described above, for all observers and conditions. For example, observer EH in Experiment 1 had average Thouless ratios of 0.72, 0.66, and 0.83 in the three reference stimulus conditions, which are similar to the fitted Thouless ratios of 0.75, 0.67 and 0.87 (Figure 2.3).

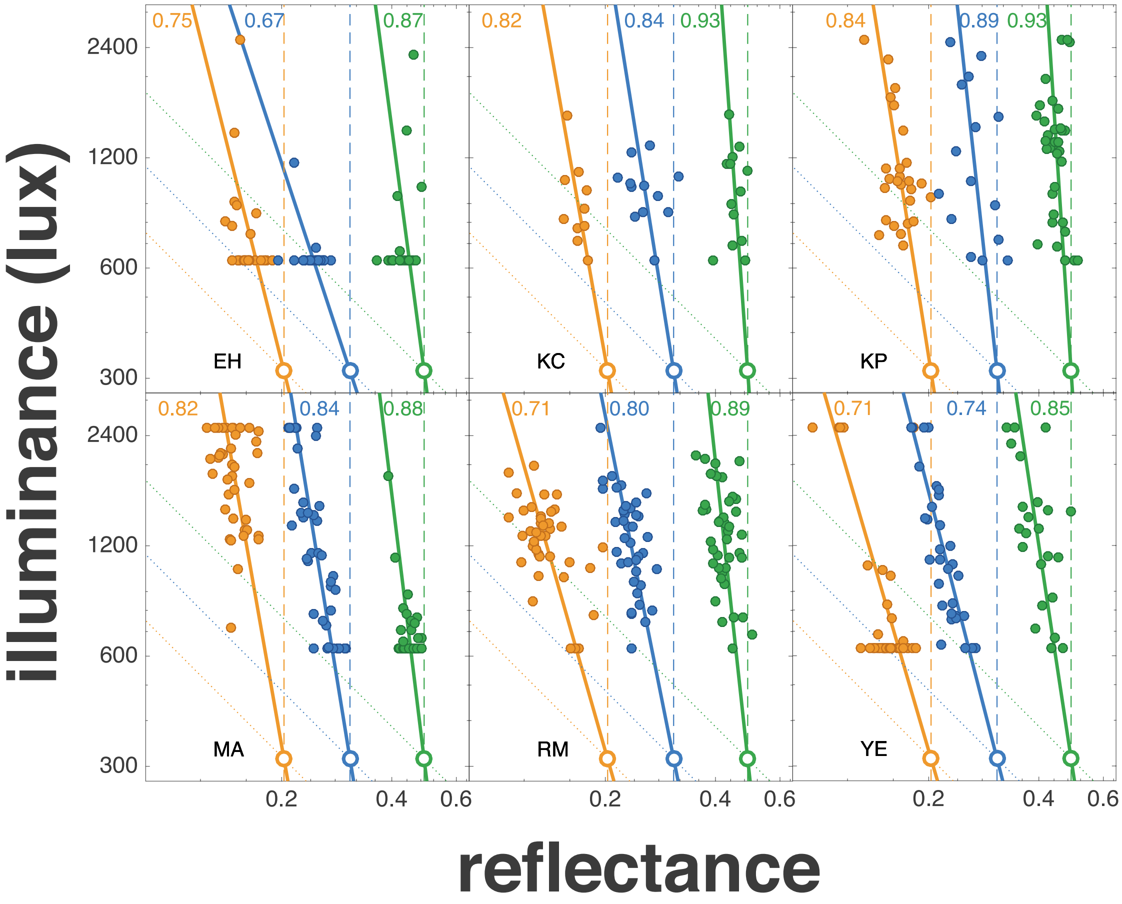


Figure 2.3. Results of Experiment 1. The small circles show reflectance and illuminance matches for all six participants. Yellow, blue, and green represent conditions with reference reflectances 0.20, 0.31 and 0.49, respectively. The white dots with a coloured outline represent the reference patches. The solid lines are linear fits to the data, constrained to pass through the point representing the reference patch. Thouless ratios are

shown beside each line. The vertical dashed lines are iso-reflectance lines and the diagonal dotted lines are iso-luminance lines.

At the end of this experiment, and also at the end of Experiment 2, we informally asked observers if they were satisfied with their matches, and whether there were trials on which a good match was not possible. Observers described the task as an easy one, where it was always possible to make a good match.

Clearly observers were able to match reflectance in this task where they freely adjusted both reflectance and illuminance. This is consistent with Rutherford and Brainard (2002), who also found that observers had reasonably good lightness constancy in tasks where they could adjust both reflectance and illuminance. The set of possible match stimuli was not on a single iso-illuminance line, and so observers cannot have been making reflectance matches by maximizing similarity between the reference patch and the match patch, using a similarity measure like the one shown in Figure 2.1. It appears that instead, observers compute reflectance (e.g., Gilchrist et al., 1999) or some proxy of reflectance (e.g., Blakeslee & McCourt, 1999), and choose the match stimulus for which this estimate has approximately the same value as for the reference stimulus.

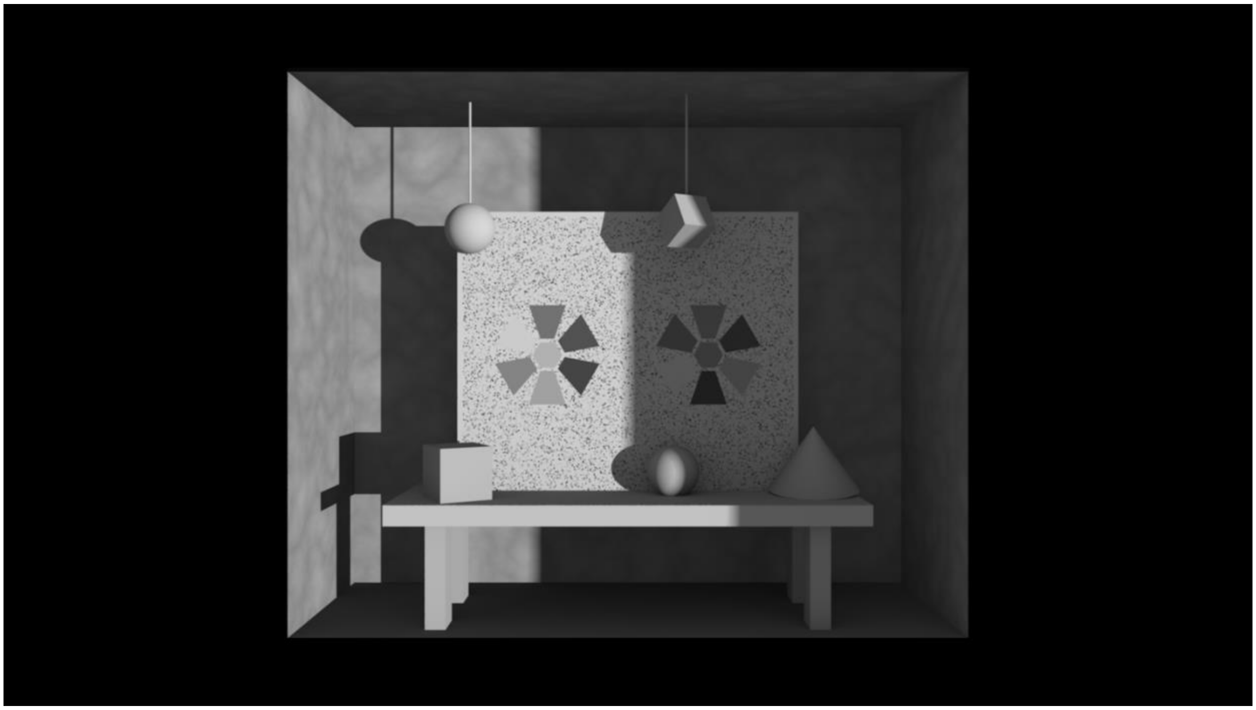
2.3 EXPERIMENT 2

Having built an apparatus to run Experiment 1 with real paper and light sources, we had an opportunity to test whether we would have found similar results using computer-generated stimuli. It is increasingly common for vision researchers to use rendering software to create precisely controlled images of complex scenes (e.g., Morgenstern, Murray, & Harris, 2011), but there have been concerns that even high-quality computer-generated images may not be realistic enough to elicit the same visual processing as real scenes (Morgenstern, Geisler, & Murray, 2014; Radonjić et al., 2016). To investigate this issue, we repeated Experiment 1 using computer-generated renderings of a similar apparatus.

Methods

Participants. There were six observers. All were unaware of the purpose of the experiment and participated for payment. None had participated in Experiment 1. The observers were 20 to 37 years old, and four were female.

Stimuli. The stimulus was rendered in RADIANCE (Ward, 1994), and was modelled after the apparatus used in Experiment 1. The stimulus depicted a box containing a frontoparallel panel and some geometric objects on a small table (Figure 2.4). The panel had the same surface texture as the panel in Experiment 1, and had trapezoids and hexagons with the same relative sizes and in the same arrangement. The panel, trapezoids, and hexagons were the same size in degrees of visual angle as those in Experiment 1, and the simulated reflectances of the panel and trapezoids were also the same. The simulated reflectances of the two hexagons were adjustable, as described under Procedure. We added the geometric objects, which were not present in Experiment 1, in order to give observers a more vivid sense of the spatial arrangement and lighting conditions of the rendered stimulus. Experiment 1 used a paper-and-light stimulus, located on a desk and surrounded the nearby furniture and walls of the testing room, so the spatial arrangement and lighting conditions were very clear to observers. We added the geometric objects to Experiment 2 to compensate somewhat for the reduced viewing conditions with rendered stimuli.



<u>Figure 2.4.</u> Computer-generated stimulus in Experiment 2. The hexagon in the middle of the darker, right-hand side of the panel is the reference patch, and the hexagon in the middle of the brighter, left-hand side is the match patch.

We rendered the scene under simulated lighting consisting of a small point-like light source and an ambient source. The point-like source consisted of 20 infinitely distant point light sources, all with the same intensity, distributed over an arc of 0.4°, behind the virtual camera and to the right. This light source illuminated the left half of the display, and cast a shadow on the right half due to an opaque surface that was not visible in the scene. We distributed the point-like source over a small range of directions (0.4°) in order to give the shadow on the panel a clearly visible penumbra, so that it was unlikely to be mistaken for a reflectance boundary. The intensity of the point-like light source was adjustable, as described under Procedure. The ambient light source illuminated the scene uniformly from all directions with a luminance of 135 cd/m^2. That is, the ambient light source functioned as an extended sky, with a visible luminance of 135 cd/m^2 in all directions.

We showed the stimuli on an LCD monitor in a dark room at a viewing distance of 45 cm. The background luminance surrounding the stimulus on the monitor was 0.08 cd/m^2. The monitor had a resolution of 2560 × 1440 pixels, a pixel size of 0.248 mm, and a nominal refresh rate of 60 Hz. We characterized the monitor's gamma function, using a Minolta LS-110 photometer to measure the relationship between greyscale RGB value and luminance, and the stimulus display software inverted a fitted gamma function to display the required luminances (Brainard, Pelli & Robson, 2002, p. 179).

Procedure. Each observer ran in one 20-minute session that consisted of 120 trials. At the start of each trial, the simulated reflectance of the reference hexagon (Figure 2.4, right-hand side) was randomly set to one of the three reference reflectances used in Experiment 1 (0.20, 0.31, or 0.49), with the constraint that the same reflectance was not shown on consecutive trials. The simulated illuminance at the reference hexagon was 136 lux, meaning that a surface with a reflectance of 1.0 would have a luminance of 43 cd/m^2. Also at the start of each trial, the reflectance of the match hexagon (Figure 2.4, left-hand side) was set to a random value in the range 0.06 to 0.80, and the intensity of the point-like light source was set so that the illuminance at the match hexagon was a random value in the range 275 to 1102 lux. The observer's task was to adjust the reflectance and illuminance at the match hexagon so that the reflectance appeared to be the same as at the reference hexagon, but the illuminance appeared to be different. Moving a mouse left or right changed the reference hexagon's reflectance in the range 0.06 to 0.80, and moving it up or down changed the illuminance in the range 275 to 1102 lux. Because of the monitor's limited luminance range, the simulated illuminances were lower in this experiment than the real illuminances in Experiment 1, but they were proportional: here the simulated illuminance was the same for each reference stimulus, and the lower and upper limits of the adjustable illuminance range for the match stimulus were two and eight times the reference stimulus illuminance, as in Experiment 1. On a few practice trials before the main experiment, we trained the observer to adjust both reflectance and illuminance on every trial. The observer had unlimited time to respond, and clicked the mouse button to indicate that they had finished the trial. A short beep acknowledged the observer's response, and then the next trial began. The median duration of the free-adjustment phase of the trials was 6 seconds, across all observers and trials. We discarded all trials where observers did not adjust both reflectance and illuminance. On average, observers adjusted both dimensions on 85% of the trials.

Results and discussion

Figure 2.5 shows the reflectance and illuminance of observers' match settings, as well as lines of best fit and Thouless ratios. The results are broadly similar to those of Experiment 1. Thouless ratios ranged from 0.69 to 1.26, and bootstrapped 95% confidence intervals ranged from ±0.02 to ±0.11.

However, the mean Thouless ratio was significantly higher in Experiment 2 than in Experiment 1 ($M_1 = 0.82$, $SD_1 = 0.08$; $M_2 = 0.94$, $SD_2 = 0.13$; independent samples t test, $t(34) = 3.12$, $p < 0.01$). This is surprising, because previous studies have typically found weaker lightness constancy with computer-generated stimuli than with real stimuli (e.g., Morgenstern et al., 2014). Observers may sometimes have weaker constancy with computer-generated stimuli because such stimuli are not usually completely realistic, and so observers may tend to match image luminance rather than the simulated reflectance of objects and surfaces depicted in rendered images. In the general discussion we speculate as to why observers in Experiment 2 showed the opposite pattern of results, i.e., seemingly better lightness constancy with a simulated experimental apparatus than with a real apparatus.

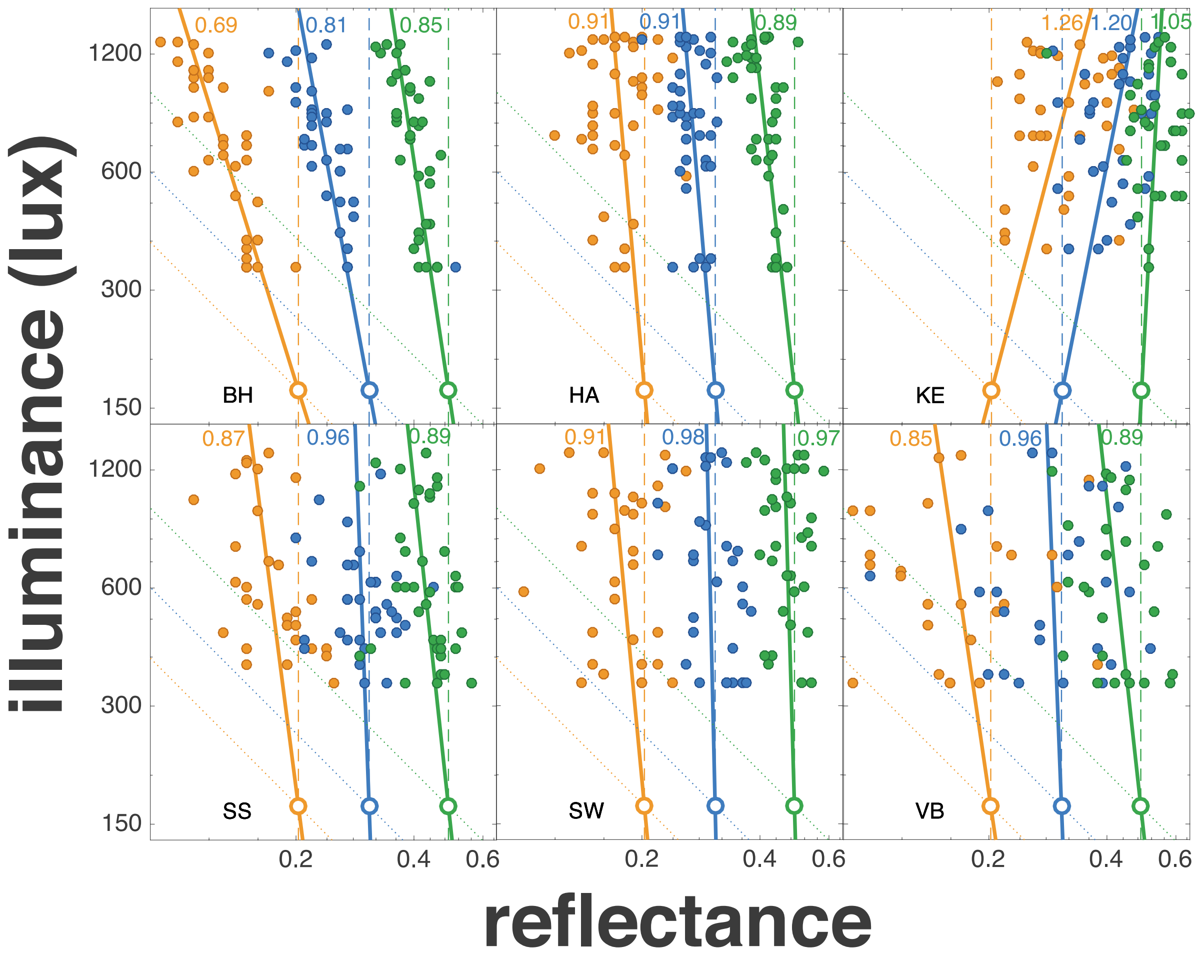


***Figure 2.5*.** Results of Experiment 2. See caption of Figure 2.3 for details.

2.4 GENERAL DISCUSSION

Logvinenko and Maloney (2006) advance two hypotheses about perception of achromatic surfaces. The first is that achromatic surfaces have at least two perceptual dimensions: lightness and surface brightness. This hypothesis is based on a multidimensional scaling analysis of observers' dissimilarity ratings of achromatic surface patches, which resulted in a perceptual space with two dimensions roughly corresponding to reflectance and illuminance. Thus Logvinenko and Maloney define lightness and brightness differently than other authors, namely as dimensions that emerge from a multidimensional scaling analysis. For most other authors, lightness is perceived reflectance, and brightness is perceived luminance. To emphasize this difference, Logvinenko, Petrini, and Maloney (2008) refer to the dimensions that emerge from multidimensional scaling as *surface lightness* and *surface brightness*.

To provide some context, it is worth noting that later studies using multidimensional scaling have found different perceptual spaces for achromatic surfaces. Logvinenko, Petrini, and Maloney (2008) arrived at a one-dimensional space for percepts of surface patches in the snake illusion (Adelson, 2000), possibly because the stimuli in those experiments did not contain lighting boundaries. Madigan and Brainard (2014) also arrived at a one-dimensional space, perhaps because they instructed one group of observers to judge the similarity of reference patch reflectance, and another group to judge the similarity of reference patch luminance (i.e., reflected light intensity), but none judged overall similarity as Logvinenko and Maloney's (2006) observers did. We also note that observers in the second group judged the *luminance* of two reference patches, whereas in Logvinenko and Maloney's account the surface brightness dimension corresponds most closely to *illuminance* (i.e., incident light intensity); observers may be more likely to conflate reflectance and luminance than reflectance and illuminance, which may account for the one-dimensional perceptual space found in that experiment. Finally, Logvinenko (2015) found that in achromatic scenes that contained both cast shadows and attached shadows, perceptual spaces were three-dimensional. In summary, there has been progress in understanding percepts of achromatic surfaces, but the question of what dimensions such percepts have, and under what conditions, is far from settled.

Logvinenko and Maloney's second hypothesis is that the perceptual dimensions of lightness and surface brightness are not separately accessible to observers, and that observers match lightness by minimizing a dissimilarity metric that depends on both reflectance and illuminance. In the experiments reported here we find that observers can match reflectance reasonably well in a task where both reflectance and illuminance are freely adjustable. The second hypothesis is not easily reconciled with this finding. However, the alternative view, that observers form a separate estimate of surface reflectance, and that their performance in lightness matching tasks is based on this estimate, fits well with the behaviour we observed in these experiments.

Logvinenko and Maloney note that in lightness matching tasks, observers sometimes report that no available match stimulus, over the whole range of physically possible reflectances, is a good match for the reference stimulus. Their second hypothesis gave an explanation for this phenomenon, namely that the perceptual dissimilarity between reference and match stimuli cannot be reduced to zero when the two stimuli are seen under different lighting conditions (Figure 2.1). If we do not accept the second hypothesis, how can we understand the difficulty that observers sometimes have when making lightness matches?

We suggest that the explanation lies in the first hypothesis, that achromatic surfaces have at least two perceptual dimensions. If a reference surface has two or more perceptual dimensions (e.g., lightness, gloss, translucency, surface brightness), then even if a match surface can be equated to the reference surface on one of these dimensions by adjusting a one-dimensional physical property such as luminance, it may not be matched on the other dimensions. Thus, even when the targeted perceptual dimension (e.g., lightness) is the same for reference and match surfaces, observers may be aware that the overall appearance of the two surfaces remains quite different.

A vivid example of this is the haze illusion, shown in Figure 2.6. The plus-shaped regions A and B are physically identical, but they have very different appearances. Region A appears to be a clear aperture, whereas B appears to depict a partly transparent material. Unsurprisingly, we find that increasing or decreasing the luminance of region A does little to make the regions appear more similar, because at no luminance does region A have the cues to transparency that determine the appearance of region B (e.g., the pattern of luminances at X-junctions; Metelli, 1970). We can judge whether region A appears brighter or darker than region B, but even when the two regions look about equally bright (as in Figure 2.6), their total appearance remains very different. In a similar way, we suggest that when observers in a lightness matching task adjust the match stimulus so that it has the same lightness as the reference stimulus, the two stimuli may still differ on other perceptual dimensions, such as surface brightness, with the result that there is an imperfect match in overall appearance. As mentioned earlier, we did not find this to be a problem in our experiments, and whether it occurs presumably depends on the perceptual dimensions of the stimuli being used.

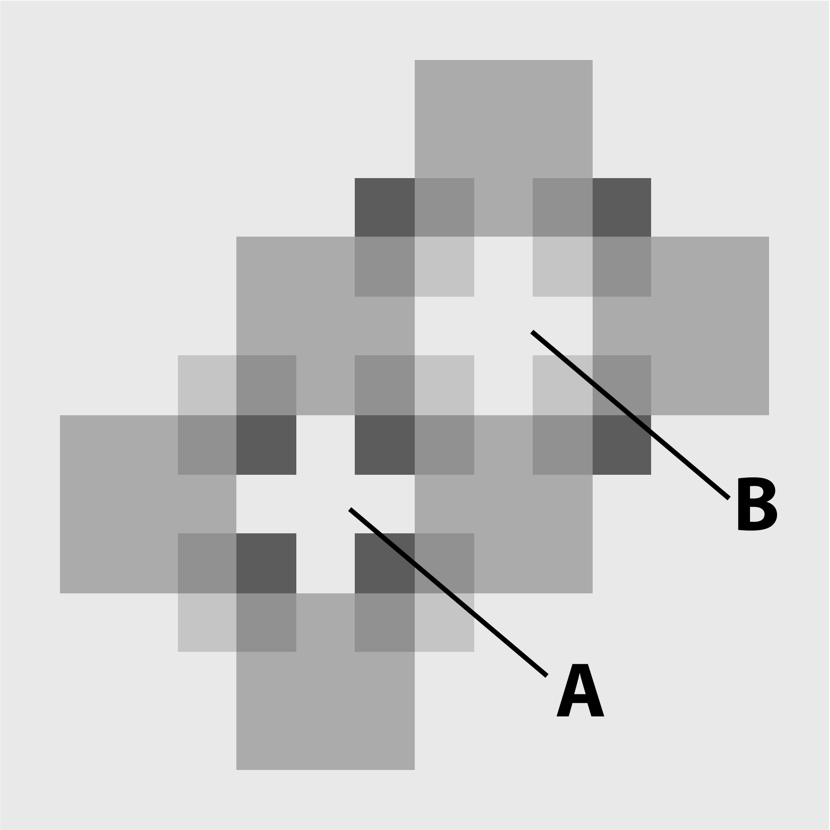


Figure 2.6. The haze illusion (after Adelson, 2000). The two plus-shaped regions, A and B, are physically identical. Region A appears to be an empty aperture through which the background is visible, whereas region B appears to be a partly transparent material.

Our second main finding is that observers had good lightness constancy when viewing computer-generated scenes in Experiment 2. The existing literature suggests that we should be cautiously optimistic about using rendered scenes in perceptual experiments. Morgenstern et al. (2014) compared the results of previous experiments on lightness constancy, some of which used computer renderings of naturalistic scenes and some of which used physical objects as stimuli. They found that observers showed qualitatively similar lightness matching behaviour in the two types of experiments, although constancy was weaker with computer-generated stimuli. In experiments on the closely related topic of illumination discrimination, Radonjić et al. (2016) found that observers showed nearly identical performance with computer-rendered stimuli and real scenes.

However, our findings are surprising in that observers showed significantly *higher* Thouless ratios, and so significantly *better* lightness constancy, with computer-generated scenes than with real scenes. One observer (KE) in Experiment 2 even had Thouless ratios greater than one, which suggests that the high Thouless ratios we found for most observers in Experiment 2 may not have been simply due to the computer-generated stimuli being perceived as very realistic. These results may be related to Agostini and Bruno's (1996) finding that contrast effects are larger in small, bright images in dark environments, such as the computer-generated stimuli in Experiment 2. Also related is McNamara, Chalmers, Troscianko, and Gilchrist's (2002) finding that observers may or may not perceive the intended reflectance in computer-generated images of surfaces, depending on seemingly minor adjustments in the rendering algorithm, such as the number of simulated interreflection bounces.

We speculate that the large, white, textured panel in the rendered images (Figure 2.4), which was the brightest object in the scene and in the otherwise unlit testing room, may have been perceived to be self-luminous, i.e., to be emitting light instead of only reflecting light from external sources. Bonato and Gilchrist (1999) found that when the luminance of a surface that observers perceived to be glowing was increased, the strength of perceived glow increased, and the perceived reflectance of other, non-glowing surfaces in the scene *decreased*. (See Murray (2013) for a Bayesian explanation of why this might be a rational response for the visual system.) If observers had perfect lightness constancy, then as the illuminance on the left-hand side of the panel increased, they would have held the match reflectance constant. Instead, Figure 2.5 shows that as the illuminance on the left increased, observer KE *increased* the reflectance of the match hexagon in order to maintain a constant perceived lightness (i.e., to match the perceived reflectance of the reference hexagon). This is precisely what Bonato and Gilchrist's findings would predict, if the panel appeared self-luminous. Furthermore, we do not believe that observer KE was completely anomalous, because in pilot sessions where the simulated reflectance of the panel was 1.00 (in Experiment 2 it was 0.78), we found that it was relatively common for observers to have Thouless ratios greater than one.

Computer monitors are light-emitting surfaces, and in many experimental settings that use computer-generated images as stimuli, this is perceptually obvious: even if stimuli are rendered and displayed realistically, as judged by current standards, observers are unlikely to mistake virtual objects for real objects, and they can see that virtual objects are not simply reflective surfaces. This was certainly the case in our Experiment 2: we carefully constructed the stimuli by rendering the apparatus from Experiment 1 using physically accurate rendering software, but nevertheless the stimuli were greyscale images viewed at close range (45 cm), without stereo cues, in a dark room. As is usual in such experiments, it was obvious to observers that they were seeing a representation generated by a light-emitting monitor.

Thus we speculate that the unusually high Thouless ratios we found for most observers in Experiment 2 were due to typical (e.g., average Thouless ratio of 0.82 in Experiment 1) or weaker-than-typical lightness constancy being artificially inflated by the glow effects described by Bonato and Gilchrist. It will take further experiments to test this hypothesis, but we provisionally suggest that the high Thouless ratios in Experiment 2 are not strong evidence for normal lightness perception with computer-generated scenes. Furthermore, these findings suggest a problem that will need to be addressed when evaluating the realism of computer-generated displays, namely the possibility that two failures of lightness perception -- weak lightness constancy and glow artifacts -- may combine to produce behaviour that mimics typical lightness constancy.

2.5 APPENDIX

Here we show that in a lightness matching task, the set of match points corresponding to a fixed Thouless ratio τ forms a straight line with slope $(\tau - 1)^{-1}$ on axes where log illuminance is plotted against log reflectance, as in Figures 3 and 5.

In a lightness matching task, the reference patch has reflectance r_R, illuminance i_R, and if the patch is Lambertian, luminance $l_R = r_R \, i_R/\pi$. Similarly, the observer's match setting has reflectance r_M, illuminance i_M, and luminance $l_M = r_M \, i_M/\pi$. An observer who has perfect lightness constancy makes match setting $r_M = r_R$. An observer who has no lightness constancy, and simply matches image luminance, makes match setting r_0, given by $r_0 \, i_M/\pi = r_R \, i_R/\pi$, or $r_0 = r_R \, i_R/i_M$. The Thouless ratio is given by Equation (1) in the main text, and using the above definitions is equal to

$$\tau = \frac{\log(r_M) - \log(r_R \, i_R/i_M)}{\log(r_R) - \log(r_R \, i_R/i_M)}$$

Solving for $\log i_M$ as a function of $\log r_M$,

$$\log i_M = (\tau - 1)^{-1}(\log r_M - \log r_R) + \log i_R$$

Thus for a fixed Thouless ratio τ, the set of all match points $(\log i_M, \log r_M)$ forms a line through $(\log r_R, \log i_R)$ with slope $(\tau - 1)^{-1}$.

Chapter 3

TEXTURE AND COLOUR CUES FOR EMISSIVE SURFACE DETECTION

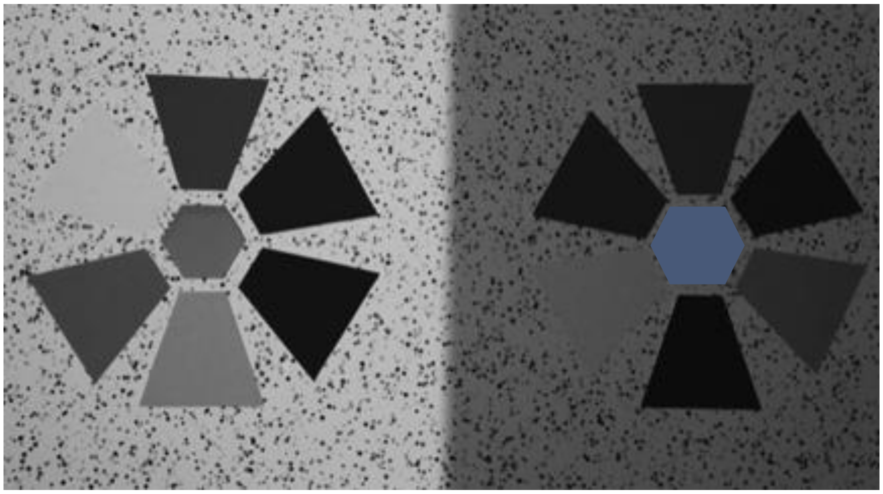
High-resolution LCD screens can display realistic scenes, but even under restricted viewing conditions (e.g., monocular, stationary), we can usually tell that the objects shown are not real. One explanation may be that we recognize that the screen *emits* light instead of *reflecting* incident light. Having shown that lightness perception differs in the real world vs. computer screen, in Chapter 2, we now investigate visual cues that allow us to differentiate between real and computer-rendered surfaces.[2]

---

[2] Parts of the following chapter have been presented at the Vision Sciences Society annual meeting (Patel, Palatnic, & Murray, 2017) and the European Conference on Vision Perception (Patel, Palatnic, & Murray, 2017).

3.1 INTRODUCTION

Originally, to test the prediction of Logvinenko and Maloney (2006) as described in Chapter 2, I intended to use an LCD screen embedded in a physical scene rather than build the motor driven apparatus that displayed real paper as shown in Figure 3.1. Pilot observations revealed that it is surprisingly difficult to make rendered surfaces indistinguishable from real surfaces. For example, when a textured piece of paper is scanned or photographed and displayed on an LCD screen, we are easily able to identify that it is a virtual image of paper and not a physical piece of paper placed on an LCD screen. To test if an LCD screen could be used for a matching task, I cut two hexagonal apertures on a poster board as shown in Figure 3.1, which was illuminated by a projector brighter on the left side compared to right side. A piece of paper was placed behind the left-hand aperture, and through the second aperture an LCD screen could be seen. During pilot observations it was also evident that matching any gray shade displayed on the LCD screen to the real paper was challenging as it was clear that the LCD screen appeared to glow even when surrounded by a rich physical scene. Next, I placed a dark translucent paper on top of the computer screen which was viewed through the aperture. Again, no gray shade seemed to be a good match to the real paper, even though the translucent paper added a physical texture to the LCD screen. The observation that the LCD screen could not be matched to a physical paper patch set the course for the experiments described in this chapter. In a set of three exploratory experiments I test whether colour and texture cues are effective in making computer-generated stimuli appear realistic, and more difficult to distinguish from real paper. Findings from this chapter are applied in Chapter 4 to make an LCD screen appear more realistic in a lightness matching task using an apparatus similar to the one shown in Figure 3.1.



<u>Figure 3.1</u> An example of the apparatus used for pilot observations. The left-hand hexagon has a piece of paper behind the aperture and the right-hand hexagon has an LCD screen behind it.

Some previous work has examined the factors that affect realistic presentation of textured materials on artificial displays. Factors such as the test patch size and its local surround have been found to be key factors in the way that an emissive or reflective surface is perceived. For carefully replicated conditions between a patch of monitor and for a patch of reflective surface, Brainard and Ishigami (1995) found the effect of illumination to be the same. In local illumination context, such as a CRT monitor placed in a light booth, Fairchild and Lennie (1992) found that illumination does not affect the color appearance of the monitor. However, Brainard and Ishigami (1995) found that in a global context, where illumination covers the entire room where the experiment is being run, illumination does affect the color appearance of the monitor. This effect was stronger when the stimuli occupied the entire monitor (here no local context, just the global illumination context is present) and weaker but still present if the stimuli did not occupy tne entire monitor.

One explanation for our ability to distinguish between light-reflecting and light-emitting surfaces may be that the chromaticity of the emissive surface (i.e., the LCD screen) does not exactly match the room white point. *White point* refers to the chromaticity of a white paper or surface under certain lighting. There are two different ways to match the white points: either to match the chromaticity of a patch on an LCD screen to the white point of the room or match the white point of the room to match the chromaticity of a patch on an LCD screen. Either method should be equally effective as long as the chromaticity of the LCD screen patch matches the white point of the room.

Another reason emissive surfaces like an LCD screen look different from paper may be that subtle shadow and lighting cues from a texture displayed on the screen may not match the lighting conditions of the scene in which the emissive surface is placed. This may be the telltale cue to distinguish emissive from reflective surfaces. Two ways to adjust the texture of the screen are to reproduce the texture of physical paper on the screen (e.g., by scanning a piece of paper) or to place a translucent textured surface over the screen. In the first case, it is very difficult to get the local highlights and shadow cues correct as flatbed scanning of a piece of paper uses a light perpendicular to the surface, while the lighting in the room is usually from the above. The second option, placing a translucent paper over the LCD screen, might be more convincing as it provides a real 3D texture, but it is a dark overlay which would limit how bright the brightest patch could be.

In the three experiments reported in this chapter, I test the importance of colour and texture cues for equating the appearance of patches presented on an LCD screen to patches of coloured paper. In a nine-alternative forced choice (9-AFC) task, observers were presented with a combination of eight real paper patches and one LCD screen patch. The observers' task was to identify which of the 9 patches was the LCD screen, as the colour and texture cues were manipulated.

In Experiment 1, the chromaticity of a patch on an LCD screen is adjusted to match the white point of the room and a digital texture is applied to the screen patch from scanned pieces of paper. In a two-way factorial design, the effect of the colour and texture cues is tested individually and in combination. In Experiment 2, in a one-way design, the effect of matching the chromaticity of a patch on an LCD screen to the room white point and matching the room white point to the factory-set white point of the LCD screen are compared to each other and to a baseline condition in which the chromaticity of the LCD screen and room white point are not matched. In Experiment 3, the chromaticity of a patch on an LCD screen is matched to the white point of the room and a physical texture is applied to the screen in the same lighting setting as Experiment 1. Additionally, the experiments provide information about how subtle shadow and highlight cues, or lighting direction, might be used by our visual system to distinguish between emissive and reflective surfaces.

3.2 EXPERIMENT 1: Testing Digital Texture and Colour Cues

In Experiment 1, I tested whether matching the chromaticity of a patch displayed on an LCD screen to the room white point would improve the realism of surfaces displayed on the LCD screen in a 9-AFC task. I also tested the role of texture cues, using scans of physical papers displayed on an LCD screen. There were four conditions in this experiment: in a two-way factorial design, I test the effect of colour and texture cues individually and in combination.

Methods

Participants

Five naïve observers participated for compensation. The observers were 20 to 32 years old, and four were female. In all experiments reported in this chapter, observers gave written informed consent before the experiment and reported normal or corrected-to-normal vision. All procedures were approved by the York University Office of Research Ethics.

Stimuli

Observers viewed a cardboard board that measured 26.5 cm horizontally and 33.0 cm vertically, at a viewing distance of 3.2 meters, subtending 4.8° × 6.0° of visual angle (Figure 3.2). There were 27 such boards. Each board had nine 3.2 × 3.2 cm square apertures, each of which subtended 0.6° × 0.6°. Patches of grey paper were attached behind eight randomly-selected apertures on each board; the Appendix lists the manufacturers and names of these paper samples. The paper patches were picked randomly without replacement from the 12 samples shown in Figure 3.3. The ninth aperture was left empty; this could be at any of the nine locations and is referred to as the reference aperture (e.g., top left aperture in Figure 3.2). Three sets of boards were used, each with nine boards with the reference aperture from the first to the ninth window, yielding 27 randomized cardboard boards (3 sets × 9 empty aperture positions = 27 boards). The three sets of boards differed from one another, as the eight apertures other than the reference aperture showed different random samples of paper.

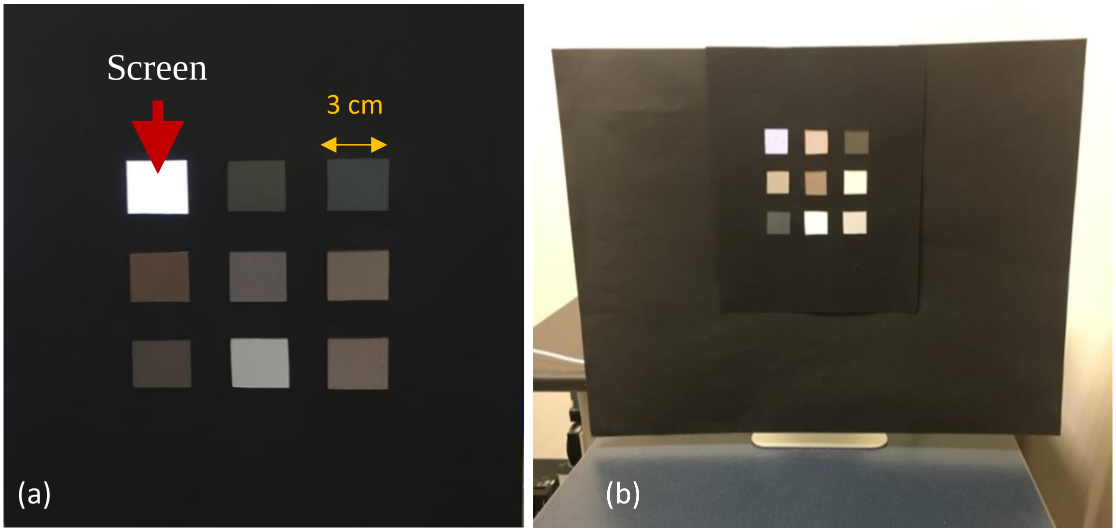


Figure 3.2 Example of a display used in the experiments. (a) shows a zoomed-in version of the board with nine apertures; (b) shows the experimental display.

Figure 3.3 Samples of the twelve paper types used in the experiments reported in this chapter. The Appendix section includes the details about the manufacturers and names of these paper samples.

There were four conditions in a two-factorial design with colour and texture as the two factors: C0T0, C1T0, C0T1 and C1T1. C0 in this nomenclature denotes a non-colour match, while C1 means that the white point of the LCD screen was matched to room white point. T0 indicates the absence of a texture on the LCD screen and T1 denotes that a texture was present. In the non-colour-match conditions (denoted by C0), the portion of the LCD screen visible through the aperture displayed a grey region with luminance matched to a randomly-chosen paper sample from the set shown in Figure 3.3, so only the brightness of the screen was matched. In the colour match conditions (denoted by C1), the CIE XYZ coordinates of the patch of LCD screen was matched to a randomly chosen paper in Figure 3.3. In the texture-absent (T0) conditions, the LCD screen displayed a plain, untextured patch, but in the texture-present (T1) conditions, the LCD screen displayed a flatbed scan of one of the randomly-chosen real papers from Figure 3.3.

The flatbed scans of paper were acquired using an Epson Perfection 2480 PHOTO scanner (300 dpi), down-sampled and shown on the LCD screen at the original size. The LCD screen model was late-2013 21.5" iMac, had a resolution of 1920 × 1080 pixels, a pixel size of 0.248 mm, and a nominal refresh rate of 60 Hz. The LCD screen's XYZ values were measured using a PR-655 SpectraScan spectroradiometer to measure the relationship between a range of RGB values and CIE XYZ coordinates, and the stimulus display software inverted this relationship to display the required colours.

In all four conditions, the source of lighting in the room was overhead fluorescent lights that were located 0.6m in front of the reference panel at a height of 2.8m. Under these lights, a Labsphere Spectralon 99% diffuse reflectance standard (model SRS-99-020) placed at the same location as the central aperture in Figure 3.2, had CIE Yxy coordinates $Y=63.31\text{cd/m}^2$, $x=0.4063$, $y=0.3965$. The chromaticity points of all of the stimuli, the room lighting, and the LCD screen can be visualized on the CIE xy chromaticity space shown in Figure 3.4. For both non-colour-calibrated conditions (C0T0 and C0T1), only the luminance of the stimuli was matched, so the chromaticity of the displayed patches (grey circles in Figure 3.4) was the same as the LCD screen's factory-set white point (denoted as ×). However, in the colour-calibrated conditions (C1T0 and C1T1) the chromaticity of the displayed patches (orange circles in Figure 3.4) was matched to the room white point (denoted as +).

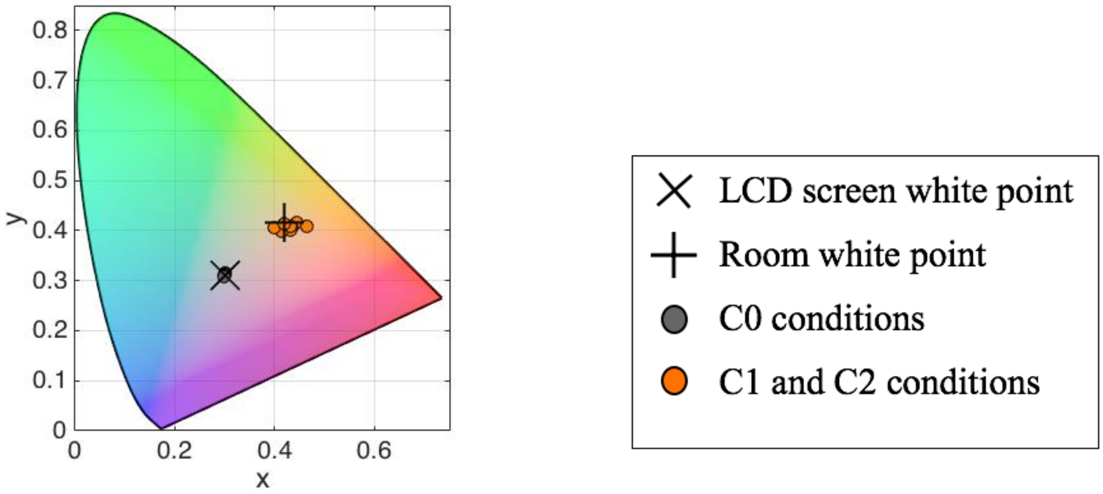


Figure 3.4 CIE xy chromaticity coordinates for non-colour-matched (C0) and colour-matched (C1 and C2) conditions. The stimuli in C0 were not colour matched, so their xy coordinates are close to the LCD screen's factory-set white point (x). In the C1 conditions the chromaticity of the patch displayed on the LCD screen was matched to the room white point (+). The xy coordinates of the stimuli from conditions C1 are close to the room white point because the LCD screen and room lights are colour matched; these stimuli are slightly spread around the room white point because the stimuli (as shown in Figure 3.3) are not all perfectly gray, so the distribution reflects these slight changes in hue. The C2 condition is from Experiment 2, and will be explained below.

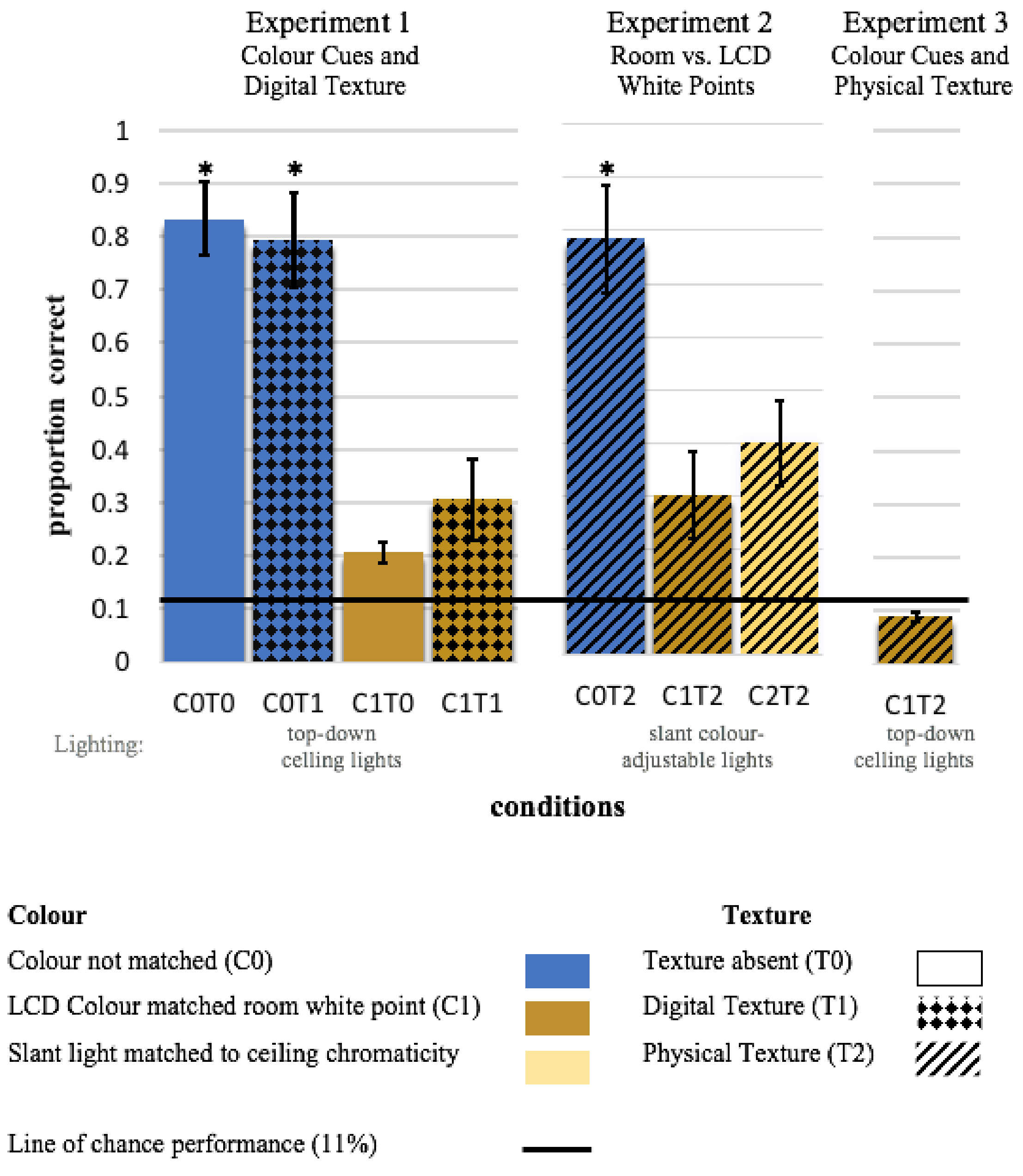
Procedure

Each observer ran in three one-hour sessions that included 108 trials per condition for a total of 432 trials. In each of the four conditions, the twelve simulated paper samples were displayed in all nine possible reference apertures once, which yielded the 108 trials per condition (9 possible empty apertures × 12 possible digital matches of paper samples = 108 trials). The 108 trials trials from the four conditions were interleaved in random order.

At the beginning of each trial, a motor-controlled black curtain blocked the experimental apparatus from the observer's field of view while the experimenter placed a new board, similar to the one shown in Figure 3.2, in front of the LCD screen. The curtain only blocked the observer from seeing the LCD screen; observers still received lighting cues about the scene from around the curtain in between trials. Once the board was placed, the curtain was rolled up and the observer performed a 9-AFC task. The observers judged which aperture had the LCD screen behind it and reported their choice by pressing one of nine keys. Observers typically took two to three seconds to make a judgment. At the end of each trial the curtain was drawn back down and the experimenter placed the board for the next trial. Observers were given 15 practice trials before the start of data collection. Observers used a chin rest placed at the same height as the LCD screen throughout the experiment to prevent head movements, as changes in viewing angle can cause changes in brightness and colour on the LCD screen.

Results and discussion

Figure 3.5 shows the proportion of correct responses in picking the aperture with the LCD screen behind it. Since it was a 9-AFC task, an accuracy of 11% is chance performance, meaning that observers were equally likely to pick the reference aperture or any of the eight physical papers. The solid black line in Figure 3.5 is the chance performance line. Mean proportion correct is quite high when the colour is not matched: C0T0 (M=84% and SD=15%) and C0T1 (M=80% and SD=20%), for no-texture and texture conditions, respectively. However, the proportion correct was lower in conditions where colour was matched C1T0 (M=21% and SD=4%), and C1T1 (M=31% and SD=17%), for no-texture and texture conditions, respectively. A 2×2 repeated measures ANOVA was conducted to test the influence of colour (2 levels: C0 and C1) and texture (2 levels: T0 and T1) on observers' mean proportion correct. The analysis reveals a significant main effect of colour ($F(1, N=24) = 24.32$, $p<0.01$). There was no significant main effect of texture or a significant interaction between the factors colour and texture. Post hoc t-tests using Bonferroni corrections revealed both C0 conditions, C0T0 and C0T1, had significantly higher mean proportion correct than both C1 conditions, C1T0 and C1T1, but C1T0 and C1T1 condition means were not significantly different from each other, with all comparisons at the $p<0.01$ level. Additionally, the addition of texture cues or the interaction between colour and texture did not have a significant effect on observers' mean proportion correct in this task. To conclude, the factor of colour had a significant effect on performance, so it can be concluded that matching the colour of the LCD screen patch to the room white point makes it more difficult for observers to choose the aperture with the LCD screen behind it.



<u>Figure 3.5</u> Results of Experiments 1-3. The stars on bars indicate performance significantly different from chance, assessed by one-sample t-tests with Bonferroni corrections which individually compare all eight conditions' mean proportion correct to the chance performance level of 11% (* $p < 0.05$).

Table 3.1

Bonferroni-corrected one-sample t-tests for all eight conditions in experiments 1, 2, and 3. The mean proportion correct from each of the eight conditions is compared to the chance level of 11%.

	Conditions	Chance Level = 11%					
		t	df	Bonferroni-Corrected p Value	Mean Difference (%)	Confidence Interval Lower(%)	Confidence Interval Upper(%)
Exp1	C0T0	11.00	4	0.003*	72.7	50.0	118.3
	C0T1	7.69	4	0.012*	68.7	32.8	126.6
	C1T0	4.77	4	0.071	9.5	10.0	31.0
	C1T1	2.64	4	0.459	19.2	-8.7	70.7
Exp2	C0T2	6.65	4	0.021*	67.5	25.2	131.8
	C1T2	2.29	4	0.664	19.1	-13.5	73.7
	C2T2	3.61	4	0.179	29.0	-2.1	82.2
Exp3	C1T2	-2.56	4	0.501	-2.1	4.5	13.2

$*p<0.05$

I conducted one-sample t-tests with Bonferroni corrections to individually compare all conditions' mean proportion correct in this chapter (eight significance tests) to chance performance level of 11%. If the mean proportion correct for any of the conditions is significantly different from chance performance, this means observers can distinguish between the LCD screen and real paper. The mean proportion correct in conditions C0T0 and C0T1 is significantly greater than the chance level of 11% ($t(4) = 11.00$, $p<0.05$ and $t(4) = 7.69$, $p<0.05$, respectively). Thus, in the colour-not-matched conditions C0T0 and C0T1, observers are able to distinguish LCD screens from real paper. On the other hand, the colour-matched conditions C1T0 and C1T1 are not significantly different from chance. However, because this is a null result, we cannot decisively conclude that the mean proportion correct in the C1T0 and C1T1 conditions is equal to chance.

3.3 EXPERIMENT 2: Testing the Importance of Room and LCD screen White Points

The purpose of Experiment 2 is to test whether colour matching either by adjusting the white point of the room lighting to the chromaticity matching on the LCD screen or vice versa would significantly improve the realism of the displayed stimuli as compared to not colour matching. It is possible that the textures shown in Experiment 1 were not sufficiently realistic due to the subtle shadow and highlight cues caused by different lighting directions in the room compared to the flatbed scaner. Therefore, for Experiment 2 a physical texture was used – namely, a piece of dark transluscent tracing paper. In addition to using a different texture cue, the lightning condition in Experiment 2 was also different from Experiment 1. In some conditions in Experiment 1, I colour calibrated the stimuli in order to match the displayed stimuli's chromaticity to the room white point. In a new condition in Experiment 2, I tried the reverse; that is, I adjusted the room white point to match the LCD screen's factory-set white point. I aim to compare a colour-non-matched condition and a colour-match condition under fluorescent lights, to a colour-non-matched condition in which the room light is adjusted to the LCD screen's white point. An additional goal is to compare whether both methods of matching the chromaticity of the LCD screen patch matches to the room white point are equally effective at improving the realism of the displayed stimuli.

Methods

Participants

Five naïve observers, none of whom had participated in Experiment 1, participated for compensation. The observers were 20 to 28 years old (three were female).

Stimuli

Observers viewed the same display as Experiment 1. The overhead fluorescent lights used in Experiments 1 were turned off and replaced by nine 60W Philips Hue White and Color Ambiance bulbs. Colour details for the bulbs are given after the descriptions of relevant conditions. The nine bulbs were placed 2 m in front of the experiment board (Figure 3.2) at a height of 2.8 m. The nine bulbs in this experiment were approximately evenly distributed over a distance of 1.8 m and produced lighting that was noticeably more diffuse than the overhead fluorescent lights used in Experiment 1. The lower placement of the bulbs in this experiment made the direction of lighting more perpendicular to the display than in the previous experiment.

In Experiment 1, stimuli were displayed on the LCD screen through the reference aperture. For all conditions in Experiment 2, a dark translucent paper (LCI Paper Co., Hudson, MA; 30 lb ebony black translucent vellum, CC811-10) was placed directly behind the reference aperture. The LCD screen could be seen through the translucent paper, which provided a physical texture. This physical texture is denoted by T2 in the condition name. The translucent paper was only placed behind the reference aperture; the other eight physical samples were taped directly behind the other apertures in the board, as in Experiment 1, and as shown in Figure 3.2.

Only the factor colour was varied across conditions, by either not matching the chromaticity of the LCD screen and room white point (condition denoted by C0) or by matching the chromaticity of the LCD screen and room white point (conditions denoted by C1 and C2). In the C0T2 condition, the greyscale RGB values of the screen were adjusted so that the luminance of the translucent paper matched a randomly chosen paper sample, and the chromaticity matched the factory-set white point of the LCD screen, similar to condition 1 in Experiment 1. The room white point was adjusted so that the xy-chromaticity of the bulbs was set to a florescent colour, to match the chromaticity of the overhead lighting in Experiment 1. For C1T2 the colour (i.e., CIE XYZ coordinate values) of the screen was adjusted so that the colour of the translucent paper matched a randomly chosen paper sample, similar to condition C1 in Experiment 1; again, the room white point was adjusted so that the xy-chromaticity of the bulbs was set to a florescent colour. In conditions C0T1 and C1T2, the nine bulbs were set to the same Yxy coordinates as the florescent lights in Experiment 1 in order to keep the colour of the lighting consistent with Experiment 1. In condition C2T2, the bulbs were set to the same Yxy as the factory-set LCD screen white point. The LCD screen white point was determined by measuring the xy chromaticity coordinates of greyscale RGB values across the range 0-255 displayed on the LCD screen. A Labsphere Spectralon 99% diffuse reflectance standard (model SRS-99-020) placed at the same location as the central aperture in Figure 3.2 had a Yxy reading of Y=63.31cd/m^2, x=0.4063, y=0.3965 in conditions C0T2 and C1T2. For the C2T2 condition, the greyscale RGB values of the screen were adjusted so that the luminance of the translucent paper matched a randomly chosen paper sample, and colour matching was accomplished by matching the white point of the room to the factory-set white point of the LCD screen. The 99% reflectance standard had a Yxy of Y=63.31cd/m^2, x=0.3049, y=0.3228 in condition C2T2.

Procedure

The procedure was the same as in Experiment 1. Since this experiment had three conditions, observers were run in three blocked 45-minute sessions of 108 trials per session, totaling 324 trials overall. Each condition was run in a separate session. The order of conditions was randomized for each observer. The experimenter changed the associated light colour in the room at the beginning of each session, before the observers entered the room.

Results and Discussion

Figure 3.5 shows the mean proportion correct for observers in the condition C0T2 (M=79% and SD=23%), C1T2 (M=30% and SD=19%), and C2T2 (M=40% and SD=18%). A one-way repeated measures ANOVA was conducted to test the influence of colour (3 levels: C0, C1, and C2) on observers' mean proportion correct. The analysis reveals a significant main effect of colour ($F(2, 12) = 8.30$, $p<0.01$). Post hoc tests using Bonferroni corrections reveal that condition C0T2 has significantly higher mean proportion correct when compared to either condition C1T2 or C2T2, with all comparisons at $p<0.01$ levels. However, the mean proportion correct in conditions C1T2 and C2T2 do not significantly differ from each other. Therefore, it be can concluded that matching the room white point to the white point of the LCD screen makes observers significantly worse at picking the aperture with the LCD screen behind it, compared to a condition with no white point calibaration (C0T2).

As shown in Table 3.1, one-sample t-tests with Bonferroni corrections across all eight conditions from this chapter revealed that the mean proportion correct in condition C0T2 is significantly different from chance ($t(4)=6.65$, $p<0.05$). However, the means conditions C1T2 and C2T2 are not significantly different from chance ($t(4)=2.29$, $p>0.05$ and $t(4)=3.61$, $p>0.05$, respectively). Therefore, in condition C0T2, observers are able to distinguish the LCD screen from a real piece of paper. On the other hand, in conditions C1T2 and C2T2, there is no evidence that the LCD screen can be distinguished from paper.

Based on the ANOVA and the t-test we can conclude that matching the chromaticity of the LCD screen and the room white point reduced observers' mean proportion correct compared to not matching the chromaticity of the LCD screen and the room white point. Additionally, the mean proportion correct for C0T2 is significantly higher than chance performance. However, observers are unable to distinguish virtual and real paper in C1T2 and C2T2 conditions.

From Experiments 1 and 2 it appears that adding a realistic texture and matching the chromaticity of a displayed patch to the room white point are important to improve the realism of patches displayed on LCD screen. Nonetheless, comparing the texture adjustment across Experiment 1 and 2 is difficult because the lighting direction is another factor that changed which might be the cause of differences in observer performance rather than texture; the lighting direction in Experiment 1 is top-down, whereas in Experiment 2 it is more perpendicular to the display. To compare the difference in mean proportion correct values across the two texture conditions T1 and T2, the C1T2 from this experiment is replicated in the next experiment with a different texture T2.

3.4 EXPERIMENT 3: Testing Physical Texture and Colour Cues

The purpose of this experiment is compare the mean performance to chance performance when the LCD screen chromaticity is matched to the room white point and a physical texture is added. The current experiment consisted of one condition (denoted by

C1T2) in which a piece of dark transluscent paper (T2) was used as a texture under the same lighting conditions as in Experiment 1.

Methods

Participants

Five naïve observers, who had not participated in the previous two experiments, participated for compensation. The observers were 19 to 37 years old (four were females).

Stimuli

Observers viewed the same display under the same ceiling lights as in Experiment 1. A dark translucent piece of paper was placed directly behind the reference aperture, so the LCD screen could be seen through the translucent paper, as in Experiment 2. Additionally, the colour (i.e., CIE XYZ coordinate values) of the screen was adjusted so that the colour of the LCD screen seen through the translucent paper matched a randomly-chosen paper sample. Conditions C1T2 from Experiment 2 and the current condition C1T2 are similar in all regards except for the lighting direction: rather than using the colour adjustable lighting from experiment 2, the current experiment uses the same top-down celling lights as Experiment 1. This condition and the condition C1T1 from Experiment 1 were also similar in all aspects, except that here the texture cue came from a piece of translucent paper placed over the screen in C1T2, instead of a digital texture from a flatbed scan. For Experiment 3, the stimuli xy coordinates also lay close to the room white point (+), similar to the orange circles in Figure 3.4 in Experiment 1.

Procedure

I followed the same procedure as in Experiment 1. Since this experiment had only one condition, observers were run in one 45-minute session of 108 trials.

Results and discussion

Figure 3.5 shows the mean proportion correct for observers in the condition C1T2 (M=8.9%, SD=2%). One-sample t-tests with Bonferroni corrections across all eight conditions, reveals that the mean proportion correct in condition C1T2 is not significantly different from the chance performance of 11% ($t(4) = -2.56$, $p>0.05$, 95% CI [4.5,13.2]). Therefore, when the transluscent tracing paper was added as texture alongside the colour calibaration observers were not significantly above chance at choosing the LCD screen as compared to real paper, but it cannot be decisively concluded that the mean proportion correct for the C1T2 condition is actually at chance, as it is a null result.

A paired sample t-test was also conducted to compare the mean proportion correct when using the scans as texture in condition C1T1 from Experiment 1 and a piece of translucent paper as texture in the current C1T2 condition. A significant difference in the mean proportion correct was found when using the two different types of texture $t(4) = 4.04$, $p<0.05$. Therefore, using a piece of translucent paper may be a better choice for providing texture cues, if realism is the goal.

As noted earlier, the C1T2 condition in Experiment 2 is similar to the C1T2 condition in the current experiment in all aspects expect for the lighting direction. On an informal note, at the end of Experiment 2 most observers reported always picking the patch that appeared more textured than the rest, which was not the case in either Experiment 1 or 3. A paired sample t-test conducted to compare the mean proportion correct for both C1T2 conditions in Figure 3.5, revealed that the current C1T2 condition with the top down lighting is significantly different from the C1T2 condition with a more fronto-parallel lighting direction t(4) = 2.86, $p<0.05$. The possible reason that the translucent paper texture was more noticeable in Experiment 2 was because the lighting used in Experiment 2 was more ambient in nature as the nine bulbs were spread over a distance of 1.8m. Additionally, the light direction from these 9 bulbs was not top-down but more perpendicular to the display than in Experiment 2. In Experiments 2, the combination of the lighting being more ambient and coming from a lower angle relative to the screen made real paper surfaces appear more matte than when the overhead lights were used in Experiments 1 and 3. However, in the reference patch, the LCD screen light shone through the transluscent tracing paperand would have appeared equally textured in both Experiments 2 and 3. Consequently, for experiment 2 the texture of the reference patch and the the rest of the patches were noticably different, which could have acted as a confound.

3.5 GENERAL DISCUSSION

The experiments in this chapter allow us to isolate the cues that enable observers to distinguish between LCD screens and physical paper surfaces. It is evident from Experiment 1 that colour is the driving cue in improving the realistic appearance of LCD screens. From Experiment 2 we learn that matching the LCD screen chromaticity to the room white point is important, but it is less important how the matching takes place. From Experiment 3, it is evident that a translucent paper is a better texture cue than scans of paper displayed on an LCD screen. The findings provide a basis for experimenters to use a translucent piece of paper placed in front of a computer and colour matching to improve the realism of displayed stimuli in experiments.

Making an LCD screen indistinguishable from paper is a much more difficult task than expected. With regards to texture cues in Experiment 1, it is evident that the observers are more sensitive than expected to the subtle highlights and shadow cues on the displayed digital texture. Since the scans used to provide the texture in Experiment were flatbed (or front-facing), while the lighting in the room for Experiment 1 was top-down, perhaps the highlights and shadows generated by these different lighting directions were visible to the observers. Recent development of lighting sensitive displays which render a 3D scene such that it always appears to be lit by the lighting from real environment (Nayar, Belhumeur & Boult, 2004) could be used to illuminate the flatbed scans for a more realistic texture appearance, in future studies. Additionally, if a translucent piece of paper texture is to be used in a design similar to this chapter, the lighting direction is crucial for the texture overlay to appear similar to other real paper textures. In hindsight, it seems that a design other than a 9-AFC task may be better suited to test if LCD screens can be made indistinguishable from paper, as the choice of texture and lighting direction acted as confounds. Our long-term goal in this project is to test whether it's feasible to use a patch of virtual paper in a lightness matching experiment. As already discussed in Chapter 2, computer generated stimuli often do not produce the same level of lightness constancy as real stimuli. I aim to improve methodology, so computer displays appear more realistic. Chapter 4 builds on these exploratory findings and extends the applications of adjusting colour and texture cues in a lightness matching experiment, using a free adjustment paradigm similar to Chapter 2.

3.6 APPENDIX

The following is a list of the twelve paper samples used as stimuli. 3.2 × 3.2 cm square apertures of these papers were attached to eight windows in the experimental display as shown in Figure 3.2, and the XYZ chromaticity of these papers were used to generate the colour matches on the LCD screen in the three experiments.

Paper Number	**Paper Name**
1	Canson MI-tenites #111 ivory
2	Canson MI-tenites #120 pearl gray
3	Fabriano bristol board FAL-702
4	Canson flannel gray #122
5	Strathmore cool gray #500
6	Bainbridge white core cardboards BN22-4
7	Canford paper gun metal
8	Novacore whitecore warm gray BN32-4
9	Canson MI-tenites sepia #133
10	Strathmore dark gray #345
11	Fabriano brstol board FAL-122
12	Fabriano tiziano 30 antracite

Chapter 4

CUES FOR EQUATING LIGHTNESS MATCHING PERFORMANCE ON REAL AND VIRTUAL SURFACES

In Chapter 3, observers' ability to distinguish between real and virtual patches of paper was measured. Here, the aim is to test whether observers make equivalent lightness matches with real and virtual patches of paper when colour and texture cues are adjusted.[3]

---

[3] Portions of the following chapter have been presented at the Vision Sciences Society annual meeting (Patel & Murray, 2018). The body of this text has been submitted to Journal of Vision.

4.1 INTRODUCTION

As shown in Chapter 3, matching the chromaticity of an LCD screen to the room white point and displaying screen patches through a physical texture improved the realism of a patch displayed on an LCD screen. In a lightness matching task, such as the one in Chapter 2, an observer matches the shade of grey of a reference patch when the match and reference patches are placed under different lighting. As evident in Chapter 2, observers are partially lightness constant in such a lightness matching task with physical surfaces. In this chapter, the knowledge about the role of colour and texture cues in the perception of an LCD screen is applied to see if it is possible to get equivalent lightness matching performance in a screen-based task and a task using physical paper.

Katz (1935) introduced the concept of articulation and how it affects lightness constancy. Articulation refers to the degree of complexity present in a scene, namely the number of surfaces in a scene possessing a range of reflectances, shapes, and 3D curvatures, and how much area they occupy. A Mondrian (Figure 4.1) is a well-known example of a highly articulated surface used in many lightness and colour matching experiments (Gilchrist & Annan, 2002; Land, 1977; Rudd, 2010). Despite the implications of the term ‘lightness constancy,’ asymmetric lightness matching (lightness matching under differently-lit areas) is not perfect. As is evident from Chapter 2, when observers match shades of gray under different illuminants, illumination affects their choice of gray shade. Katz reports that as the degree of articulation increases, the strength of lightness constancy also increases in a scene with varying illumination. Researchers in the field generally accept the idea that articulation strengthens lightness constancy (Adelson & Pentland, 1996; Gilchrist & Annan, 2002; Goldstein, 1987). Consequently, it is standard practice in the field to test theories of lightness in highly-articulated scenes. The experiment in this chapter tests the degree of lightness constancy using a highly-articulated physical scene where the reference patch is displayed on an LCD screen.

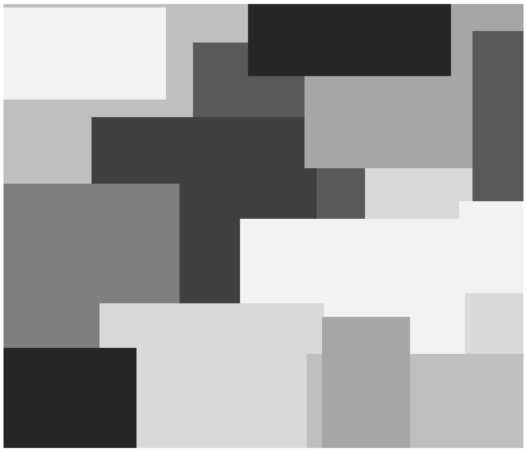


Figure 4.1 An achromatic version of a Mondrian (based on Rudd, 2010).

In order to test lightness constancy with high external validity, it is ideal to perform experiments with physical materials of known reflectances. Historically, when using a physical display, researchers have tested theories of lightness by asking observers to perform lightness matching between pieces of papers of different shades of gray under different lighting conditions. As explained in Chapter 2, matching physical surfaces is slow, therefore we use a free adjustment task in this design. Also, as seen in section 2.2, assembling a physical setup can be cumbersome. This has led to many experiments in psychophysics being performed entirely on CRT monitors or LCD screens. However, as we noted in Chapter 2, the lightness constancy of human observers looking at simulated surfaces on screens can be significantly different from that observed in real scenes.

In the past, researchers have used a combination of real and virtual surfaces to test colour and lightness related theories. For example, Rutherford and Brainard (2002) used an LCD monitor through a gray gelatin filter and plastic diffuser to perform colour matching experiments. Kraft and Brainard (1999) used adjustable spotlights shone on a small rectangular patch in a physical scene for their colour matching task. Researchers have also occasionally used a simulated match patch as a part of a physical apparatus (Jacobsen & Gilchrist, 1988). In this chapter, I investigate the use of an LCD screen as part of a larger physical setup. An articulated scene as shown in Figure 4.2 is used to measure lightness constancy in an asymmetric lightness matching task; observers adjust a patch on the LCD screen to match a reference patch in the real world. In the experiments described here, I used the results of Chapter 3 to modify the LCD screens, to make convincing lightness matches in a realistic scene. Lightness constancy is also measured using physical paper in place of the LCD screen, which makes it possible to acquire baseline lightness constancy with real paper matches. In a two-factorial design, colour and texture manipulations are made to LCD screen, similar to Experiment 1 in Chapter 3. The colour matching is accomplished by matching the chromaticity of the LCD screen to room white point and the texture cue used is a dark transluscent paper. I compare the lightness constancy for each of the LCD screen manipulations to the lightness constancy for physical paper matches. This comparison tests whether the addition of colour and texture cues are important to obtain similar lightness constancy levels on LCD screens and physical paper .

4.2 METHODS

Participants. Six naïve observers participated for payment. The observers were between 19 and 33 years old; three were female. Observers gave written informed consent and reported normal vision. All procedures were approved by the Office of Research Ethics at York University.

Stimuli. Observers viewed a panel of poster paper mounted behind a wooden frame as shown in Figure 4.2. The wooden panel and its supports were painted using acrylic matte gray paint with a reflectance of 0.09 (i.e., 9%). The poster panel measured 100 cm horizontally and 50 cm vertically at a viewing distance of 150 cm, subtending 18.9° × 36.9°. The poster panel had a background reflectance of 0.25 and was printed with a field of randomly-placed circles, rectangles, and triangles with heights ranging from 2 to 8 cm and reflectances ranging from 0.08 to 0.75. The panel also had two 7-cm long by 5-cm high apertures, which subtended 2.7° × 1.9°. The distance between the centers of the two apertures was 40 cm (15.2º). The left-hand rectangle was the reference aperture and the right-hand rectangle was the match aperture. Two cubes and pyramids made from matte printed paper were also placed in front of the poster panel as shown in Figure 4.2. The reflectances of these geometric objects were (from left to right): 0.20, 0.25, 0.46 and 0.58. A chin rest was placed 150 cm away and at the same height as the LCD screen throughout the experiment to prevent head movements which can cause changes in brightness and colour on the LCD screen.

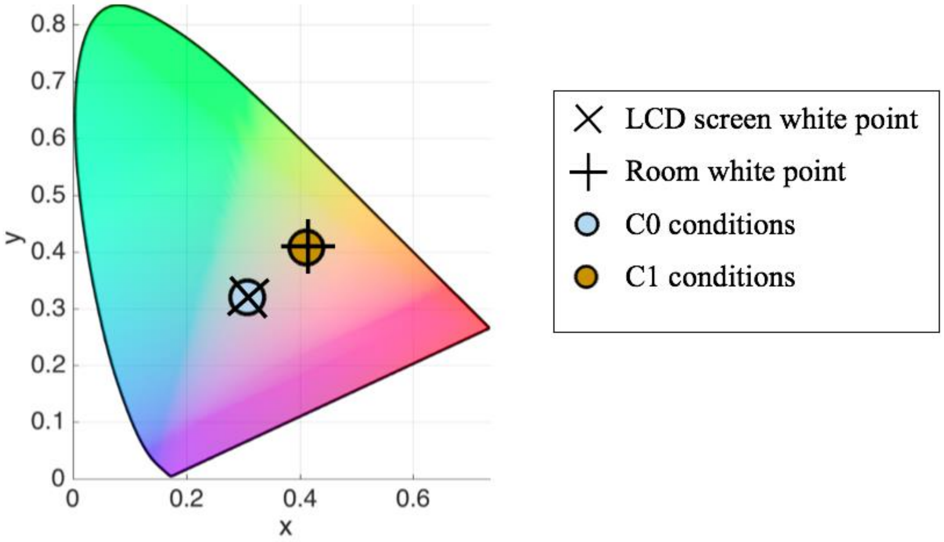
Immediately behind the reference aperture, flat against the back of the main panel described above, was a circle of poster paper with a diameter of 31.0 cm. The circle was divided into three equal sectors with luminance values of 38.1, 60.8 and 101.6 cd/m^2. A computer-controlled motor rotated the circle so that a uniform section of paper with one of the three reflectances was visible through the reference aperture.

Figure 4.2 Photograph of the experimental apparatus. The left rectangle is the reference aperture and the right rectangle is the match aperture.

Figure 4.3 All five match patch conditions shown in a magnified version of the match aperture in Figure 4.2. The physical match condition has a real paper sample behind the aperture, while the other four conditions have an LCD screen behind the match patch aperture.

There were five conditions which differed in whether the match patch displayed through the aperture was made of real paper or was displayed on an LCD screen. Figure 4.3 shows a sample match patch from each condition. In the physical match condition, immediately behind the match aperture, flat against the back of the panel, was a conveyor-belt-like loop of poster paper, 152 cm long and 11 cm wide. The reflectance of the conveyor belt ranged continuously from 0.06 to 0.80. Reflectance increased nonlinearly along the belt, in such a way that reflectances were evenly spaced according to the Munsell scale (Schanda, 2007). A computer-controlled motor could advance the conveyor belt so that any section of the belt was visible through the match aperture. Vertically, just 5 cm of the 152 cm belt was visible through the match aperture at any time, so the visible segment of paper always had approximately uniform reflectance, and the reflectance gradient was not perceptible. For all the other conditions, an LCD screen was placed flat against the back of the panel. The LCD screen model was a late-2013 21.5" iMac with a resolution of 1920 × 1080 pixels, a pixel size of 0.248 mm, and a nominal refresh rate of 60 Hz. The LCD screen's colour characteristics were measured using a PR-655 SpectraScan spectroradiometer to measure the relationship between RGB values and CIE XYZ coordinates, and the stimulus display software inverted this relationship to display the required colours.

In conditions in which the LCD screen was displayed through the match aperture the two factors, colour (denoted by C in the name of the condition) and texture (denoted by T) were crossed over four conditions: the chromaticity of the LCD screen was either matched to the room (denoted by C1) or not matched (denoted by C0), and a textured film was either present over the screen (denoted by T1) or absent (denoted by T0). In the C1 conditions, in which colour was matched, all match patches were displayed at the room chromaticity (i.e., the chromaticity of a white patch of paper at the locations of the test and match apertures), x=0.4063 and y=0.3965 (CIE coordinates). In the C0 conditions, in which colour was not matched, the factory set chromaticity of the LCD screen x=0.3049 and y=0.3228 were used; the luminance for these patches could be adjusted. Figure 4.4 shows the CIE xy chromaticity of the LCD screen for the colour match (C1) and colour non-match (C0) conditions which are matched to the room white point and the LCD screen factory set white point, respectively. In the T1 condition, a dark translucent piece of paper (LCI Paper Co., Hudson, MA; 30 lb ebony black translucent vellum, CC811-10) was placed in front of the LCD screen behind the match aperture and provided the texture to the stimulus displayed on the LCD screen. In the without-texture condition (T0), there was no paper covering the LCD screen.



<u>Figure 4.4</u> CIE xy chromaticity coordinates for colour non-match (C0) and colour match (C1) conditions. The stimuli in C0 were matched to the LCD screen's factory set white point (x) and C1 is matched to the room white point (+).

The poster panel, three-sector circle, and conveyor-belt-like loop of poster paper were created by printing them on a poster printer that we had characterized by measuring the mapping from greyscale RGB values to reflectances. Reflectances were measured using a PR-655 SpectraScan spectroradiometer and a Labsphere Spectralon 99% diffuse reflectance standard (model SRS-99-020). Under constant illumination of the target surface the reflectance was computed as following:

$$\text{reflectance} = 0.99 \times \frac{\text{luminance of the target surface being measured}}{\text{luminance of the reflectance standard}}. \qquad (2.1)$$

The apparatus was illuminated by overhead fluorescent lights, and by an LCD data projector (Epson PowerLite HC2040). The overhead fluorescent lights were placed 0.30 m in front of the reference panel at the same height as the bottom of the poster panel. Under the fluorescent lighting, a 99% diffuse reflectance standard placed at the same location as the apertures in Figure 1 had an Yxy reading of Y=82.03cd/m^2, x=0.4063, y=0.3965. The projector was placed xx cm in front of the panel. The projector's gamma function was characterized using a 99% diffuse reflectance standard and a PR-655 SpectraScan spectroradiometer to measure the relationship between RGB values and colour (CIE XYZ coordinates), and the stimulus display software inverted a fitted gamma function to display the required colour (CIE XYZ coordinates). Throughout the experiment, lighting on the left-hand side of the main panel was fixed, with an illuminance of 320 lux at the reference aperture (Figure 4.2, left-hand aperture), meaning that a Lambertian surface with a reflectance of 1.0 would have a luminance of 102 cd/m^2. Lighting on the right-hand side of the main panel was two times brighter and fixed at 637 lux or 203 cd/m^2. The lighting boundary between the left- and right-hand sides was created by the data projector, which projected a two-tone image that was brighter on the left than on the right. The projector was slightly defocused so that no pixelation was visible in the projected image, and so that the lighting boundary was penumbra-like and unlikely to be mistaken for a reflectance boundary.

Procedure. Each observer ran in five sessions of approximately 45 minutes that included 90 trials per condition for a total of 450 trials. Each session consisted of one of the five conditions, with the order of sessions randomized for all observers. In each condition, the 90 trials consisted of 30 trials per reflectance level of the reference patch (reflectances of 0.19, 0.30 and 0.50), which were interleaved in random order.

At the beginning of each trial, the computer-controlled motor rotated the circle placed behind the reference aperture so that a uniform section of paper with one of the three reflectances was visible through the reference aperture. Observers used the up and down arrow keys on a keyboard to increase or decrease the luminance of the match patch. They were instructed to adjust the shade of grey of the match patch until it matched the reference patch by asking them to ‘try to make it look like the two like the two apertures were cut out of the same paper’. Once satisfied with their choice, observers pressed the return key to submit their response. The average reaction time for each trial was 14 seconds. The overhead lights and the projector were kept on throughout the experiment. Observers were given 10 practice trials before the start of data collection to get them accustomed to the appratus and ensure they understood the task. At the end of each condition observers were asked to rate how satisfied they were with their matches throughout the session on a scale from 1 to 10, with 10 being the most satisfied. They were also given a chance to revise their ratings at the end of the experiment after they had seen all of the conditions.

4.3 RESULTS

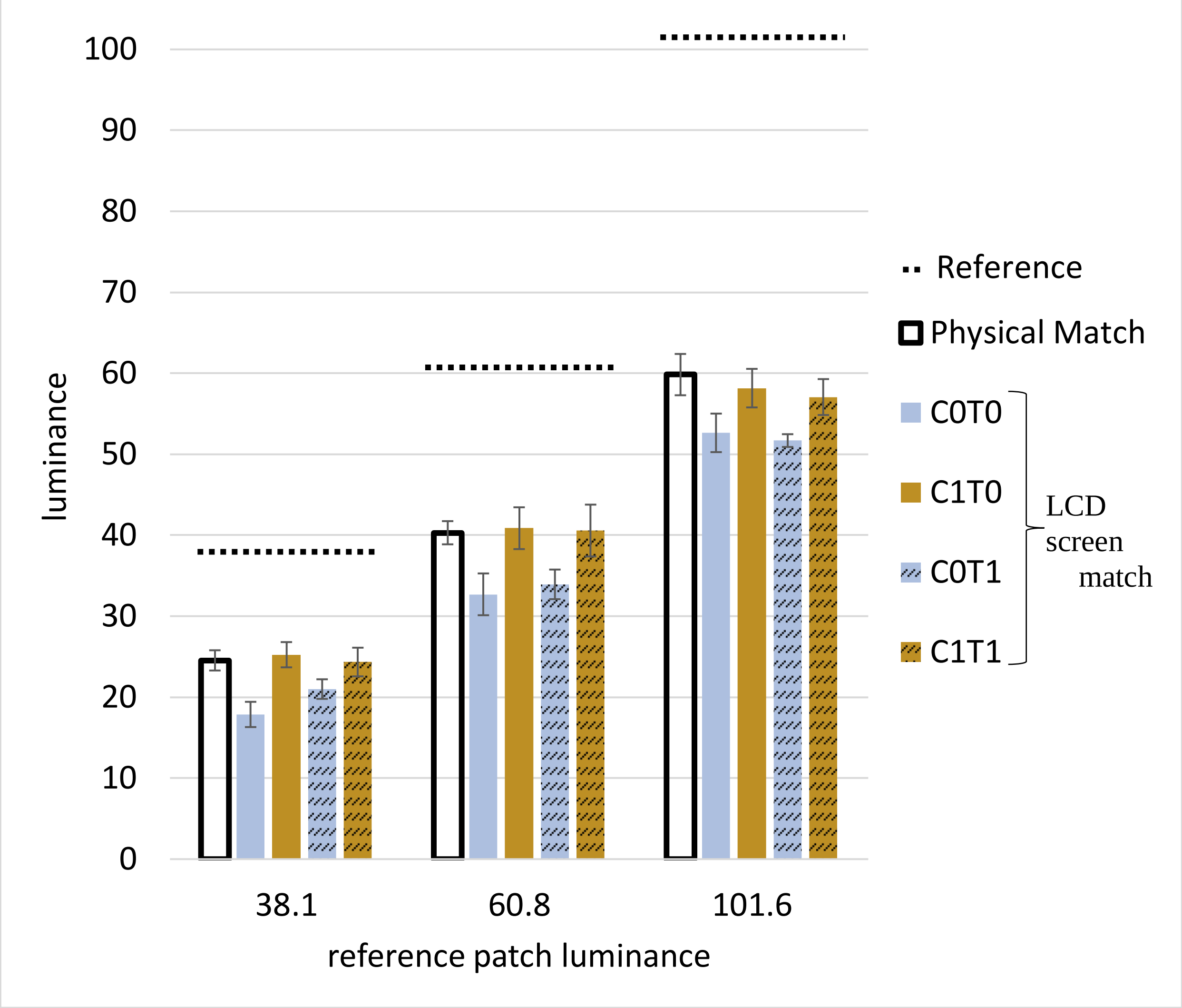


Figure 4.5 Average match patch luminance values for six observers for three reference patch luminance values 38.1, 60.8 and 101.6 cd/m^2 (dark to light), across five conditions: one physical paper condition and four LCD screen manipulations (C0T0, C1T0, C0T1, and C1T1). The reference patch luminance is indicated by the dotted lines.

Figure 4.5 shows the average luminance values produced by observers for the physical match condition and the four LCD screen match conditions, for the three possible reference patch luminances. The dotted black lines show the reference luminance values. It is only possible to adjust the luminance of an LCD screen and not reflectance. However, it is difficult to measure lightness constancy by comparing the reference and match luminance values; some measure of reflectance is needed to visualize lightness matching across different lighting. Therefore, as the patch of LCD screen is viewed under different but constant illumination, *equivalent reflectance* values can be calculated. A similar formula to equation 2.1 is used to calculate the equivalent reflectance from the reference and match luminance (equation 4.1).

$$\text{reflectance} = 0.99 \times \frac{\text{luminance of the reference or match patch}}{\text{luminance of the reflectance standard}} \qquad (4.1)$$

Figure 4.6 shows the average equivalent reflectance values of the observer matches for all three reference patch reflectances. The dotted black lines show the reference reflectances. The first bar from the left in each set shows the average reflectance of the match patch when physical paper is used, across all six observers. Across all three gray shades, observers pick a higher reflectance or a lighter patch than the reference patch; this is to be expected due to imperfect lightness constancy as discussed in Chapter 2. The match patch is under a darker illumination than the reference patch, so observers would pick a lighter patch to compensate for the dark illumination. This first bar in Figure 4.6 is the physical match condition, which provides a baseline measure of lightness constancy with a real paper match. All the LCD screen manipulations are compared to this baseline. The remaining four bars in each set are the average match patch reflectance values across all six observers for the four LCD screen conditions.

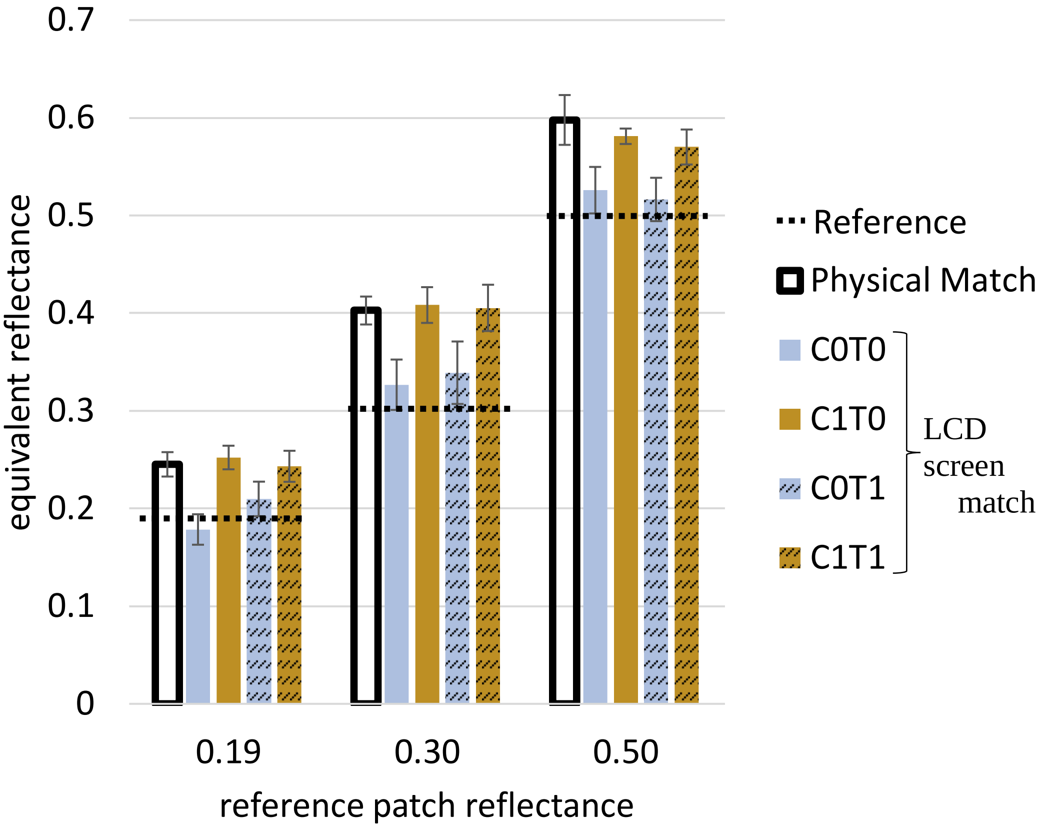


Figure 4.6 Average match patch reflectance values for six observers for three reference patch reflectance values 0.19, 0.30 and 0.50 (dark to light), across five conditions: one physical paper condition and four LCD screen manipulations (C0T0, C1T0, C0T1, and C1T1). The reference patch equivalent reflectance is indicated by the dotted lines.

To quantify the degree of lightness constancy in each condition, I introduce the *proportionate lightness constancy* as a means to compare lightness constancy across the four LCD screen conditions, as compared to the baseline physical condition. Let the luminance of the reference patch from Figure 4.5 be L0 and let the luminance of the physical match patch be L1. When the luminance of the LCD match condition is L, the proportionate lightness constancy is defined as:

$$\text{proportionate lightness constancy} = \frac{\log(L)-\log(L_0)}{\log(L_1)-\log(L_0)} \qquad (4.2)$$

The proportionate lightness constancy values are normalized such that a match whose luminance is identical to that of the reference patch would have a proportionate lightness constancy of 0 and the match in the physical condition would have a proportionate lightness constancy of 1. If observers are luminance matching, then $L = L0$ and resulting ratio will be equal to 0. On the other hand, if the luminance match performance on the LCD screen is the same as matching with paper, then $L = L1$, which will result in the ratio being 1. The proportionate lightness constancy is similar to the Thouless ratio (Thouless 1931), which is a common measure of lightness constancy as described in Chapter 2. The distinction between the two is that in cases where the physical match is not a perfect reflectance match (which would yield a Thouless ratio of 1), the proportionate lightness constancy allows for comparisons between various virtual conditions and a baseline physical condition.

Figure 4.7 shows the proportionate lightness constancy values for all three gray levels, computed within observers, then averaged across six observers for all four LCD screen manipulations (C0T0, C1T0, C0T1, and C1T1). A 2×2×3 mixed model ANOVA was conducted to test the influence of colour (2 levels: C0 and C1), texture (2 levels: T0 and T1), and gray level (3 levels: light, medium, and dark) on the proportionate lightness constancy values. The analysis revealed a significant main effect of colour ($F(1, 72) = 33.42$, $p <.001$). When the LCD screen is colour-matched to the room chromaticity, proportionate lightness constancy values are significantly lower than when the LCD screen is not colour matched to the room chromaticity. On the other hand, no significant differences in mean proportionate lightness constancy values were found for either the three gray levels or for whether texture was present or absent. There were no significant 2- or 3-way interactions between the factors.

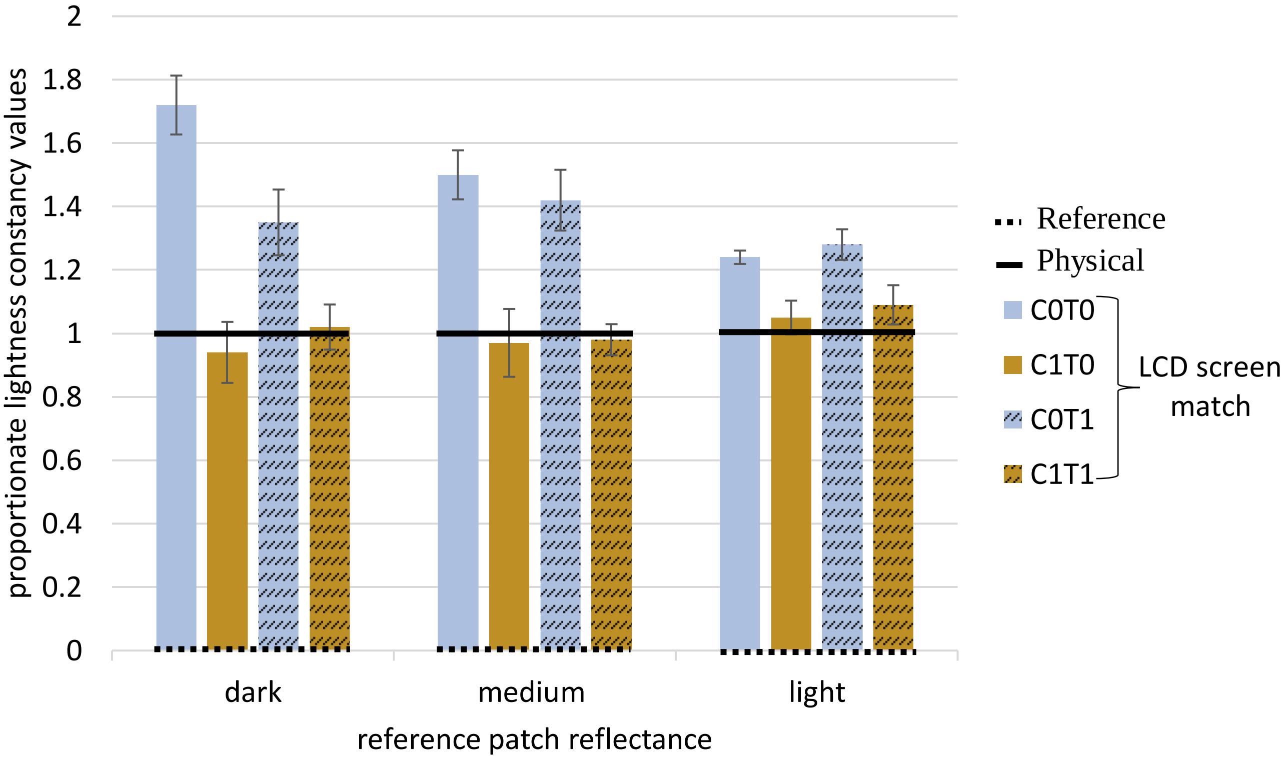


Figure 4.7 Proportionate lightness constancy values for three gray levels, averaged across six observers for all four LCD screen manipulations (C0T0, C1T0, C0T1, and C1T1). The error bars are the standard error of the mean. The proportionate lightness constancy values are normalized such that the reference patch would have a proportionate lightness constancy of 0 and the match patch from the physical condition would have a proportionate lightness constancy of 1.

One-sample t-tests with Bonferroni corrections were used to individually compare proportionate lightness constancy values for the non-colour matched (C0) and colour matched (C1) conditions (pooled across different reference patch reflectance levels and textures) to the proportionate lightness constancy of physical paper match (proportionate lightness constancy of 1). This allows us to determine if the mean proportionate lightness constancy values for the C0 or C1 conditions is significantly different from a proportionate lightness constancy of 1, which indicates the luminance match is the same as would be observed with physical paper. The proportionate lightness constancy values for the colour non-matched conditions C0 ($M = 1.29$, $SD = 0.28$) is significantly different from the proportionate lightness constancy of a physical paper match ($t(35) = 5.44$, $p<0.001$, 95% confidence CI [1.18, 1.39]). However, the proportionate lightness constancy values for the colour-matched conditions C1 ($M = 1.01$, $SD = 0.18$) is not significantly different from the proportionate lightness constancy of a physical paper match ($t(35) = 0.41$, $p>0.05$, 95% confidence CI [0.95, 1.07]). Therefore, in our asymmetric lightness matching task, the pooled proportionate lightness constancy values in the non-colour-matched conditions C0 are significantly different from the proportionate lightness constancy when observers perform the task with physical paper. Conversely, in the C1 conditions, the proportionate lightness constancy values were not significantly different. Thus, based on the ANOVA and the t-tests, it is evident that in a lightness matching task in which the chromaticity of the LCD screen is not matched, proportionate lightness constancy values are significantly different from those obtained with physical matches. Additionally, for the colour not matched conditions, the proportionate lightness constancy is significantly different from 1, which means that observers are not lightness matching as they would with physical paper surfaces. The results are consistent with constancy but it cannot be decisively concluded that the proportionate lightness constancy for the colour-matched conditions is the same as paper, as this is a null result.

4.4 GENERAL DISCUSSION

The reflectance of the physical match patch is not the same as the reflectance of the reference patch (Figure 4.6) because observers exhibit partial lightness constancy. This is in line with the partial lightness constancy shown in Chapter 2 and in almost all previous studies (e.g., Kraft & Brainard, 1999). In the present experiment, the match patch was under a lower illumination, and therefore observers picked a slightly higher match patch reflectance than the reference patch, so it appears the same lightness as the reference patch. When comparing the LCD screen manipulations, colour (whether or not the screen white point was matched to the room lighting) was only the factor which affected lightness constancy relative to a physical paper match.

In Figure 4.7, it is clear the colour-not-matched C0 conditions yield higher proportionate lightness constancy values than the colour-matched C1 conditions and a t-test shows that C0 also has significantly higher mean proportionate lightness constancy values compared to the physical match condition. Equation 4.2 shows that in order for proportionate lightness constancy values to be greater than 1, the difference between the luminance of the LCD screen match patch (L) and reference patch (L_0) must be higher than the difference between the luminance of the physical match patch (L_1) and reference patch (L_0). In C1, L is higher than L_1; therefore, observers are picking brighter matches on average as compared to the physical matches. It is possible that an isoluminant (same physical luminance) C0 patch appears dimmer to observers than a physical patch, when placed under the same illumination level. Thus, observers pick a higher luminance for C0 match patches to compensate for their dimmer appearance as compared to the physical match patch. Future studies need to be done to establish if there is a difference in the apparent brightness of isoluminant patches of different chromaticity to validate this claim. Simply comparing reflectance, with no baseline values can be misleading, therefore baseline measures (such as the physical match condition in this experiment) should be included when comparing lightness constancy to ensure that a brightness artifact is not erroneously interpreted as better lightness constancy, as may be the case for C0 in this experiment.

At first glance, from Figure 4.6 one may conclude that the conditions where the colour is not matched (C0) observers have better lightness constancy than conditions where the chromaticity is matched (C1), as observers seem to be choosing a reflectance which is more closely matched to the level of the reference patch in the (C0) conditions. However, it is important to keep in mind that this is a lightness matching task, so the reference patch and the matches are placed under different lighting. As seen in chapter 2, and in the physical match condition of this experiment, when observers make lightness matches with physical stimuli they do not perfectly match reflectance; i.e. they are only partially lightness constant. Thus, the colour-matched (C1) conditions are better in the sense that they produce lightness matching performance more similar to that observed with real paper.

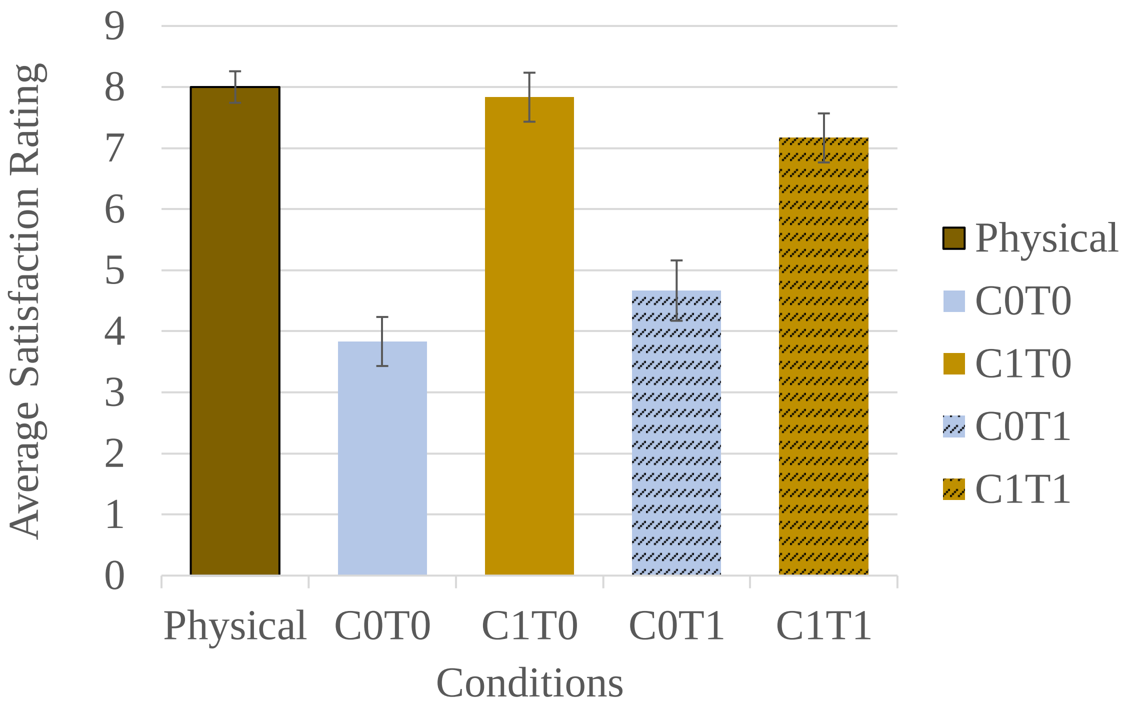


Figure 4.8 Average satisfaction ratings across observers for all five conditions. The error bars show standard error of the mean. At the end of each session observers were asked to rate how satisfied they were with their matches throughout that particular condition. They were given a chance to revise their ratings at the end of the experiment after they had seen all five conditions.

Figure 4.8 shows the average satisfaction rating across the six observers for each of the five conditions. The physical match and C1 conditions seem to have higher satisfaction ratings than the C0 conditions; although additional significance tests would need to be done to confirm this. As noted in chapter 2, in a lightness matching task where the goal is to match the targeted dimension (lightness), the two stimuli may still differ on other perceptual dimensions, such as gloss, translucency, or surface brightness, which may result in an overall imperfect match patch appearance. Logvinenko and Maloney (2006) commented on this to explain why observers may be dissatisfied with their match choices.

Perhaps the chromaticity difference in this experiment is perceived as a difference in brightness, which produces low satisfaction ratings for the colour-not-matched conditions. It may come as a surprise that observers rate the colour-not-matched C0 condition lower, even though the C0 condition showed better lightness constancy (matches closer to the true reflectance value).

In Chapter 3, it was shown that a good texture match was important to make an LCD screen patch look similar to a piece of paper. However, adding texture cues in this experiment did not produce a significant difference in the proportionate lightness constancy values. Texture may have been less important in this experiment because the reference patch was made of matte paper, so no texture was necessary to generate convincing lightness matches.

All in all, it is evident in this chapter that the factor colour, as in the chromaticity of the LCD screen is matched to the white point of the surrounding environment, plays a crucial role for obtaining realistic levels of lightness constancy using an LCD screen. The colour-corrected LCD screens can be used as a part of a more extensive physical setup to test theories of lightness constancy using lightness matching tasks such as a free adjustment task.

Chapter 5

CONCLUSION

This thesis investigated lightness constancy with physical and simulated surfaces to address three major goals. The studies in Chapter 2 contribute to the lightness literature both as an empirical test of a theoretical hypothesis and a comparison of effects in real and computer-generated stimuli. Chapter 3 examined the importance of colour adjustment and texture addition as cues to improvise the realism of computer-generated stimuli. Finally, in Chapter 4 an LCD screen was used as a part of a larger physical setup and it was shown that by matching the colour of the LCD screen to the room white point, it is possible to obtain lightness constancy levels similar to those observed with paper surfaces.

Summary of major findings

Achromatic surfaces have many perceptual dimensions, namely, lightness, gloss, translucency, surface brightness. There is much debate about whether the human visual system has access to these different dimensions in all situations, or whether this access varies under different conditions or based on the instructions provided on a task (Arend & Spehar, 1993). Logvinenko and Maloney's (2006) hypothesis was based on a multidimensional perceptual space with two dimensions roughly corresponding to reflectance and illuminance. Logvinenko and Maloney's theory proposed that observers have no perceptual access to either dimensions. In Chapter 2, we provide evidence that observers are fairly good at estimating reflectance when given complete freedom over both the dimensions of reflectance and illuminance (more evidence for this is also provided in Chapter 4). However, more work is needed to be done to better understand the perceptual difference between other dimensions such as brightness and lightness, which are easily perceptually confused. The second main finding in Chapter 2 was that observers had significantly better lightness constancy when performing the task in computer-generated scenes as compared to real scenes. This finding was surprising because there is prior evidence showing that lightness constancy in computer-generated scenes is either similar to the real world or worse (e.g., Morgenstern, Murray, & Harris, 2011; Radonjić et al., 2016). We postulate that our result is due to a combination of weak lightness constancy and glow-related artifacts that mimic typical – or in the case of some observers, better – lightness constancy for computer-displayed surfaces as compared with real paper.

The experiments in Chapter 3 demonstrate that, with colour adjustment and the addition of texture cues, a patch of LCD screen can appear realistic. Furthermore, although it is important that the chromaticity of the LCD screen is matched to the room white point, the manner in which this matching is accomplished is not important. Namely, either matching the chromaticity of the LCD screen to the room white point or vice versa results in realistic computer-generated stimuli. We also found that lighting direction plays a more important role than initially expected in a discrimination task because the virtual stimuli were compared to real pieces of paper. However, in a lightness constancy task as in Chapter 4, where an LCD screen is used in place of a paper surface and the goal is to achieve the same lightness constancy levels, factors such as lighting direction may be less important. More research is needed to understand the effect of lighting on a discrimination task, where a wider variety of lighting directions, as well as the lighting being ambient versus spotlight are manipulated as discussed in Chapter 3 and in lightness matching tasks as discussed in Chapters 2 and 4.

In Chapter 4, an LCD screen was used to measure lightness constancy in a highly-articulated physical scene. Extending the knowledge gained in Chapter 3, colour and texture cues were adjusted to test if lightness matching results using an LCD screen are equivalent to those obtained with paper stimuli. It is evident that colour plays a crucial role in achieving realistic levels of lightness constancy using an LCD screen. The colour-corrected LCD screens can be used as a part of a more extensive physical setup to test theories of lightness constancy using tasks such as a free adjustment task. A novel analysis to compare lightness constancy performance between physical and simulated surfaces is also provided. These results demonstrate that LCD screens can be used in lightness matching experiments so long as the chromaticity of the LCD screen is matched to the white point of the surrounding environment. Although, as shown in Chapter 2, lightness constancy levels in a fully virtual scene are not equivalent to the lightness constancy levels observed using a physical apparatus, LCD screens can be incorporated into a larger physical apparatus to achieve realistic levels of lightness constancy.

Chapter 6

REFERENCES

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science, 262*, 2042-2044.

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 339-351). Cambridge, MA: MIT Press.

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

Arend, L. E., & Spehar, B. (1993a). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics*, *54*(4), 446-456.

Arend, L. E., & Spehar, B. (1993b). Lightness, brightness, and brightness contrast: 2. Reflectance variation. *Perception & Psychophysics*, *54*(4), 457-468.

Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & psychophysics*, *35*(5), 407-422.

Blakeslee, B., & McCourt, M. E. (1999). A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. *Vision Research, 39*(26), 4361-4377.

Blakeslee, B., & McCourt, M. E. (2003). A multiscale spatial filtering account of brightness perception. In L. R. Harris & Jenkin (Eds.), L. Springer-Verlag.

Blakeslee, B., & McCourt, M. E. (2008). Nearly instantaneous brightness induction. Journal of Vision, 8(2), 1–8.

Blakeslee, B., & McCourt, M. E. (2012). When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision Research, 60,* 40-50.

Blakeslee, B., Reetz, D., & McCourt, M. E. (2009). Spatial filtering versus anchoring accounts of brightness/lightness perception in staircase and simultaneous brightness/lightness contrast stimuli. *Journal of Vision*, *9*(3), 22-22.

Bloj, M., Ripamonti, C., Mitha, K., Hauck, R., Greenwald, S., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision, 4*(9), 735–746.

Bonato, F., & Gilchrist, A. L. (1999). Perceived area and the luminosity threshold. *Attention, Perception, & Psychophysics, 61*(5), 786-797.

Brainard, D. H., & Ishigami, K. (1995, January). Factors influencing the appearance of CRT colors. In *Color and Imaging Conference* (Vol. 1995, No. 1, pp. 62-66). Society for Imaging Science and Technology.

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. Hornak (Ed.), *Encyclopedia of Imaging Science and Technology.* Hoboken, NJ: John Wiley and Sons, Inc.

Bressan, P. (2006). The place of white in a world of grays: A double-anchoring theory of lightness perception. *Psychological Review, 113*, 526–553.

Bressan, P., & Actis-Grosso, R. (2001). Simultaneous lightness contrast with double increments. *Perception, 30*, 889–897.

Cornsweet, T. N. (1970). *Visual perception.* New York: Academic Press.

Economou, E., Zdravkovic, S., & Gilchrist, A. (2007). Anchoring versus spatial filtering accounts of simultaneous lightness contrast. *Journal of Vision*, *7*(12), 2-2.

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

Gilchrist, A. L. (2006). *Seeing black and white.* New York, NY: Oxford University Press.

Gilchrist, A. L., & Annan Jr, V. (2002). Articulation effects in lightness: Historical background and theoretical implications. *Perception*, *31*(2), 141-150.

Gilchrist, A., et al. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*(4), 795-834.

Horn, B. K. (1974). Determining lightness from an image. *Computer graphics and image processing*, *3*(4), 277-299.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing-and test-field luminances. *Journal of experimental psychology*, *50*(2), 89.

Helmholtz, H. V. (1866). *Concerning the perceptions in general. Treatise on physiological optics, 3.*

Hering, E. (1874/1964). Outlines of a theory of the light sense. (L.M.H.D. Jameson, Trans.). Cambridge, MA: Harvard University Press.

Jacobsen, A., & Gilchrist, A. (1988). The ratio principle holds over a million-to-one range of illumination. Perception & Psychophysics, 43, 1–6.

Kingdom, F. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision research*, *51*(7), 652-673.

Kosslyn, S. M., & Andersen, R. A. (Eds.). (1995). *Frontiers in cognitive neuroscience*. MIT Press.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences, 96*(1), 307-

312.

Krawczyk, G., Myszkowski, K., & Seidel, H.P. (2005). Lightness perception in tone reproduction for high dynamic range images. *Computer Graphics Forum, 24*, 635-645.

Kim, M. (2016). Shape and luminance cues for the visual perception of glow (Unpublished doctoral dissertation). New York University, New York.

Land, E. H. (1977). The retinex theory of color vision. *Scientific American*, *237*(6), 108-129.

Land, E. H. (1983). Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(16), 5163.

Land, E.H. & McCann, J. J. (1971) Lightness and retinex theory. *Journal of the Optical Society of America, 61*. 1-11

Logvinenko, A. D. (2015). The achromatic object-colour manifold is three-dimensional. *Perception, 44*(3), 243-268.

Logvinenko, A. D., & Maloney, L. T. (2006). The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics, 68*(1), 76-83.

Logvinenko, A. D., Petrini, K., & Maloney, L. T. (2008). A scaling analysis of the snake lightness illusion. *Perception and Psychophysics, 70*(5), 828-840.

McNamara, A., Chalmers, A., Troscianko, T., & Gilchrist, I. (2000). Comparing real and synthetic scenes using human judgements of lightness. In B. Péroche & H. Rushmeier (Eds.), *Rendering Techniques 2000* (pp. 207-218). Vienna: Springer.

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, *13*(1), 59-66.

Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2014). Human vision is attuned to the diffuseness of natural light. *Journal of Vision, 14*(9):15, 1–18.

Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences of the U.S.A., 108*(30), 12551-12553.

Murray, R. F. (2013). Human lightness perception is guided by simple assumptions about reflectance and lighting. In B. E. Rogowitz, T. N. Pappas, & H. de Ridder (Eds.), *Proceedings of SPIE 8651, Human Vision and Electronic Imaging XVIII.* San Francisco, CA, February 4-7, 2013.

Murray, R. (2018). A computational mid-level model of lightness perception. Poster presented at *Vision Sciences Society* 2018, Florida.

Nayar, S. K., Belhumeur, P. N., & Boult, T. E. (2004). Lighting sensitive display. *ACM Transactions on Graphics (TOG)*, *23*(4), 963-979.

Patel, K., Palatnic, L., & Murray, R. (2017, May). Why do LCD screens appear to glow? (conference abstract from Vision Sciences Society, Florida). *Journal of Vision*, *17*(10), 662-662.

Patel, K., Palatnic, L., & Murray, R. (2017, August). How to make LCD screens look like paper? Poster presented at the *European Conference on Visual Perception* 2017, Berlin, Germany.

Patel, K. Y., Munasinghe, A. P., & Murray, R. F. (2018). Lightness matching and perceptual similarity. *Journal of vision, 18*(5), 1-1.

Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision research*, *90*, 10-14.

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., Brainard, D. H., & Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision, 16*(11):2, 1-18.

Rudd, M. E. (2010). How attention and contrast gain control interact to regulate lightness contrast and assimilation: a computational neural model. *Journal of vision*, *10*(14), 40-40.

Rudd, M. E., & Zemach, I. K. (2005). The highest luminance rule in achromatic color perception: Some counterexamples and an alternative theory. *Journal of Vision, 5*, 983–1003.

Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: a direct test of the illumination estimation hypothesis. *Psychological Science, 13*(2), 142-149.

Schanda, J. (2007). *Colorimetry: understanding the CIE system.* Hoboken, NJ: John Wiley & Sons, Inc.

Shapiro, A., & Lu, Z. L. (2011). Relative brightness in natural images can be accounted for by removing blurry content. *Psychological science*, *22*(11), 1452-1459.

Thouless, R. H. (1931). Phenomenal regression to the real object. I. *British Journal of Psychology, 21,* 339-359.

Tse, P. U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research*. *45*(9), 1095-1098.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of experimental psychology*, *38*(3), 310.

Wallach, H. (1976). On Perception. New York: Quadrangle/The New York Times Book Co.

White, M. (1979). A new effect of pattern on perceived lightness. *Perception*, *8*(4), 413-416.

Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. *Computer Graphics* (Proceedings of SIGGRAPH 1994), 459-472.