Dimensionality of the Perceptual Space of Achromatic Surface Colors

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

Vorgelegt von

Nora Umbach

aus Hannover

Tübingen

2013

Tag der mündlichen Qualifikation:

Dekan:	Prof. Dr. Rosenstiel
1. Berichterstatter:	Prof. Dr. Jürgen Heller
2. Berichterstatter:	Prof. Dr. Felix Wichmann

Contents

Abstract

The perception of lightness is a much discussed topic. Lightness perception is concerned with how we perceive achromatic colors. Achromatic colors include black and white and all shades of gray. Most theories in color perception try to describe phenomenological experience by perceptual spaces. In a perceptual space, differences between colors can be described quantitatively.

Up until now, a perceptual space for achromatic colors has not been investigated systematically. Recent evidence suggests that a perceptual space for achromatic colors needs to be at least two-dimensional, even though traditional views consider the perception of achromatic colors as one-dimensional (with the end points white and black). It has been shown that this traditional view cannot account for all perceptual phenomena experienced when looking at achromatic surface colors (Logvinenko & Maloney, 2006; Niederée, 2010). With stimulus situations that are more complex than a single light presented in a dark room we get qualitative differences between stimuli that cannot be explained by a one-dimensional perceptual space for achromatic colors. These qualitative differences are, e. g., associated with the color of the stimulus (surface color) and how much light falls on the stimulus (illumination color). A formal perceptual space would allow us to get a clearer picture of what people actually perceive in more complex stimulus situations.

The focus in this thesis will be on achromatic surface colors. In everyday situations, we usually consider color to be a property of an object unaware that we construct color in our head. In order to investigate surface colors we need experimental settings that go beyond the often used dark room conditions. We can only perceive surface colors in an illuminated room or when several surfaces are present. Therefore, all experiments were conducted under constant illumination conditions. Subjects performed same-different judgments on gray patches with or without surrounds presented in an illuminated room.

In the first experiment, subjects judged simple gray patches without any context. Results show that we need a single perceptual dimension to discriminate between these simple stimuli. In the second and third experiment, local context effects in the form of uniform surrounds were introduced. Results show that two perceptual dimensions are needed to discriminate between stimuli embedded in different surrounds.

The results emphasize how important it is to distinguish between different stimulus situations and investigate achromatic colors under controlled illumination conditions. The relationship between stimulus configuration and the dimensionality of a perceptual space for achromatic surface colors needs to get much more attention in future experiments.

Zusammenfassung

Helligkeitswahrnehmung ist ein viel diskutiertes Thema. Sie beschäftigt sich damit, wie wir achromatische Farben wahrnehmen. Achromatische Farben beinhalten schwarz und weiß und alle Graustufen. Die meisten Farbwahrnehmungstheorien versuchen phänomenologisches Empfinden über Wahrnehmungsräume zu beschreiben. Unterschiede zwischen Farben können mit Wahrnehmungsräumen quantitativ beschrieben werden.

Bisher wurde ein Wahrnehmungsraum für achromatische Farben noch nicht systematisch untersucht. Bisherige Befunde zeigen, dass ein Wahrnehmungsraum für achromatische Farben mindestens zweidimensional sein muss, obwohl die traditionelle Sichtweise diesen als eindimensional ansieht (mit den beiden Endpunkten Schwarz und Weiß). Es wurde gezeigt, dass diese traditionelle Sichtweise nicht allen perzeptuellen Phänomenen Rechnung tragen kann, wenn wir achromatische Oberflächenfarben betrachten (Logvinenko & Maloney, 2006; Niederée, 2010). Stimulussituationen, die in ihrer Komplexität über einfache Lichtreize, die in einem dunklen Raum dargeboten werden, hinausgehen, können nicht mit einem eindimensionalen Wahrnehmungsraum für achromatische Farben beschrieben werden. Diese qualitativen Unterschiede hängen z. B. mit der Farbe des Stimulus (Oberflächenfarbe) oder der Menge an Licht, das diese Oberfläche trifft (Beleuchtungsfarbe), zusammen. Ein formal definierter Wahrnehmungsraum würde es uns erlauben uns ein klareres Bild davon zu machen, was Menschen tatsächlich wahrnehmen, wenn sie mit komplexen Stimulussituationen konfrontiert sind.

Die vorliegende Dissertation konzentriert sich auf achromatische

Oberflächenfarben. In unserem Alltag betrachten wir Farbe normalerweise als eine Eigenschaft von Objekten, ohne uns bewusst zu werden, dass Farbe im Kopf entsteht. Um Oberflächenfarben untersuchen zu können, brauchen wir experimentelle Gegebenheiten, die über die meist verwendeten Dunkelraumbedingungen hinausgehen. Wir können Oberflächenfarben nur in beleuchteten Räumen oder neben anderen farbigen Flächen wahrnehmen. Deswegen wurden alle Experimente unter konstanten Beleuchtungsbedingungen durchgeführt. Die Versuchspersonen gaben Gleich-Ungleich-Urteile über graue Flächen ab, die mit oder ohne Umfeld in einem erleuchteten Raum präsentiert wurden.

Im ersten Experiment beurteilten die Versuchspersonen einfache graue Felder ohne Kontext. Die Ergebnisse zeigen, dass wir eine einzige perzeptuelle Dimension benötigen, um zwischen diesen einfachen Stimuli zu unterscheiden. Im zweiten und dritten Experiment wurden lokale Kontexteffekte als einfache Umfelder eingeführt. Die Ergebnisse zeigen, dass wir zwei perzeptuelle Dimensionen benötigen, um zwischen Stimuli zu unterscheiden, die mit unterschiedlichen Umfeldern präsentiert werden.

Die Ergebnisse heben hervor, wie wichtig es ist zwischen unterschiedlichen Stimulussituationen zu unterscheiden und achromatische Farben unter kontrollierten Beleuchtungsbedingungen zu untersuchen. Die Beziehung zwischen Stimuluskonfiguration und der Dimensionalität eines Wahrnehmungsraums für achromatische Oberflächen muss in zukünftigen Untersuchungen noch stärker beachtet werden.

Chapter 1

Introduction

The perception of color has been studied extensively for hundreds of years (see, e. g., Purves & Lotto, 2011; Wyszecki & Stiles, 2000, for overviews). However, there is still an ongoing discussion what really constitutes color and how it should be dealt with on a theoretical level.[1] When talking about colors in everyday life, we are certain that objects have a 'true' color and that this color does not change with different illumination or different context. We use color to define objects (Should I wear my blue or my black dress?), never aware that color is something that we construct and not a property of objects. Visual perception comes so naturally to us, that we never doubt what we see is the 'real' world; completely unaware that we construct this world from very limited information (Hoffmann, 1998). This construction of the world and its colors within cannot be influenced consciously.

All theories trying to explain the perception of all perceivable colors make assumptions about the perception of achromatic (gray) colors (see Volbrecht & Kliegl, 1998, for a review that focuses on the perception of blackness). Gilchrist (2006) defines achromatic surface colors as: "In the case of surfaces, *achromatic* refers to colors along the scale of grays from black to white" (p. 375). This defines the traditional view of an achromatic color space: All shades of gray

[1] "Among the many different attributes of visual experiences the attribute of color appears to be the most enigmatic with respect to our attempts to deal with it theoretically" (Mausfeld, 1998, p. 219).

lie between two end points (black and white) on one dimension.

Achromatic color perception is usually referred to as lightness (or brightness) perception. When studying lightness separately, researchers usually focus on black-and-white stimuli. There is a vast amount of research literature on this topic (see, e. g., Gilchrist, 2006; Volbrecht & Kliegl, 1998, for overviews) using all kinds of simple and elaborate black-and-white stimulus configurations. However, this research produced a plethora of contradicting results (Gilchrist, 2006).

The goal of this thesis is to try to understand how simple achromatic stimuli are perceived. Traditional psychophysical methods building on early theoretical considerations will be used in an attempt to systematically investigate what a perceptual space for achromatic surface colors looks like. Staying within one paradigm using clearly defined stimulus situations and building on theoretical assumptions might help to understand the underlying psychological processes of lightness perception. In the light of the many different approaches and results found in lightness perception, it seems important to build on a solid theoretical framework and start small with stimuli increasing in complexity over several experiments in order to understand and integrate different concepts (a procedure suggested by Mausfeld, 1998).

When defining color in physical terms, we consider electromagnetic radiation that is visible to the human eye. The wavelengths of visible light range from about 380 to 780 nm. When defining light one has to distinguish between radiometric photometry and photometric photometry. Radiometry measures radiant energy emitted from or transferred through a surface. Radiance is the amount of radiant energy reflected by a surface. Photometry measures light in terms of how bright a surface appears to an observer. The photometric equivalent to radiance is luminance. Hereby, the radiant power is weighted by a standard luminous efficiency function which is often called the Standard Observer function. This function was obtained by having subjects judge lights of different radiant energy according to their brightness and is closely related to the spectral sensitivity of our receptors. Most (spectro-)photometers calculate

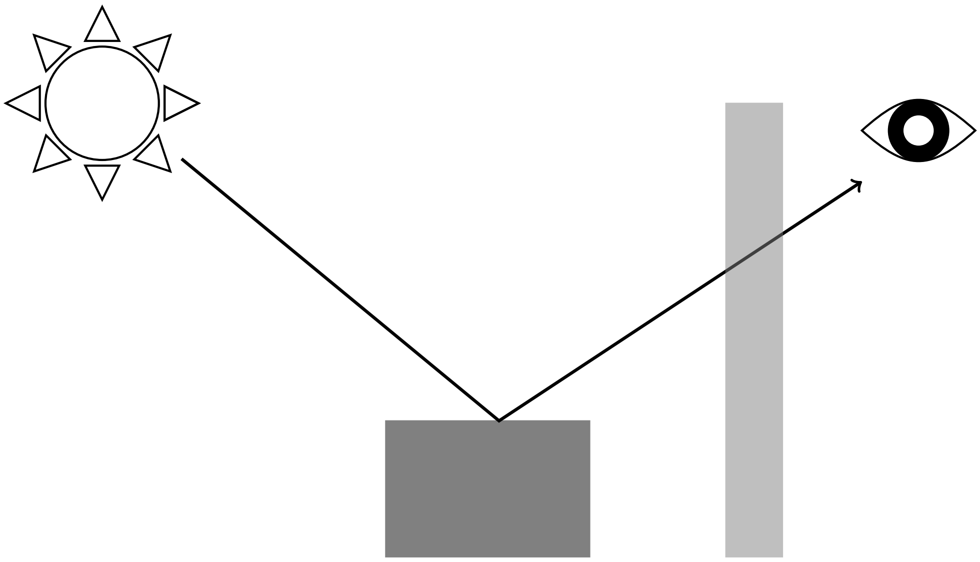


Figure 1.1: Physical properties influencing the perception of colors: Illumination, reflectance, and transmittance.

luminance by first measuring radiance and then integrating over the spectrum of visible light ($\lambda = 380 \ldots 780\,\mathrm{nm}$) to obtain luminance

$$L_v = K_m \int_\lambda L_{e,\lambda} V(\lambda)\, d\lambda \tag{1.1}$$

(Wyszecki & Stiles, 2000, p. 259); where $L_{e,\lambda}$ is radiance, $V(\lambda)$ is the Standard Observer function for photopic vision and K_m is a constant which is set to $K_m = 683\,\frac{\mathrm{lm}}{\mathrm{W}}$ for practical applications (see Wyszecki & Stiles, 2000, p. 258 for more details). Luminance is therefore a psychophysical property of light and achromatic stimuli can be characterized exclusively by luminance in physical terms. Equation 1.1 shows that the physical property radiance determines luminance exclusively. It is common in psychophysics to define stimuli by their luminance (usually measured in candela per square meter, $\frac{\mathrm{cd}}{\mathrm{m}^2}$).

Most of the time, we are not aware of the complex interplay of different factors when we perceive colors or lightness. Figure 1.1 shows the minimal circumstances our visual system has to take into account when perceiving the color of a surface: First, we are constantly confronted with illumination changes. These can be successive changes over the day or when we switch on the light in a room, or simultaneous illumination changes like shadow borders.

Secondly, the reflectance properties of the surface play a role. A surface that reflects more light will look lighter to us. Thirdly, the stimulus might be seen through some kind of medium (haze, fog, dirty window, etc.). Thus, the light that actually reaches the eye is influenced by many factors; light may have the same physical properties, but was generated by different physical circumstances (e. g., less light falling on a lighter surface vs. more light falling on a darker surface). Despite of these ambiguities the visual system excels at disentangling these different circumstances and reliably generates an interpretation of the scene that helps us understand what we see. Most theories on vision (and especially theories on color vision) try to explain how the visual system achieves this outstanding performance.

Achromatic color space is often understood to be one dimension (namely brightness) of a three-dimensional chromatic color space with the other two dimensions being hue and saturation (see, e. g., Evans, 1964; Izmailov & Sokolov, 1991; Wallach, 1963). When talking about lightness or brightness perception, it is important to define certain concepts since literature on lightness and brightness perception uses many concepts and expressions in different ways. Brightness and lightness are both psychological concepts that describe what we perceive. *Lightness* refers to the perceived amount of light the surface of an object apparently reflects. In the context of achromatic colors, lightness can be considered as the color of a gray patch. *Brightness* is the perceived luminance of an object. It depends on the lightness (color) of an object and the amount of light incident on this object. Luminance, illumination, and reflectance are understood to be (psycho-)physical properties of the stimuli presented.

Furthermore, in order to understand the many aspects playing a role in color perception, it is important to distinguish between different encoding types. Color perception can be roughly divided into an early (physiological) encoding of stimuli (often referred to as sensation and associated with trichromacy as postulated by the Young-Helmholtz theory introduced in Section 2.3) and a later encoding (perception) that can take context effects and environmental variables like illumination into account. The early encoding is associated with light stimulating the different receptor types and the later encoding is supposed to be located in cortical structures (Webster, 2003).

Mausfeld (1998) argues that the visual system distinguishes between two modes of color perception: illumination and object colors (also called light and surface colors). This distinction was made by several authors (e. g., Evans, 1964; Heggelund, 1992; Niederée, 1998). According to Mausfeld, both color modes create separate color codes. One color code represents the color of a surface and one is associated with the illumination of the room. These two color codes cannot be traded-off. Under normal conditions, the color appearance of a surface cannot be changed by changing the illumination of the room. Evans (1964) explicitly states that "object-color perception [. . .] can occur only when there is more than one object present, using the word object generally, i. e., perception of the illumination as distinct from the objects *must be possible*" (p. 1468). This summarizes the achievement of the visual system of separating the color of an object and the illumination in the room. This achievement is a prerequisite to experiencing color constancy (see Section 2.4).

Surface (or object) colors are especially relevant when investigating lightness perception, since light colors can only differ in brightness. Wallach (1963) points out that "[l]ightness or darkness is a property of surfaces, and the investigator of neutral-color perception must concern himself with white or gray or black surfaces" (p. 278). To encourage that stimuli will be perceived as surface colors, they should be presented under controlled illumination conditions. This has been widely neglected in the literature up until now (Gilchrist, 2006).

Research on color perception can be divided into different approaches. Most theories combine psychophysical data and assumptions and physiological approaches. Mausfeld (1998) differentiates between psychophysical and physiological color codes and emphasizes that there is "a logical gap between quantitative psychophysical notions of color codes that refer only to psychological relations on the one hand, and the neurophysiological interpretations of these codes in terms of neural codes on the other" (p. 229). When considering psychophysical color codes, the focus lies on understanding psychological processes underlying the perception of color. A psychophysical approach focuses on the psychological processes created by physical stimuli without interpreting results in terms of neurophysiological processes. As put by Mausfeld (1998): "[...] psycho-physics deals with the interplay of phenomenology and physics and with the abstract 'strategies' that the visual system employs for achieving tasks, without embarking on speculations about neural mechanisms. To reveal the underlying strategies is first and foremost a psychological or psychophysical task, since only when we have an idea of the basic 'logic' of the system can we speculate on neurophysiological implementation [...]" (p. 231).

Chapter 2

Theoretical Background

The following chapter gives an introduction of relevant theoretical concepts. First, concepts that are relevant to understand the background on color and lightness perception are introduced: The distinction between a perceptual space and a stimulus space will be explained and infield-surround configurations and their meaning in color and lightness research will be introduced. Then, a short overview of color vision and lightness perception will follow. Some of the traditional psychophysical experiments conducted over the last decades will be introduced and two of the most investigated effects, lightness constancy and simultaneous lightness contrast, will receive particular attention. Then, theoretical work done in lightness perception and how it connects to a perceptual color space of achromatic colors will be presented. Thereafter, the influence of cognitive processes on the perception of lightness will be discussed and why presenting stimuli under controlled illumination conditions might be in order.

2.1 Perceptual Space vs. Stimulus Space

In psychophysics, stimuli are defined by physical properties. It is important to distinguish between a *stimulus space* obtained by these definitions on the one hand and a *perceptual space* on the other hand. When investigating the perception of loudness, for example, the sound source can be characterized by physical properties like sound pressure level (SPL), frequency, and duration. These are all definitions of the stimulus space. On the other hand, there are relevant properties of the perceptual representation. For this example, these might be loudness and pitch. A common mistake in psychophysics is to directly derive the perceptual or psychological characteristics from the physical ones (Mausfeld, 2002). Mausfeld (2002) uses the term "measurement device (mis-)conception of perception" to emphasize that the assumption that the goal of the perceptual system is to measure the 'true' (physical) situation is fundamentally flawed. A lot of misunderstanding results from not separating psychological and (psycho-)physical dimensions clearly (Evans, 1964; Mausfeld, 2002). Loudness, e.,g., can be influenced by changes in sound pressure level as well as changes in frequency (Moore, 2008), which illustrates that we do not have a one-on-one mapping of the physical characteristics into perceptual space. A stimulus space can have several physical dimensions by which it is defined. These dimensions can but do not have to be associated with perceptual dimensions. Moreover, the number of dimensions does not have to be identical. For example, we can have a two-dimensional stimulus space with a single perceptual dimension (see Dzhafarov & Colonius, 1999, and Section 5.2). Generally, a stimulus space will be continuous, but for the use in an experiment a discrete realization of this stimulus space will be used. In our experiments, we are limited to 1024 shades of gray by the monitor used.

A perceptual space is the mental representation of a stimulus space (or a set of stimuli). Gradual changes in stimulus space should result in gradual perceptual changes (Mausfeld, 1998). When talking about dimensionality of a perceptual space, the dimensions of this space refer to perceptual dimensions (as opposed to physical dimensions). Here, we are interested in the perceptual (or psychological) dimensions of color space.

Perceptual spaces for different observers do not have to be identical. They can differ quantitatively and qualitatively. In research on lightness perception aggregated data are often used (e. g., Izmailov & Sokolov, 1991; Logvinenko & Maloney, 2006; Wallach, 1963; Whittle & Challands, 1969). This has been

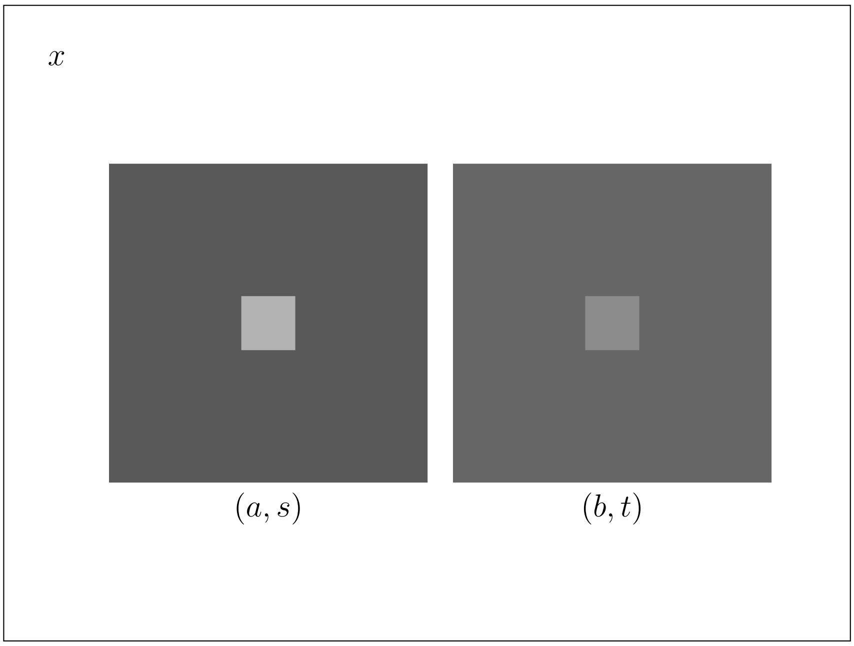


Figure 2.1: Two infield-surround configurations showing increments in an illuminated room. Infields and surrounds differ for both configurations, but background (illumination) x is the same. Stimuli are defined as (a, s, x) and (b, t, x) with $a, s, b, t < x$ and $a > s,\ b > t$.

criticized by several authors (e. g., Ekroll & Faul, 2009; Heller, 2001; Ronacher & Bautz, 1985). Perception is a process influenced by many different factors. Low-level as well as high-level mechanisms, like individual experience, influence our perception (Purves & Lotto, 2011). It is therefore not plausible to assume that perception is identical for different observers. Individual data analyses might provide us with a more distinguished picture about the underlying perceptual processes.

2.2 Infield-Surround Stimuli

The traditional way to investigate lightness and brightness perception are infield-surround stimuli (see, e. g., Heinemann, 1955; Hess & Pretori, 1970/1894; Wallach, 1948). In their simplest form, infield-surround stimuli consist of a small homogeneous color patch placed on a larger homogeneous color patch. These configurations are often called center-surround, aperture-surround, or target-surround stimuli and can have different shapes. Circles (e. g., Heinemann, 1955; Wallach, 1948) and squares (e. g., Hess & Pretori, 1970/1894), or a combination of both (e. g., Whittle & Challands, 1969) are most common. These configurations will be called infield-surround stimuli here, since the notion of a center or target, and especially an aperture, might be misleading in some contexts. Infield-surround stimulus refers to a uniform patch on a uniform surround and will be defined as (a, s, x). Variables a and s represent luminance of the infield and the surround, respectively, and variable x represents the luminance of the background.[1] In our experimental setting (see Chapter 4), the background contains the complete visual field of subjects.

Infield-surround configurations can be divided into two categories: *Increments* (see Figure 2.1) and *decrements* (see Figure 2.2). For increments, infields have a higher luminance than surrounds ($a > s$, for all stimuli) and for decrements vice versa ($a < s$, for all stimuli). Most experiments on lightness perception are conducted in dark rooms, i. e., x is generally smaller than a and s. Here, stimuli will be presented in an illuminated room and therefore $x > a$ and $x > s$, for all stimuli. Increments and decrements are supposedly processed and encoded differently (Gilchrist, 2006; Niederée, 1998; Whittle, 1986). There is a strong qualitative perceptual difference when looking at increments and decrements in a dark room. Increments will give the impression that the infield is a self-luminous object. Decrements, on the other hand, will look like a patch of gray paper (Gilchrist, 2006). When presenting infield-surround configurations in an illuminated room, they should always look like surface colors (like patches of gray paper), no matter if they are increments or decrements.

The distinction between object colors and illumination colors should be

[1] Traditional experimental settings, where stimuli are presented in a dark room, would be represented as $(a, s, 0)$ in our notation, which is usually reduced to (a, s). It should be mentioned that stimuli presented on a monitor will always have a background luminance, since it is impossible to present a background with a luminance of $0\,\frac{\mathrm{cd}}{\mathrm{m}^2}$ on a monitor. Nevertheless, this is often neglected when defining stimuli or interpreting results.

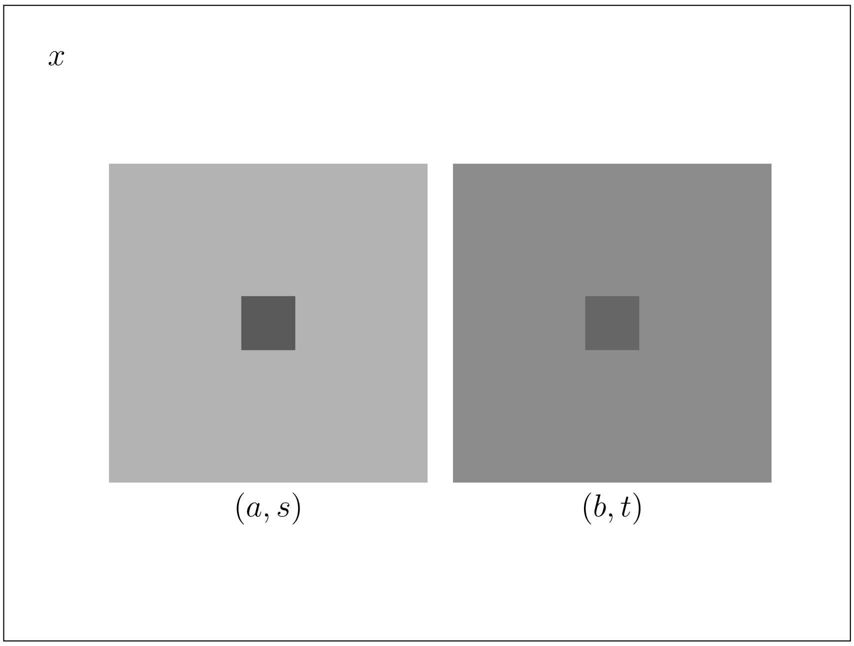


Figure 2.2: Two infield-surround configurations showing decrements in an illuminated room. Infields and surrounds differ for both configurations, but background (illumination) x is the same. Stimuli are defined as (a, s, x) and (b, t, x) with $a, s, b, t < x$ and $a < s,\ b < t$.

given some attention here in order to prevent misunderstandings. These two color codes are sometimes associated with incremental and decremental stimuli. When Mausfeld (1998) talks about different color codes, he does not refer to the difference between increments and decrements, but he rather refers to one color perception that is associated with the color of a surface and one color perception that depends on the (perceived) illumination of a room. Our visual system does not only distinguish these two modes of color perception, but perceives them at the same time attributing different aspects of a scene to different origins. This is what Mausfeld (1998) calls 'a dual code of color perception' and he regards "center-surround configurations as minimal stimuli for triggering a dual code for 'object' and 'illumination' colors" (p. 224). In an illuminated room, stimuli should appear like surface colors, regardless of whether one looks at increments or decrements. This does not imply that increments and decrements are necessarily processed in identical ways (Gilchrist, 2006).

For the concepts of background, surround, and illumination, there is again much variability how terms are used in the literature. We are talking about infield-surround configurations here. By surround, a larger field completely surrounding the infield is meant. Background denotes the whole visual field including the background of the monitor. In many papers, the term background describes the surround. Sometimes, it can even be found that the surround is called illumination (e. g., Whittle & Challands, 1969). When presenting infield-surround configurations in a dark room, it is hypothesized that the surround is interpreted as illumination and the infield judged as being presented under this illumination (see, e. g., Sawayama & Kimura, 2012; Soranzo & Agostini, 2006). Perceived differences between infields with identical luminance presented on surrounds with different luminance are attributed to this perceived difference in 'illumination.' For an experimental setting where stimuli are presented in a completely dark room, this assumption might hold and subjects might interpret this ambiguous stimulus situation accordingly. But other interpretations may be possible, and even plausible under these viewing conditions (Heller, 2001). Surround, in our experiments, always refers to a local context effect. Since only surface colors are considered, surrounds should not be perceived as illumination. The background x, on the other hand, will be perceived as illumination since it fills the whole visual field. Therefore, the terms *background* and *illumination* may be used interchangeably.

One should keep these distinctions in mind when interpreting results. Walraven (1976) coined the phrase *discounting the background* which has been widely used in the literature. It describes the effect that subjects match an infield to another infield with a different surround by only considering the ratio between infield and surround (Whittle & Challands, 1969, replicated these results for achromatic stimuli). It is called discounting the background since subjects do not take the brightness of the surround into account, but only how much the infield differs from the surround. When surround is interpreted in terms of illumination (as described above) this means that the lightness of the infield is deduced without taking into account that the infields are 'under different illuminations.' As mentioned before, this might hold when presenting stimuli in a dark room, but is certainly not true when presenting stimuli in an illuminated room. So talking about illumination in this context seems misleading. The experiments by Walraven (1976) and Whittle and Challands (1969) were conducted with Maxwellian-view optical systems where input for both eyes is independent. Hence, the results for these experiments may look very different under different conditions. Nevertheless, this framework is used to explain phenomena like simultaneous lightness contrast and color constancy (see Section 2.4), even though its applicability to these situations seems highly questionable.

The *ratio principle* (as first described by Wallach, 1948) makes the same predictions as the concept of discounting the background but forgoes the interpretation of the surround as illumination. The ratio principle states that the lightness of the infield solely depends on the luminance ratio between infield and surround; independent of the luminance of the surround. It has been argued that the receptors encode relative luminance and not luminance of the infield itself (Gilchrist, 2006). This view has been criticized on several grounds (see Gilchrist, 2006, Chapter 5), but the most common opinion is still that "[w]ithin the range of surface grays from white to black and within the range of normal illumination, the ratio principle appears to be the rule" (Gilchrist, 2006, p. 101). Jacobsen and Gilchrist (1988) claim that the ratio principle "holds over a million-to-one range of illumination" (p. 1). But considerations above show that the concepts are not always clearly distinguished and it might be in order to look at this claim in more detail and for different stimulus and illumination conditions.

2.3 Color Perception

The perception of color has a long and extensive research tradition (e. g. Wyszecki & Stiles, 2000). Nevertheless, much of the mechanisms that try to explain how color is perceived, are still not understood well. Color perception results from an interplay of multiple visual (physiological and psychological) processes as well as physical properties of objects and light. In the nineteenth century, several theories of color perception were introduced (see Mausfeld, 1998; Volbrecht & Kliegl, 1998; Wyszecki & Stiles, 2000, for overviews). The theories of Helmholtz and Hering were most influential.

The theory postulated by Helmholtz (1911) states (building on theoretical ideas formed by Young) that color vision is based on three different receptor types[2] and that color is not a physical property of objects, but something that we create during the process of perception. The theory is known as the Young-Helmholtz theory today. It was further formalized by Grassmann (see Mausfeld, 1998) and Krantz (1975), and is a psychophysical theory building on metamerism (see below). One of Helmholtz' assumptions was that 'unconscious inferences' like interpreting a stimulus situation in light of illumination conditions, three-dimensional shape, or similar are used to determine which color is seen when looking at objects.

Hering postulated a theory of opponent processes stating that color perception consists of three opponent channels: red–green, blue–yellow, and white–black. Both theories were supported by experimental data and Hurvich and Jameson (1957), among others, integrated both approaches in their opponent-process theory of color vision (Wyszecki & Stiles, 2000, call these theories 'zone theories'). This theory could account for a lot of experimental data (Volbrecht & Kliegl, 1998) and states that color sensations coming from the three types of receptors are combined at a second stage (or zone) to form the three channels mentioned above.

The signal coming from the receptors cannot be mapped to a unique physical stimulus. In theory, every signal coming from the receptors could be evoked by an indefinite number of physical stimulations (see Figure 1.1). This means that we perceive different physical stimuli as being of identical color. This phe-

[2]Helmholtz (1911, p. 119) does not call them receptors, yet, but nervous fibres—"Nervenfasern."

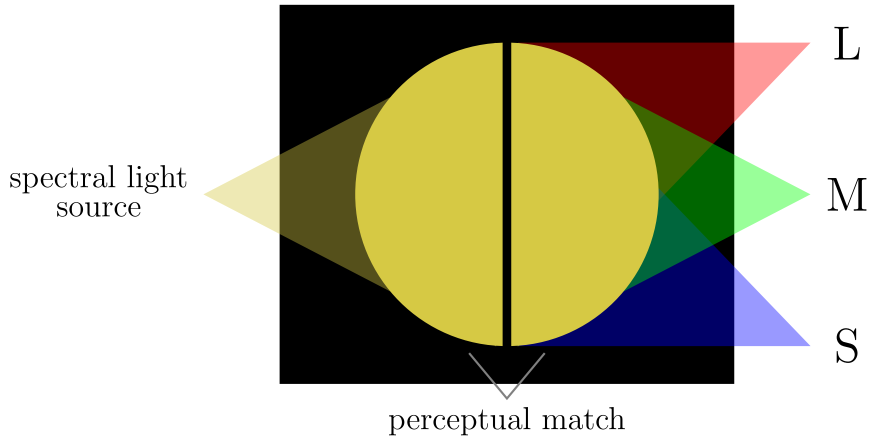


Figure 2.3: A typical setup for colorimetric matching experiments; adapted from Figure 3.3 in Purves and Lotto (2011).

nomenon is known as *metamerism*. Two stimuli are called metameric when a person would match them to look the same even though the stimuli have different physical properties. Metamerism was used to construct perceptual color spaces which were supposed to describe which colors are perceived when looking at different spectra. Irtel (1991) describes how a perceptual color space can be constructed from same-different judgments for stimuli without or with a constant background. He is aware of the fact that this stimulus situation is very restricted, but points out that psychological investigations of color perception have only been successful when focusing on color and object perception separately.[3] This view emphasizes that color spaces constructed this way only apply to limited stimulus situations. In the following, these color spaces will be called trichromatic and referred to as the 'traditional view' (following Niederée, 1998).

Figure 2.3 shows a typical matching experiment used to create a color space that allows us to define different colors as vectors in a three-dimensional space (Purves & Lotto, 2011). Subjects had to adjust the right field using

[3] "Die Objektorientierung unserer Wahrnehmung führt dazu, daß wir Farbe nicht als eine allein dem Subjekt zugängliche Empfindung, sondern als äußere Objektqualität erleben. Für die Psychologie war die Analyse der Farbwahrnehmung allerdings nur dort erfolgreich, wo Farb- und Objektwahrnehmung getrennt wurden."(Irtel, 1991, p. 3 f.)

three light sources of long (L), medium (M), and short (S) wavelengths so it would look like the left field which showed lights with different spectra. These experiments were conducted in the dark and the results are therefore only valid for colors presented in the dark (which means colored lights) or on a uniform background. The best known of these spaces is the CIE 1931 color space (Wyszecki & Stiles, 2000), but there are many others that mostly build on the same empirical assumptions. These color spaces are equivalent up to linear transformations. The underlying assumption of a continuous color space is that “color experiences are tied in a lawful way to properties of the physical world” (Mausfeld, 1998, p. 219). When the physical properties of a stimulus are changed gradually, the perception of its color should change gradually as well.

Building on this assumption, Niederée (1998, 2010) gives theoretical proof that a chromatic color space for infield-surround stimuli presented in a dark room has to be at least four-dimensional. He claims that it is not always possible to match an infield in a (homochromatic) surround to a field without a surround. Infield-surround configurations lead to color perceptions of the infield that are outside of the three-dimensional color space to which colors without a surround belong to. Following a continuous path, he shows that we have to go through a fourth dimension to get to the color outside of this three-dimensional color space. Niederée (1998) stresses explicitly that his dimensionality results can only account for infield-surround stimuli presented in a dark room. More complex settings provide the observer with more context information which are not dealt with by his theoretical considerations. This does not necessarily mean that his conclusions do not hold for these situations, but it should be kept in mind when drawing conclusions for data in an experimental setting different from the ones Niederée includes in his considerations. Niederée (1998, 2010) does not challenge traditional views which claim that color space is three-dimensional. He argues, that color space is reduced to three-dimensions when there is no or a constant surround (cf. Irtel, 1991).

Niederée (1998) emphasizes that a proper topological coordinatization leads to a gradual change of color impressions (what should be expected from a perceptual color space).[4] By proper topological coordinatization he means that, e. g., a multidimensional scaling (MDS) result suggesting $N = 4$ does not necessarily mean that the underlying structure has $N = 4$ dimensions (see Chap. 5.6 in Niederée, 1998, for a critical assessment of MDS techniques to determine the dimensionality of color space). A three-dimensional object embedded in a four-dimensional space would still suggest that there are three underlying perceptual dimensions.

Izmailov and Sokolov (1991) use MDS to show that their discrimination data for equibright stimuli has to be represented in three-dimensional Euclidean space (and not two-dimensional as would be expected by traditional views). However, their MDS solution is a sphere embedded in three-dimensional space. The topological structure of a sphere is two-dimensional, and it is therefore questionable if the results of Izmailov and Sokolov (1991) support a color space with more than three-dimensions (cf. Niederée, 1998, Chap. 5). Furthermore, their stimuli do not fulfill the requirements for a color space with more than three dimensions, since they show their equibright stimuli with constant surrounds which would lead to two dimensions and is in accordance with a trichromatic color space (Niederée, 1998). In their third experiment, Izmailov and Sokolov (1991) show that chromatic colors having different brightness can be represented as a three-dimensional sphere embedded in four-dimensional space. This result does not challenge trichromacy either, since the topological structure is three-dimensional. Again, this would have been expected since the stimuli used by the authors were presented without surrounds on a dark background. If they would have used infield-surround configurations (with different surrounds), their results might have looked different. That they argue against a three-dimensional color space is due to a misunderstanding of the topological structure of color space and what it means

[4] "[...] *topologisch angemessen* in dem Sinne, daß einer stetigen Ortsveränderung in [einer Ebene] E eine stetige Veränderung der zugehörigen Koordinatisierung entspricht." (Niederée, 1998, p. 94)

for its dimensionality. Nonetheless, Izmailov and Sokolov (1991) go about the characterization of color space in a very systematic way. This leads to a good understanding of the underlying variables as can be seen by their comprehensive interpretation of the dimensions they found.

According to Evans (1964), the missing rigorous separation of psychophysical and psychological variables leads to the misunderstanding that only three psychological variables that are connected to the psychophysical ones are needed. Evans (1964) claims that "all possible color perceptions can *not* be described by the use of only three psychological variables. At least four independent variables are required for the 'simplest possible general case,' and there may be others not yet investigated" (p. 1467). He calls these four variables H (hue), S (saturation), B (brightness), and G (gray content) and claims that H, S, and B vary with dominant wavelength, purity, and luminance, respectively, but that we "have to conclude that the *mechanism underlying* gray content is not simply represented by the psychophysical variables so far considered" (p. 1472). He further argues that, therefore, one has to define a new psychophysical variable, but points out: "[...] since it is apparent that gray as a perception *requires the presence of a surround* in the stimulus, the psychophysical variable involved may be of a somewhat different *kind* than the others, since it almost necessarily is a ratio of something in the aperture stimulus to something in the surround stimulus" (p. 1472). Evans' notion of an aperture stimulus can be understood as an infield embedded in a surround here.

These theoretical considerations show that a three-dimensional chromatic color space only holds for specific stimulus situations, e. g., colored lights presented in the dark, and can therefore be considered a special case. For surface colors, the ones that are associated with objects and which are seen with surrounds and under different illuminations, a more general framework that includes context effects and other environmental circumstances like illumination conditions is needed (Mausfeld, 2002).

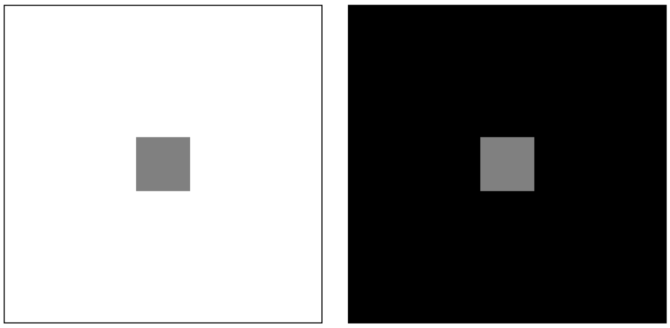
2.4 Lightness Perception

Two effects that have been extensively studied using infield-surround configurations are color constancy and simultaneous color contrast. The following section will give a short overview of these two phenomena focusing on lightness constancy and simultaneous lightness contrast.

Lightness constancy usually refers to the phenomenon that the lightness of objects is perceived as same under different illuminations. Our percept is influenced by a combination of the spectral reflectance properties of a surface and the intensity of light radiating from that surface (see Figure 1.1). In order to obtain lightness constancy, the visual system has to infer the reflectance properties of the surface to deduce its lightness (or color). Identifying the factors in a product of two unknowns is a mathematical impossibility (Adelson, 2000). Nonetheless, we experience a very stable world of colors (Foster, 2003).

When investigating lightness constancy with infield-surround configurations the distinctions between surround, background and illumination established above have to be kept in mind. The term surround is often used synonymously with illumination and it is unclear if subjects really interpret infield-surround stimuli accordingly. Lightness constancy is related to discounting the background and the ratio principle. Both theories imply that two infields are perceived as having the same lightness when the ratio to their surrounds is the same (independent of the luminance of the surround). This has been empirically shown to hold (Wallach, 1948), but there is also evidence that lightness constancy fails (Gilchrist, 2006).

Gilchrist (2006) distinguishes between *background-independent constancy* and *illumination-independent constancy*. This is one of the few times when illumination and background (in the sense of surround) are clearly distinguished. The way lightness constancy is explained above describes illumination-independent constancy. Background-independent constancy refers to the fact that a colored patch does not change color when moved around on backgrounds (surrounds) with different colors.



(a, s) (a, t)

Figure 2.4: Example of simultaneous lightness contrast. Infields have the same luminance.

In order to understand the underlying mechanisms of lightness constancy and to distinguish between background-independent and illumination-independent constancy, experimental settings are needed that let us clearly separate these concepts. In order to understand which perceptual dimensions are involved in lightness constancy, they need to be precisely defined based on empirical results obtained with stimuli presented under controlled conditions which go beyond infield-surround configurations presented in the dark.

Simultaneous lightness contrast refers to the phenomenon that two patches which have the same luminance are perceived as having different lightness when they have different surrounds. The best known version of a simultaneous lightness contrast is shown in Figure 2.4. Infields have the same luminance, but the one embedded in a white surround looks much darker than the one embedded in a black surround. Changing this stimulus slightly increases or decreases the effect, sometimes quite drastically (see, e. g., Chap. 10 in Gilchrist, 2006, for examples). There is a multitude of theories trying to explain this effect (see, e. g., Adelson, 2000; Gilchrist, 2006; Jameson & Hurvich, 1964; Volbrecht & Kliegl, 1998). The most common explanation is lightness (also called brightness) induction. This refers to the phenomenon that a patch of constant luminance appears to get darker when its surround gets lighter (Gilchrist, 2006). The term induction implies a 'transfer' from lighter to darker (or vice versa). The effect is explained by lateral inhibition. Hereby, the luminance of the surround leads to a physiological response that influences the lightness of the infield (Vladusich, Lucassen, & Cornelissen, 2007). Gilchrist (2006) refers to these theories as contrast theories and defines them "[...] as relatively simple, low-level models based on a simple conception of lateral inhibition" (p. 85). Contrast refers to the luminance difference between infield and surround; high contrast means a high luminance ratio. The concept of lateral inhibition in itself leads to misunderstandings when trying to apply a psychophysical frame of reference. Additionally, lightness induction as a physiological process shows its limitations, when looking at effects where simultaneous lightness contrast does not occur or goes in the opposite direction as predicted (again, see Gilchrist, 2006, Chap. 10 for several figures showing a complete reversal of the effect as predicted by contrast theories). This emphasizes that this effect cannot singularly be attributed to low-level perceptual processes (like the encoding of contrast or ratio), but is influenced by environmental circumstances (like some kind of context, e. g., a surround or illuminated background).

Logvinenko (2005) presents evidence that the perception of lightness can be independent of contrast. He put white paper stripes in front of a black sheet of paper and adjusted illumination, so that both surfaces had a luminance of $50\,\frac{\mathrm{cd}}{\mathrm{m}^2}$. Subjects had to match the white stripes as well as the black paper in the background to Munsell chips. The white stripes were judged to be lighter than the dark background. In a second condition, subjects were instructed to focus between the white stripes on the black paper in the background. This led to a coplanar interpretation of the situation (white stripes and black paper were perceived in the same depth plane). In this setting, the stripes (white) and the paper in the background (black) were judged to be of the same amount of Munsell units. Logvinenko (2005) concluded that subjects used illumination information to judge lightness of the stripes and the paper in the background. In different depth planes, the illumination can differ and the white stripes are perceived as being much lighter. When seen in the same depth plane, the illumination has to be the same and the sheet of black paper and the white paper stripes are perceived as having the same lightness.

The ingenious part about this experiment is the fact that the retinal stimulation remains the same for both conditions. Therefore, the contrast between white stripes and background was also exactly the same (keeping in mind that there hardly was any contrast in the first place since stripes and background had the same luminance). Logvinenko (2005) concludes that the difference in perception can be accounted for by the interpretation of the subjects and that the perceptual space for achromatic surface colors has two dimensions: lightness and surface-brightness. Gilchrist (1979) found a similar effect with three-dimensional objects and concludes: “This means that the relation between the target and its background in the retinal image is irrelevant to the target’s perceived shade of gray” (p. 116). This is in strong opposition to traditional contrast theories (as defined by Gilchrist, 2006) trying to explain simultaneous lightness contrast.

The relation between lightness constancy and simultaneous lightness contrast becomes apparent in Gilchrist’s statement that simultaneous lightness “[c]ontrast is a failure of background-independent constancy” (Gilchrist, 2006, p. 152). However, it is unclear how the two effects connect and if the same or similar processes are involved. The same is true when trying to integrate results for both constancy types. This short summary shows that even with simple stimulus configurations there are already phenomena that are difficult to explain with current theories. Contrast theories and other theories that implicitly or explicitly build on assumptions like lateral inhibition are too limited to provide a big picture of what is going on.

2.5 Achromatic Color Space

As opposed to chromatic color space (see Section 2.3), up until now there has been no attempt to formulate a comprehensive perceptual achromatic color space that would allow us to define achromatic colors quantitatively. The following section will give an overview of some theoretical and empirical work that has been done so far.

There are several reasons why an achromatic color space did not get much attention until now. First, traditional views assume (often implicitly) that achromatic colors range from white to black over all shades of gray. This assumption might not induce a strong need to characterize the perceptual space in a better way. As with chromatic color space, the influence of context effects was often neglected. Secondly, achromatic color space (or the perception of lightness) is often considered to be one dimension of chromatic color space (e. g., Wallach, 1963), which might again hinder a more rigorous investigation of the matter. This, of course, is in sharp contrast to the fact that there are almost two centuries of research on lightness and brightness perception. Furthermore, it became obvious within this research that the perception of lightness is a rather complex topic with a plethora of phenomena only partly understood. Characterization of the color space for achromatic colors would lead to a better understanding of the topic. Insights from these theoretical considerations will without doubt contribute to explaining the perception of chromatic colors as well (see, e. g., Wallach, 1963; Whittle, 1994, for some considerations on the generalization of lightness contrast on color contrast).

The traditional view of achromatic color space assumes that this space is one-dimensional. This view has usually not been justified but considered a given. Wallach (1963), e. g., writes: "But there is a family of colors the quality of which does not depend on wavelengths. These are the achromatic, or neutral, colors—white, the various grays and black—which differ from one another only in degree of lightness or darkness. The scale of lightness, in other words, is the only dimension of the neutral colors, although it is one dimension (along with hue and saturation) of the chromatic colors as well" (p. 278). This quote shows that most authors did not see any reason to give the dimensionality of achromatic color space any consideration. It was already *defined* as being one-dimensionally.

The theoretical results from Niederée (1998, 2010) reported in Section 2.3 apply in the same way to achromatic stimuli. These results imply that the perceptual space for infield-surround stimuli (presented in a dark room) has to be at least two-dimensional. Logvinenko and Maloney (2006) found evidence for two perceptual dimensions involved in the perception of lightness with different illuminations. In their experiment, subjects saw a configuration consisting of seven surface chips varying in reflectance on an articulated background under three different illuminations in a side by side display. Subjects had to rate the difference of each pair of surface chips on a 30-point scale. They applied a nonmetric MDS algorithm to the matrix of dissimilarities averaged across all observers and found a two-dimensional solution that described the distances between different patches best. They confirmed their results by fitting a model they call Maximum Likelihood Parametric Scaling (MLPS) which allowed them to test several nested hypotheses. Their data could best be represented with a City-block metric, and they found an interaction between illumination and lightness: "When illumination decreases, the lightness continuum shrinks" (Logvinenko & Maloney, 2006, p. 80). Interpreting the parameters of the model, they concluded that "[c]hanging log reflectance was roughly four times as effective in producing a perceptual difference between two surfaces as changing log illumination intensity" (p. 82). Their model suggests that different lightness of the patches can be arranged on concentric circles. Different radii correspond to different illuminations and the seven gray patches can be arranged on each circle; the patch closest to the origin is seen under the lowest illumination. They conclude that there are two distinct perceptual dimensions of achromatic surface colors and that these "two dimensions are incommensurable: A change in one dimension cannot be compensated for by a change in the other" (p. 83). That the two dimensions lightness and perceived illumination both influence the perception of lightness has been shown by Logvinenko in two experiments before (Logvinenko, 2005; Logvinenko & Ross, 2005).

In their second experiment, Izmailov and Sokolov (1991) presented achromatic stimuli with different surrounds to observers and had them rate their dissimilarity (see Section 2.3 for a description of their other two experiments with chromatic stimuli). They found a two-dimensional solution for their data, although traditional views suggest that only one perceptual dimension is needed to represent achromatic colors. However, the MDS solution shows a one-dimensional structure embedded in two-dimensions. The stimuli follow a semi-circle from the lowest to the highest luminance of the infields. Since they used different surrounds one would expect a two-dimensional solution for their data. Niederée (1998) argues that their subjects might have only paid attention to some salient aspect of the infield, thus neglecting the surrounds.[5] This seems a bit speculative, but nevertheless it is debatable whether their results can be taken as evidence for a traditional view of achromatic color space with $N = 1$. Another argument by Niederée was that the short presentation time of the stimuli ($0.5\,s$) or the low range of the surround luminance could account for the results. Izmailov and Sokolov (1991) used three different surrounds of $0.1\,\frac{\mathrm{cd}}{\mathrm{m}^2}$, $10\,\frac{\mathrm{cd}}{\mathrm{m}^2}$, and $100\,\frac{\mathrm{cd}}{\mathrm{m}^2}$ and Niederée (1998) argues that this might not be a big enough sample of different stimuli configurations to influence the perception of infields. Another important point to consider is that data were averaged across observers. This might lead to a cancellation of different influences of the surrounds for different subjects, therefore, only showing the influence of the infield in the discrimination judgments. The authors interpret their data according to traditional views by saying that the "polar coordinate of a color point in this model is interpreted as brightness[6]—a psychological characteristic of light" (p. 255). Their two-dimensional interpretation in terms of on- and off-cells is explained in more detail in Izmailov and Sokolov (2004), but of no interest for our considerations. Izmailov and Sokolov (1991) found a

[5] "Neben allgemeinen methodischen Problemen, mit denen derartige [MDS] Verfahren behaftet sind, könnte eine der Ursachen dafür, daß sich unser entsprechendes Dimensionsresultat nicht deutlicher in der betreffenden Koordinatisierung widerspiegelt, darin bestehen, daß die Versuchspersonen möglicherweise (einem abstraktiven Farbabgleich analog) nur einen bestimmten salienten Aspekt der Infeldfarbe berücksichtigt haben (wie etwa die Graustufe) oder dieser zumindest weit stärker bei der Beurteilung der Farbähnlichkeit 'ins Gewicht' fiel." (Niederée, 1998, p. 172)

[6] The term brightness used by Izmailov and Sokolov (1991) is better understood as lightness in the terminology of this thesis.

one-dimensional representation for their dissimilarity data but interpreted it as a two-dimensional solution due to the same misunderstanding about topology mentioned earlier. Nevertheless, their approach to investigating dimensionality of color space seems fruitful.

Heggelund (1992) introduces a bidimensional theory for achromatic color vision. He points out that for achromatic colors there is a difference between lightness (object color) and brightness (light color). Heggelund states that these two forms of achromatic colors are distinct (in a similar way Mausfeld, 1998, distinguishes between illumination and object color). He suggests two perceptual variables that he calls w (white) and b (black-luminous) and defines how to determine strength (s) and quality (q) by

$$s = |b| + w, \tag{2.1}$$

$$q = \frac{b}{|b| + w}. \tag{2.2}$$

According to Heggelund (1992), these two perceptual variables "are necessary and sufficient for specification of the achromatic light and object colors in a simple disk-ring configuration" (p. 2110). The strength of an infield is defined as the summation of the two perceptual variables w and b, and the impression of quality is determined by the b process weighted against strength. The quality aspect allows one to distinguish between increments and decrements. Both processes are assumed to be orthogonal with w being the lightness of a stimulus and b a luminous quality with luminous on one end and black on the other end of the dimension. Black and luminous are two ends of one dimension since they seem mutually exclusive in a way red and green are mutually exclusive in opponent-process theory (Jameson & Hurvich, 1964). He found that the b process is primarily related to contrast and w to the luminance of the infield. He then incorporates these results in his mathematical model

$$b = \frac{t^n - i^n}{t^n + i^n + c_1}, \tag{2.3}$$

$$r_1 = \frac{t^n}{t^n + c_2}, \tag{2.4}$$

$$w = r_1(1 - c_3|b|), \tag{2.5}$$

where t is the test luminance (the infield), i is the inducing luminance (the surround) and n, c_1, c_2, and c_3 are constants. The w process depends mostly on local luminance, so Heggelund suggests a two stage response model with r_1 being the response of the first stage, only depending on luminance of the infield. At the second stage some inhibitory influences of the b process can be taken into account.

Heggelund (1992) concludes that his two perceptual variables black-luminous and white are orthogonal and not opponent (as stated, e. g., by Jameson & Hurvich, 1964). Both variables contribute to the perception of an achromatic stimulus at the same time and are, therefore, not mutually exclusive. The two variables seem to explain data for infield-surround configurations quite well as was shown by Heggelund with multiple simulations. He points out that we might need more perceptual variables to properly perceive more complex achromatic scenes and that his w process is comparable to other chromatic processes. This stresses that lightness is interpreted as an object property independent of illumination or viewing conditions.

Niederée (1998) argues that Heggelund's concept of an opponent black-luminous process is theoretically flawed and that the two concepts of illumination (light) color and object color cannot be incorporated in the same theory. It is difficult to distinguish between the two processes as described by Heggelund (1992). The way he describes them gives the impression that he uses the w and the b process to explain differences between increments and decrements. But Heggelund's simulations show that the perception of increments as well as the perception of decrements are influenced by both processes. However, this is not pointed out specifically and has to be read 'between the lines.' The distinction between different color modes is important as has been repeatedly emphasized (e. g., Mausfeld, 2003). In order to understand how these two color modes interact and influence the perception of surface colors, it seems like a fruitful approach to integrate both into a single theory.

The experiments and theoretical considerations described above show that the perception of achromatic colors cannot be as simple as a one-dimensional continuum from white to black over all shades of gray. At least not for stimuli that are more complex than a single gray patch or achromatic light presented in a dark room. We have seen that simple stimulus configurations induce perceptions that cannot be explained by a one-dimensional perceptual space and that illumination influences our perception of achromatic surface colors as well.

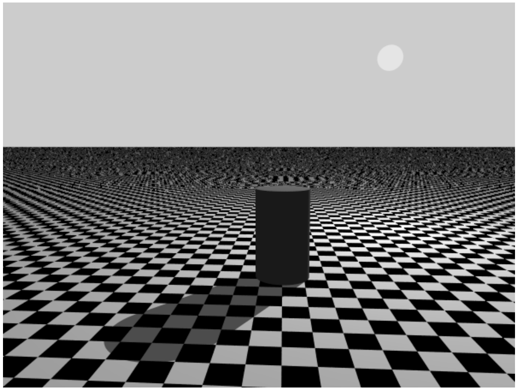
Most of the studies addressing these problems do not explicitly mention a perceptual space for achromatic colors. Thinking in terms of such a perceptual space will contribute to eliminating many conceptual problems in this research area. Defining the dimensions influencing our perception might help define clear-cut definitions of concepts like lightness, brightness, and illumination, for example.

2.6 Cognitive Processes in Color Perception

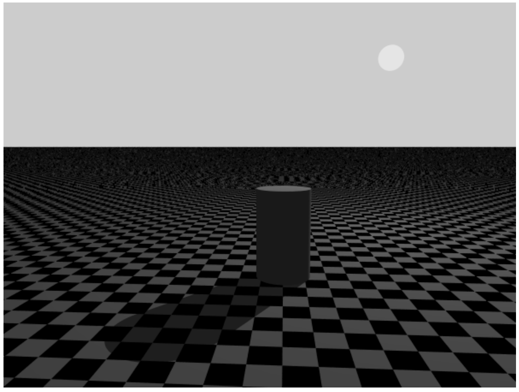
In the opponent-process theory, Hurvich and Jameson (1957) postulate two stages in color perception. Whittle (1994) likewise says that a two-stage model is needed to understand what he calls 'contrast brightness.' Adelson (2000) even distinguishes between low-level vision, high-level vision, and mid-level vision. It is commonly understood that color perception (and also visual perception in general) is a combination of perceptual (low-level) and cognitive (high-level) mechanisms. However, what exactly constitutes these levels is still under dispute. Mausfeld (2010) argues that these two-stage models follow a measurement device conception of perception (see Section 2.1) and are theoretically wrong. He distinguishes between the Sensory System and the Perceptual System. He says that "[t]he *Sensory System* [...] deals with the transduction of physical energy into neural codes and their subsequent transformations into codes that are 'readable' by and fulfil the structural and computational needs of the *Perceptual System*. [...] The *Perceptual System*, on the other hand, can be conceived as a self-contained system of perceptual knowledge, which is coded in the structure of its conceptual forms" (p. 13). The difference to the theoretical notions above is that Mausfeld (2010) claims that "[t]he conceptual forms are yielded, in a given input situation, as outputs of the *Perceptual System* are *triggered* by the codes of the *Sensory System*, rather than being computed or inductively inferred from them. We might loosely think of the triggering functions as an interface function that takes specific sensory codes as an argument and calls conceptual forms" (p. 13 f.). A distinction between perceptual and cognitive processes leaning on Mausfeld (2010) will be used in this thesis.

As mentioned before, this thesis is not concerned with neural processes of visual perception, but solely with psychophysical data and the psychological processes underlying a certain behavior, reaction, or interpretation. Therefore, only a distinction between *perceptual processes* and *cognitive processes* will be established (perceptual processes correspond to the Sensory System of Mausfeld and cognitive processes to his Perceptual System). Perceptual processes will refer to low-level mechanisms associated with phenomena that can be directly related to the stimulation of receptors (like, e. g., contrast coding, see Gilchrist, 2006, for an overview). Hereby, the light reaching the receptors influences the perception. It is well established that in very reduced stimulus situations perception can be explained well with low-level processes (Mausfeld, 2010). Cognitive processes, on the other hand, include every perception that can be accounted for by a certain interpretation of a scene. Adelson's (2000) mid-level processes, which are mostly Gestalt principles, would therefore be called cognitive processes here. Gestalt principles like grouping, coplanarity, belongingness, similarity, proximity, good continuation, and common fate will be considered to be interpretations of a scene. In a stimulus situation like the one used here, namely, infield-surround configurations presented under a constant illumination, basic perceptual processes should not suffice to explain the data.

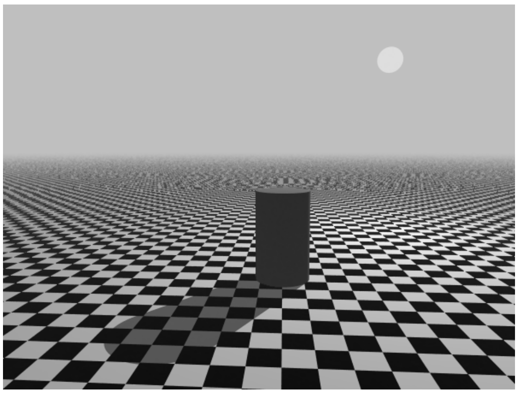
In order to get a better picture about how interpretation of a scene might influence our perception, let us take a look at Figure 2.5. It shows a cylinder on a checker board plane under different conditions (created with the ray tracing



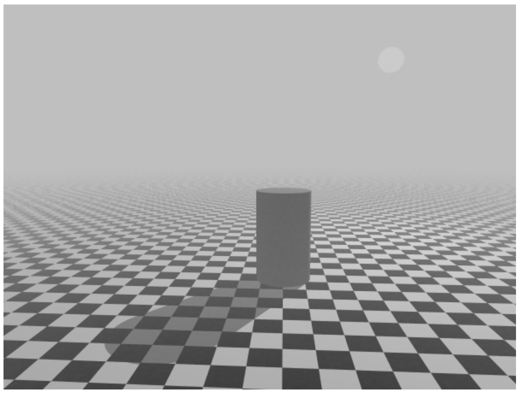
(a) Cylinder on checker board.



(b) Cylinder on low contrast checker board.



(c) Cylinder with blurred horizon.



(d) Cylinder with haze.

Figure 2.5: Cylinder on checker board (adapted after the famous Adelson checker board illusion). Subfigures show different variations, see text for details.

software POV-Ray, version 3.6). Figure 2.5a shows the cylinder illuminated by a single light source. Figure 2.5b has the exact same specifications as 2.5a, except for the contrast between the checker board tiles. One can see that the color of the cylinder looks slightly different, but more striking is how the overall impression of the scene changes. While we would probably say that we see the cylinder illuminated by the sun in Figure 2.5a, we would say that the cylinder in Figure 2.5b is viewed at night and illuminated by the moon (keep in mind that the color of the 'sky' and the light source are identical in both figures). By simply blurring the horizon in Figure 2.5c the picture looks much more natural. The most natural impression is achieved when adding some kind of haze or fog over the whole scene (Figure 2.5d). This effectively blurs the horizon and generally decreases the ratio between all colors. Usually, when we look in the distance, like we seem to do in this picture, we do not have a very clear view but see some kind of haze. For all pictures, we would attest that we are looking at the same cylinder under different viewing conditions, even though the color (lightness) of the cylinder is (or looks) different in the four pictures. The luminance of the top of the cylinder is actually identical in all pictures.

With traditional experimental settings it is difficult to manipulate cognitive processes like interpretation of a scene or the knowledge about certain illumination situations. Mausfeld (1998) claims that it is important to follow a 'continuous path' from simple infield-surround configurations to more complex stimuli and scenes, e. g., via Mondrian patterns, ending in more natural 3D-scenes. They showed in a series of experiments, in which they used what Mausfeld calls Seurat configurations, that subjects used object and illumination information to match infields. In order to avoid ambiguities caused by the use of reduced stimulus configurations, it seems sensible to apply elaborate settings where subjects are able to tell concepts like lightness, brightness, and illumination easier apart.

Helmholtz was the first to emphasize cognitive processes in color vision. He postulated so-called 'unconscious inferences' which the visual system uses to interpret a certain scene. This view has been heavily criticized. Adelson (2000), e. g., argues that the Helmholtzian approach might be overkill, since most effects can be explained with much simpler assumptions than, e. g., a three-dimensional explanation of a certain scene. He points out that simpler (what he calls mid-level) processes can account for these phenomena much easier. Volbrecht and Kliegl (1998) state that “Hering’s physiological explication of simultaneous contrast with dissimilation and assimilation seemed more parsimonious than Helmholtz’s speculations about unconscious inference” (p. 191). Nevertheless, perceptual processes cannot account for all phenomena, and interpretations of scenes seem to play a major role in color perception. After all, color perception developed with a plethora of natural scenes available to the visual system (Purves & Lotto, 2011), and it could be argued that it would be more parsimonious to trigger natural interpretations by reduced stimulus situations like infield-surround configurations (Mausfeld, 1998; Mausfeld & Niederée, 1993). Although Adelson (2000) actually argues in the opposite direction, the following quote from his paper supports this view: “Vision is only possible because there are constraints in the world, i. e., images are not formed by arbitrary random processes. To function in this world, the visual system must exploit the ecology of images—it must ‘know’ the likelihood of various things in the world, and the likelihood that a given image-property could be caused by one or another world-property. This world-knowledge may be hard-wired or learned, and may manifest itself at various levels of processing” (p. 341).

A common framework that has the potential to insert this ‘world-knowledge’ is the Bayesian approach. “The basic idea of these approaches can be described by reference to the Bayesian formula of inverting conditional probabilities: Vision is considered as being based on inferences by which scene properties are estimated from image properties. [. . .] The probability of a world scene given the image (*posterior distribution*) is basically given by the product of the probability of the image given the scene (*likelihood function*) and the a priori probability of the scene (*prior distribution*)” (Mausfeld, 2002,

p. 20). Mausfeld (2002) criticizes that these approaches often ignore that one needs to specify what consitutes a prior distribution (the a priori knowledge our perceptual system provides) in order to make fruitful theoretical contributions to the field.[7]

When looking at natural scenes it is usually effortless to discriminate between different objects or different illumination conditions, i. e., visual information is usually not ambiguous. In lightness research, one often encounters the words 'illusion' and 'error.' Adelson (2000) titles his book chapter "Lightness Perception and Lightness Illusions," and Gilchrist (2006) has a whole chapter on "Errors in Lightness" in his book. They refer to the difference between luminance and lightness of a surface when talking about errors and illusions; or put differently: How veridical our perception is. But the task of the visual system is not to represent 'reality,' but to help us orientate in the world (Mausfeld, 1998, 2002, 2010). This has sometimes been misunderstood when considering lightness and brightness effects. Gilchrist (2006), e. g., writes: "[. . .] the visual system struggles to interpret even such a highly reduced display as a set of surfaces" (p. 328). Probably, the visual system struggles to interpret the display *as a set of surfaces* because it is highly reduced and therefore artificial.

Logvinenko and Ross (2005) use several of Adelson's 'illusions' in two experiments supporting that subjects use higher level interpretations to determine lightness in simultaneous lightness contrast. Subjects rated a whole series of configurations of different complexity presented on a white background of 100 $\frac{\mathrm{cd}}{\mathrm{m}^2}$ in an illuminated room. The results show that the perception of lightness depends on the interpreted illumination condition within a certain configuration. Using different complexities starting with a simple simultaneous lightness configuration (see Figure 2.4) and adding complexity in a systematic way allowed for an interpretation of the results in terms of different perceived

[7] "This problem of choosing an appropriate set of internal primitives also extends to the specification of priors. The priors not only capture statistical dependencies between physical properties of the environment but also crucially refer to the conceptual perceptual structure of the observer." (Mausfeld, 2002, p. 21 f.)

illuminations. Logvinenko and Ross's procedure shows how important it is in lightness perception to use more complex stimuli and to try to connect results for simpler images with results found with more complex ones. They specifically point out that mid-level processes as introduced by Adelson (2000) cannot account for the results and that Helmholtz' interpretation in terms of a 'misjudgement of illumination' seems much more plausible.

In order to distinguish between different color codes like object colors and illumination colors, it seems useful to create experimental settings that allow us to disentangle concepts like lightness, brightness, and illumination. Most experiments on color and lightness perception have been conducted in dark rooms, but surface colors cannot be seen in dark rooms. The first step in constructing an experimental setting that allows us to gradually investigate color stimuli, in our case achromatic color stimuli, is to present stimuli under controlled illumination conditions.

Presenting color stimuli in an illuminated room encourages that colors are perceived as surface colors, exclusively. The perception of surface colors is the most interesting topic when investigating color perception since we usually do not perceive colors as light but as the property of objects (Evans, 1964).

2.7 Research Question

Achromatic surface colors consist of white, black, and all shades of gray. Recent evidence suggests that at least two dimensions are needed to mentally represent achromatic surface colors for infield-surround configurations (Niederée, 1998, 2010) and when stimuli are presented under different illuminations (Logvinenko & Maloney, 2006). Theoretical considerations discussed above show that illumination seems to influence the perception of achromatic colors in fundamental ways. It is unclear how surrounds influence this perception since there has been much confusion in the literature as to what is considered as illumination and, therefore, how context effects influence our perception. This thesis tries to disentangle different concepts by working with an experimental setting that allows us to distinguish between illumination and context by presenting infield-surround configurations under controlled illumination conditions. Illumination conditions are constant for all experiments in order to get a better understanding of how context effects influence lightness perception.

The experiments described in Chapter 5 address the question how many dimensions a perceptual space for achromatic surface colors has. Dimensionality of the perceptual space should be closely related to the stimuli used. Since illumination is held constant over all experiments, the influence of context (in the form of uniform surrounds) on dimensionality of the perceptual space will be investigated. In the first experiment, stimuli will be shown without surrounds under a constant illumination, and in Experiments II and III, there will be uniform surrounds. If context effects add an extra perceptual dimension to the perceptual space of achromatic surface colors, we would expect to find a second dimension in Experiment II and III, but not in Experiment I.

Chapter 3

Methodological Background

> Let us focus on better ways to collect data as well as on modeling that does not pretend respondent's judgments are themselves the measures we seek.
>
> Luce (2004, p. 6)

This chapter focuses on methodical aspects of how one can go about constructing a perceptual space from same-different judgments and what properties data must fulfill in order to do so. The theoretical background for the methods used to analyze data obtained in three experiments reported in Chapter 5 will be introduced. First, psychophysical paradigms often used in color research with a special focus on same-different judgments are discussed. Properties of same-different judgments and associated theoretical notions like the probability-distance hypothesis and observation areas will then be introduced. In order to understand the scaling procedure used in this thesis (Fechnerian scaling), the concepts of regular minimality and nonconstant self-dissimilarity will be explained, followed by an introduction of Fechnerian scaling. The last section focuses on methods (especially multidimensional scaling) which can be used to identify underlying perceptual dimensions.

3.1 Psychophysical Paradigms

Psychophysics uses a plethora of different tasks and paradigms. Two very well known paradigms are often called comparative and equality judgments. Here, we will call the two tasks associated with these two paradigms greater-less and same-different judgments, respectively. When subjects are asked to give greater-less judgments, they are usually presented with two stimuli (more details on presentation modes below) and have to judge if these two stimuli differ with respect to a specific attribute. By repeating this judgment multiple times, experimenters get estimates of probabilities of the form

$$\gamma(x, y) = P(y \text{ is judged to be greater than } x \text{ in attribute } \mathscr{P}) \tag{3.1}$$

(cf. Dzhafarov, 2003a, p. 185). One disadvantage of the greater-less task is that subjects are instructed to give their judgment with regard to a specific attribute in which the two stimuli are supposed to differ. Hence, when considering stimuli which differ in multiple attributes it is difficult to apply greater-less judgments. When judging colors, multiple attributes influence our perception, and Niederée (1998) points out that same-different judgments are better suited than other judgments to overcome some of the conceptual misunderstandings when trying to investigate the perception of color.[1]

Data obtained with same-different judgments are often referred to as *discrimination data*. When conducting an experiment where subjects have to judge if two stimuli are same or different by stating: 'x and y are the same' or 'x and y are different,' we will get a matrix of estimated *discrimination probabilities*

$$\psi(x, y) = P(x \text{ and } y \text{ are different}). \tag{3.2}$$

When performing same-different judgments, the stimulus presentation is iden-

[1] "Die Probleme, welche bereits in der knappen Gegenüberstellung von Helmholtz' und Herings Auffassung anklingen, können zum Anlaß genommen werden, bestimmte Urteile, etwa Urteile über Gleichheit und Ungleichheit (auf denen, zumindest idealiter, insbesondere Farbabgleichsexperimente beruhen), anderen Urteilen prinzipiell vorzuziehen." (Niederée, 1998, p. 92)

tical to the one for greater-less judgments but the decision task involved differs (Schneider & Komlos, 2008). Dzhafarov (2003a) points out that the probability function for $\gamma(x, y)$ is very different from the discrimination probability function $\psi(x, y)$ "[. . .] both in its mathematical properties [. . .] and in the theoretical analysis it affords [. . .]" (p. 185).

Discrimination probabilities are only one way to represent discrimination data. The most common way to collect discrimination data is to show a stimulus pair to subjects and then have them rate 'how similar' or 'how different' these stimuli are. Logvinenko and Maloney (2006) use a 30-point scale, Izmailov and Sokolov (1991) give their subjects a ten-point scale. It is well known that it is difficult to interpret the meaning of scales like these,[2] and it is questionable what subjects actually rate when using them. Additionally, the task is more demanding then a simple same-different judgment.

Another paradigm worth mentioning when talking about color perception is the matching task. This task might be the one used most often to investigate color perception. In a matching task, subjects are presented with two stimuli. One of them is the standard and the other one can be adjusted by the subject (for example by pressing buttons or adjusting a knob). As with the same-different task, there is no need to instruct subjects other than that the two stimuli are supposed to look the same after the adjustment. However, researchers have repeatedly pointed out that matching two colors in different surrounds is not possible (Ekroll & Faul, 2012a; Logvinenko & Maloney, 2006; Mausfeld, 1998; Niederée, 2010). The question arises what subjects actually do when they cannot achieve a satisfactory match. "An evident conjecture is that they select a setting point that minimizes but does not eliminate the perceived dissimilarity between two surfaces [. . .]" (Logvinenko & Maloney, 2006, p. 82).

The basic idea of psychophysical scaling is to "formulate a theory that al-

[2] It should be mentioned that Logvinenko and Maloney (2006) actually trained their subjects on how to use this 30-point scale and presented a reference scale throughout the experiment so that subjects could check their use of the scale, repeatedly.

lows the computation of perceived stimulus properties from purely physical attributes" (Irtel, 2005, p. 1628). Scaling procedures are often divided into direct and indirect scaling. Measurements of psychological magnitude are either derived indirectly via how well two stimuli can be told apart by observers or directly via observers' judgments (Gescheider, 1997). Matching is a direct scaling procedure whereas same-different judgments (and greater-less judgments) are indirect procedures. Gescheider (1988) argues that indirect procedures are the only valid method to measure sensation magnitude but points out that "[i]t has also become widely accepted in the past few years that the validity of such indirect measurements of sensation magnitude cannot be established independent of psychophysical theory" (p. 170). How same-different judgments are connected to Fechner's original theoretical considerations will be explained in the following sections.

From a theoretical point of view, same-different judgments seem to be the task of choice when investigating color perception. However, same-different judgments come for the prize of many problems as will become apparent in the sections describing the experiments conducted for this thesis.

3.2 Probability-Distance Hypothesis

Discrimination data are closely related to the *probability-distance hypothesis* (Dzhafarov, 2002c). The probability-distance hypothesis states that the probability $\psi(x, y)$ to tell two stimuli apart from each other is a function of the (mental or subjective) distance $D(x, y)$ of how different two stimuli are perceived,

$$\psi(x, y) = f(D(x, y)), \tag{3.3}$$

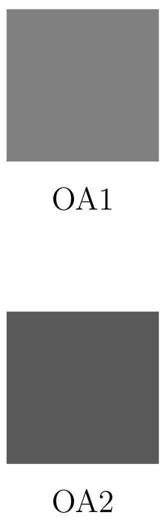
where f is a continuously increasing function. It has been postulated by several authors in several forms and is often referred to as the "Fechner Problem" since Fechner was the first to address the question "how to assign numerical values to unidimensional stimuli so that these values are equidistant for stimuli discriminated with equal probabilities" (Dzhafarov & Colonius, 1999,

p. 240). A quite general form is given by Ekman (1954) who claims that he developed the method of 'similarity analysis' for studying the dimensionality of experience. The underlying assumption of this method is "that the degree of perceived similarity is a function of the degree of overlap between those primary experiences (sensations, emotions) which are evoked by the stimuli" (p. 467). The probability-distance hypothesis itself could be viewed as a special form for discrimination probabilities of the more general form stated by Ekman (1954) for general discrimination data. Dzhafarov (2002c) shows that the distance $D(x, y)$ coincides with the Fechnerian distance $G(x, y)$ if $D(x, y)$ is an internal metric. He states that "$D(x, y)$ is internal if its value equals the infimum of the lengths of all sufficiently smooth paths connecting x and y" (p. 356). We will see in Section 3.5 that this is the definition of a Fechnerian metric and also follows Fechner's original general principle (as stated by Dzhafarov & Colonius, 2011).

3.3 Observation Areas

When comparing stimuli in a psychophysical experiment, these stimuli will belong to what Dzhafarov and Colonius (2001) refer to as *observation areas*. Stimuli belong to different observation areas as soon as they are compared to each other. For two stimuli, they can be presented side by side (see Figure 3.1a), one above the other (see Figure 3.1b), or in successive order (see Figure 3.1c). Subjects are supposed to ignore the different observation areas when judging stimuli. For two stimuli, we have a right and a left observation area, a top and bottom observation area, or a first and second (temporal) observation area, respectively. In psychophysical experiments, it is common to present stimuli in both observation areas (e. g., the reference stimulus or standard is randomly shown left or right), but then this difference is neglected when analyzing the data. However, Dzhafarov (2003a) points out that "[o]ne consequence of treating (x, y) as an ordered pair is that $\psi(x, y)$ and $\psi(y, x)$ in the "same-different" paradigm are generally different, and so are $\gamma(x, y)$ and

(a) Stimuli presented side by side



(b) Stimuli presented one on top of the other (c) Stimuli presented one after the other

Figure 3.1: Schematic overview of the concept of an observation area (OA) for typical presentation modes in psychophysical experiments.

$1 - \gamma(x, y)$ in the "greater-less" paradigm" (p. 188). In the experiments presented here, we consider the ordered pair (x, y) to be different from the ordered pair (y, x), and the pair (x, x) is not considered as a single object, but also a pair following the notation used by Dzhafarov and Colonius, in several of their papers. The discrimination probabilities will be sorted into a matrix with rows belonging to the first and columns belonging to the second observation area.

3.4 Regular Minimality and Nonconstant Self-Dissimilarity

The most fundamental property of discrimination probabilities and also the only requirement for the computation of Fechnerian distances is *regular minimality* (Dzhafarov & Colonius, 2006b). Regular minimality, in its simplest form, can be defined as follows: For any $x \neq y$,

$$\psi(x, x) < min\{\psi(x, y), \psi(y, x)\}, \tag{3.4}$$

which means that we have exactly one stimulus in the first observation area which has its Point of Subjective Equality (PSE) with exactly one stimulus in the second observation area. *Nonconstant self-dissimilarity* means that the magnitude of $\psi(x, x)$ varies for different values of x. Dzhafarov (2002d) points out that regular minimality and nonconstant self-dissimilarity are empirically corroborated for same-different judgments. Figure 3.2 demonstrates these two concepts: The different magnitude of the minima for different standards (x_1 to x_3) shows nonconstant self-dissimilarity, and the shape of the psychometric functions for same-different judgments is characteristic when regular minimality is satisfied.

Dzhafarov (2003b) develops a model that predicts regular minimality and nonconstant self-dissimilarity. The model does not make use of underlying intrinsic images (presented as random variables in perceptual space, like in Thurstonian type models), but employs 'Uncertainty Blobs' in the perceptual space. The underlying concept of this class of models is, that stimulus x in

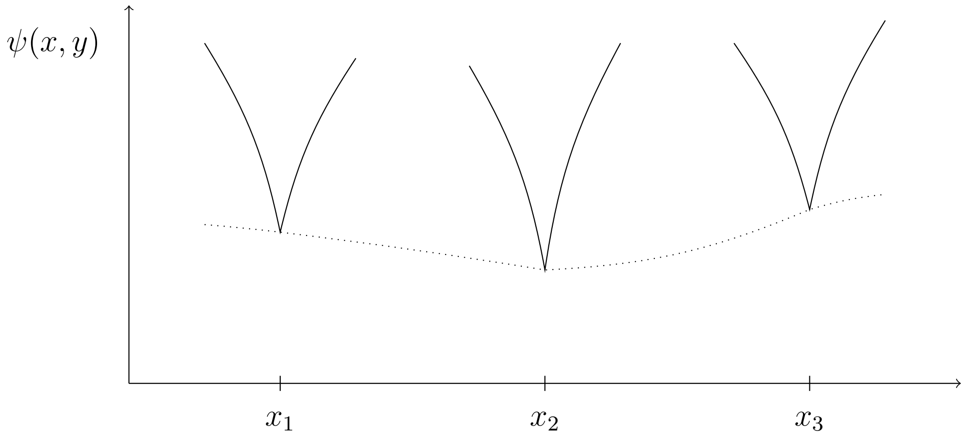


Figure 3.2: Hypothetical shape and position of psychometric functions for same-different judgments when regular minimality holds and data show non-constant self-dissimilarity.

one observation area is the minimum for exactly one stimulus y in the other observation area and vice versa, therefore, ensuring regular minimality. These two stimuli are mutual PSEs and might be relabeled x_a and y_a. The distance between two stimuli consists of a distance between two stimuli from the same observation area, $D(a, b)$, plus a self-dissimilarity R_1 for the first and R_2 for the second observation area for each stimulus. The dissimilarity between two stimuli x_a and y_b is computed as

$$S(x_a, y_b) = R_1(a) + 2D(a, b) + R_2(b). \tag{3.5}$$

Figure 3.3 shows how this can be envisioned. Stimulus x_a belongs to the first observation area (e. g., stimuli presented on the left side) and y_b belongs to the second observation area (e. g., stimuli presented on the right side). In order to determine the dissimilarity between these two stimuli, we have a distance $D(a, b)$ which is determined within one observation area and two self-dissimilarities which depend on the mutual PSEs for stimuli a and b. The probability to judge two stimuli as different, $\psi(x_a, y_b)$, is then an increasing function

$$\psi(x_a, y_b) = \beta\left(S(x_a, y_b)\right) \tag{3.6}$$

of this dissimilarity measure (note that $S(x_a, y_b)$ does not have to be a dis-

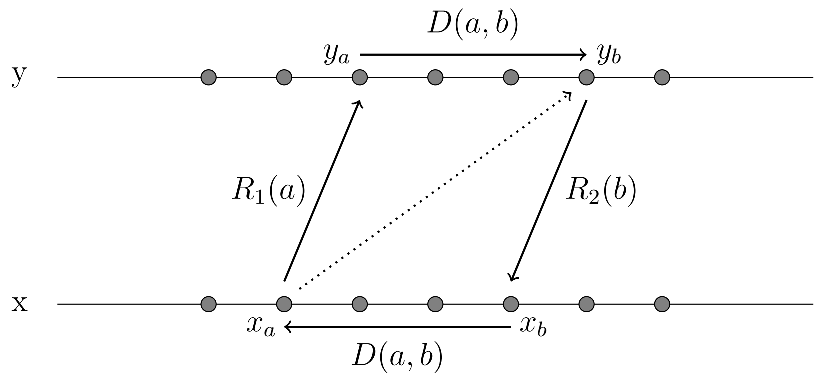


Figure 3.3: Schematic demonstration how Dzhafarov (2003b) calculates the perceptual distance between x_a and y_b in his 'Uncertainty Blobs' model; Figure adapted from Dzhafarov and Colonius (2006b, p. 29).

tance). Defining $S(x_a, y_b)$ in this way ensures that these models create discrimination probability functions which are subject to both regular minimality and nonconstant self-dissimilarity. The choice for $\beta(s)$ is of no relevance for the theoretical implications here. Shepard (1987) introduces a number of different choices for $\beta(s)$.

3.5 Fechnerian Scaling

Izmailov and Sokolov (2004) state that "[t]he idea of measuring subjective distances among stimuli is one of the most fundamental ideas in psychophysics" (p. 27). According to Whittle (1994), Fechner attempted to "construct a scale of sensation by integrating just noticeable differences" (p. 71). Dzhafarov and Colonius (2011) point out that Fechner actually never used just noticeable differences in his derivations. They summarize the "main Fechnerian idea" as the "summation of differential sensitivity values along an interval of stimulus values" (p. 128; cf. Dzhafarov & Colonius, 1999, 2006a). More precisely, they state Fechner's general principle as follows (Dzhafarov & Colonius, 2011, p. 135):

1. For each suprathreshold stimulus x, determine empirically a quantity

$H(x)$ that can be interpreted as a measure of discriminability of x from its neighboring stimuli.

2. Call $H(x)$ the sensitivity of x and define it as:

$$H(x) = \frac{D(x, x + dx)}{dx}$$

3. Integrate $H(x)$ from a to b to obtain the value of subjective dissimilarity $D(a, b)$ between stimuli a and b

$$D(a, b) = \int_a^b H(x)\, dx.$$

Dzhafarov and Colonius (1999, 2001) formalized this general principle in a mathematically rigorous way. The rest of this section will give a brief outline of this so-called Fechnerian scaling focusing on *Fechnerian scaling of discrete object sets* as outlined by Dzhafarov and Colonius (2006a). Generalizations for applying Fechnerian scaling in continuous and discrete-continuous stimulus spaces can be found in Dzhafarov and Colonius (2005a) and Dzhafarov and Colonius (2005b), respectively. Furthermore, more recent and general developments of Fechnerian scaling called "Universal Fechnerian Scaling" and "Dissimilarity Cumulation Theory" can be found in Dzhafarov and Colonius (2007) and Dzhafarov (2008a, 2008b).

Since these theoretical considerations are not relevant for an actual application of Fechnerian scaling, they will not be addressed further in this outline. Hence, whenever referring to Fechnerian scaling, I mean Fechnerian scaling of discrete object sets and the following outline of the main concepts also describes Fechnerian scaling of discrete object sets. Fechnerian scaling of discrete object sets is implemented in R (R Core Team, 2013; Ünlü, Kiefer, & Dzhafarov, 2009) and MATLAB (Rach & Colonius, 2008) for practical applications. Fechnerian scaling has been gradually developed in the following articles: Dzhafarov and Colonius (1999, 2001), Dzhafarov (2002a, 2002b, 2002c, 2002d), Dzhafarov and Colonius (2005a, 2005b, 2007), and Dzhafarov (2008a, 2008b, 2010b).

When conducting a psychophysical experiment, stimuli are necessarily chosen from a discrete and finite set of objects. A theoretical color space is usually considered to be continuous. We assume that we can gradually change, e. g., the intensity of a gray patch and that our visual system is capable of constructing a continuous mental representation of these changes. In experimental settings, we are limited to discrete stimulus spaces by hardware conditions. When applying Fechnerian scaling for discrete object sets, we assume that the underlying dimensions of a color space change continuously and that we use steps small enough to have a reasonable approximation of these continuous changes.

The theory of Fechnerian scaling deals with one of the most basic cognitive abilities, namely, to tell two stimuli apart from each other. It computes subjective distances among stimuli from their pairwise discrimination probabilities. Subjects are presented with two stimuli from a set of stimuli $\{a_1, a_2, \ldots, a_N\}$, $N > 1$, and are required to give one of two answers: 'x and y are the same' or 'x and y are different.' We compute $\psi(x, y)$ and define *psychometric increments*

$$\phi^{(1)}(x, y) = \psi(x, y) - \psi(x, x), \tag{3.7}$$

$$\phi^{(2)}(x, y) = \psi(y, x) - \psi(x, x), \tag{3.8}$$

for each observation area. Due to regular minimality (see Equation 3.4), these increments are all positive. Now, consider a chain from a to b, with at least two elements ($k \geq 2$). In a discrete stimulus space this chain consists of the sum of the psychometric increments along the path from a to b. This defines the *psychometric length of the first kind* for this chain

$$L^{(1)}(x_1, x_2, ..., x_k) = \sum_{m=1}^{k} \phi^{(1)}(x_m, x_{m+1}). \tag{3.9}$$

There is a finite number of psychometric lengths across all possible chains connecting a and b. The minimum of these chains is called the *oriented Fechnerian distance of the first kind*

$$G_1(a, b) = L^{(1)}_{min}(a, b). \tag{3.10}$$

This oriented Fechnerian distance constitutes a geodesic chain from a to b. Accordingly, one can compute the *oriented Fechnerian distance of the second* *kind* for the second observation area. Note that the oriented distances are not computed across the two observation areas, but rather within one observation area. The oriented distances satisfy all properties of a metric,

$$D(x,y) \geq 0 \qquad \text{non-negativity,} \tag{3.11}$$

$$D(x,y) = 0 \text{ iff } x = y \qquad \text{identity of indiscernibles,} \tag{3.12}$$

$$D(x,z) \leq D(x,y) + D(y,z) \qquad \text{triangle inequality,} \tag{3.13}$$

except for symmetry

$$D(x,y) = D(y,x). \tag{3.14}$$

For better interpretation, we sum up the oriented distances from a to b and from b to a and obtain the *overall Fechnerian distance* $G(a,b)$ which now satisfies all properties of a metric. Furthermore, $G(a,b)$ does not depend on the observation area. It can be shown that:

$$G(a,b) = G_1(a,b) + G_1(b,a) = G_2(a,b) + G_2(b,a) \tag{3.15}$$

(see Dzhafarov & Colonius, 2005a, for proof). This geodesic loop gives us a readily interpretable measure of the subjective distance between a and b. The theory states that Fechnerian distances follow directly from the discrimination probabilities when regular minimality is satisfied.

When trying to determine the dimensionality of a color space we can now reformulate this endeavour as the question: How many dimensions do we need to represent the distances between different stimuli? Fechnerian scaling gives us subjective distances for all stimuli and we know that these distances can in principle belong to a space of arbitrary dimensionality (Dzhafarov & Colonius, 2001, 2011). All we have to do is find the answer to the question above (an endeavour which is not trivial at all).

3.6 Identifying Perceptual Dimensions

Dzhafarov and Colonius (2011) point out that its applicability to very general stimulus spaces allows “Fechnerian scaling as a data analytic technique [to rival or complement], depending on one's preference, the widely used multidimensional scaling" (p. 137). Dzhafarov (2010a) shows how metric multidimensional scaling (MDS) performed on Fechnerian distances can reveal the underlying structure of the data (for more applications in this fashion see Dzhafarov & Colonius, 2006a; Dzhafarov & Paramei, 2010).

Using MDS to determine perceptual dimensions underlying the process of perception has a long tradition in psychophysics (see, e. g., Arabie, 1991; Ronacher & Bautz, 1985; Shepard, 1962a, 1962b, 1964, 1974). Ronacher and Bautz (1985) say that "multidimensional scaling experiments [. . .] are based on the assumption that stimuli are *analyzed* in the process of perception according to certain *attributes* and that the overall dissimilarity between stimuli is *combined* in some way from the dissimilarities determined from each component." Usually, dissimilarity data are collected to determine how different subjects perceive a set of stimuli (e. g., Logvinenko & Maloney, 2006). These discrimination data can be discrimination probabilities obtained with same-different judgments or dissimilarity ratings by having subjects judge how similar (or dissimilar) two stimuli can be rated on a scale. Then, MDS is performed on these (dis)similarity data to obtain (e. g., perceptual) distances between stimuli. The problem is, that data obtained with these methods often do not satisfy all (or even any) properties of a metric (see Equations 3.11 to 3.14). In order to deal with this problem, a multitude of non-parametric MDS techniques have been developed (and criticized on multiple levels, see, e. g., Borg & Groenen, 2005).

MDS techniques can be divided into two broad applications: Use of MDS techniques as statistical models, and use of MDS techniques as psychological models (again, see Borg & Groenen, 2005). When used as a statistical model, MDS is usually used as a data reduction technique (other techniques in this fashion are factor analysis or cluster analysis) or for the purpose of data visualization. When used as a psychological model, researchers try to determine underlying perceptual (or other psychological) dimensions. The MDS space is considered to be structurally identical to the psychological space and the distance function is considered to be a composition rule. Logvinenko and Maloney (2006) state that "traditional MDS techniques do not lend themselves to testing explicit hypotheses concerning this structure" (p. 77). They find highly structured results using a nonmetric MDS procedure and interpret their results accordingly (see Section 2.5 for more details), but strive to further support their results by developing a parametric model specifically tailored to their experiment (Maximum Likelihood Parametric Scaling, MLPS). Nevertheless, they developed the model leaning on the results found with MDS.

MDS is not used here to find distances between stimuli, but to find a spatial representation for Fechnerian distances. In this sense, we follow the statistical approach in wanting to find a visualization and the psychological approach in wanting to find the number of underlying perceptual dimensions.

Minkowski metrics are often introduced as possible composition rules for different dimensions. MDS can be applied with different Minkowski metrics

$$d_{ij}(X) = \left[\sum_m |x_{im} - x_{jm}|^p \right]^{\frac{1}{p}}, \quad (p \geq 1). \tag{3.16}$$

Dzhafarov (2002c) shows that Fechnerian distances have the structure of a Minkowski metric, when the underlying perceptual dimensions are separable. It has been suggested that the perceptual system uses the City-block metric (a Minkowski metric with $p = 1$) to represent stimuli with dimensions that are perceptually distinct and the Euclidean metric ($p = 2$) when the perceptual dimensions are integrated in some way (Shepard, 1974). When one of the dimensions dominates the other one, meaning that the other dimension is neglected perceptually, the perceptual system uses a Dominance metric ($p = \infty$, Arabie, 1991).

Borg and Groenen (2005) describe two methods to determine the 'true' Minkowski metric by MDS. The first method is based on the assumption that subjects use a City-block metric when perceptual dimensions are distinct. A City-block metric would be assumed when stimuli would differ from each other in the same way, regardless, whether we look at one or two dimensions. The problem with this method is that this is equally the case for a Dominance metric and it does not allow us to look at a continuum of Minkowski metrics with different ps. The other way to determine the best Minkowski metric is to calculate the stress,

$$\sigma(X) = \sum_{i<j\leq n} w_{ij}(d_{ij}(X) - \delta_{ij})^2, \tag{3.17}$$

where w_{ij} are weights and δ_{ij} the observed distances, for a given number of dimensions for different values of p and take the one with the lowest stress. However, there is no statistical test to check if different stress values differ significantly. So the results of this procedure can only be interpreted as giving a hint on the underlying structure of the perceptual dimensions.

Würger, Maloney, and Krauskopf (1995) show that proximity judgments of equiluminant colored light follow a City-block metric and not a Euclidean metric using a test for proximity judgments that measures the angle between intersecting lines in color space. Stimuli were presented on a neutral background and subjects had to judge which of two stimuli was more similar to a standard. Ronacher and Bautz (1985) describe a simple mathematical procedure to find the best parameter p of all Minkowski metrics for stimuli varying in size and lightness (they call it brightness). Stimuli were circles of different sizes cut out of Munsell paper presented on white filing cards. The room was darkened and only illuminated by fluorescent tubes. They found individual differences for observers that mostly followed either a City-block or a Euclidean metric and a what they call intermediary Minkowski metric with $p = 1.4$, for three of their subjects. They argue that there might be a perceptual continuum between the City-block and the Euclidean metric. However, it is difficult to conclude what these intermediary metrics mean in terms of the perceptual integration of stimulus dimensions.

These considerations show that, although finding perceptual dimensions and their underlying structure is a much discussed topic in psychophysics, there are no easy methods to answer these questions. Applying Fechnerian scaling on discrimination data overcomes many problems that applying MDS techniques directly on these kind of data has. It has a theoretical foundation dating back on Fechner's original thoughts (Dzhafarov & Colonius, 2011) and allows us to compute proper subjective distances for a specific stimulus space. But finding the number of underlying perceptual (or psychological) dimensions is difficult and a problem that still needs methodological development.

Chapter 4

Color Laboratory

The color laboratory is equipped to show stimuli under controlled illumination conditions on an LCD or CRT monitor display. The stimuli used in this thesis were all achromatic as was the illumination of the room. Figure 4.1 shows a schematic overview of the laboratory. Subjects sit in a booth and their whole visual field is limited by white walls on each side. The cut-out in the wall shows part of the monitor. The background color of the monitor has the same color and luminance as the white walls. Fluorescent tubes at the ceiling of the booth are adjusted, so that the wall and the monitor background are perceptually indistinguishable (for more details see Section 4.2). Subjects have their chin on a chin rest and are 90 cm away from the wall and 140 cm away from the monitor. The monitor sits in a black box so that no light from the tubes can fall onto the display.

4.1 Hardware

For the experiments presented in this thesis, stimuli are presented on a black-and-white 21.3" monitor of the model EIZO RadiForce GS320. These monitors are especially developed for high performance X-ray diagnostics and can display 1024 different shades of gray with a maximum luminance of 1000 $\frac{\mathrm{cd}}{\mathrm{m}^2}$. The computer is equipped with an NVidia Quadro FX 1800 graphics board. The walls of the booth are painted white (RAL 9010, semi matte) and illuminated

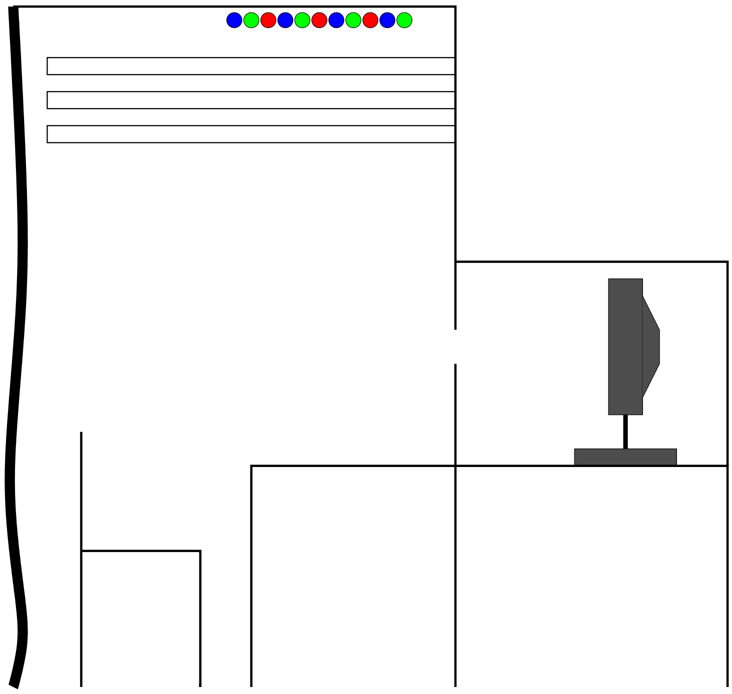


Figure 4.1: Schematic overview of laboratory.

with 12 fluorescent tubes from OSRAM in red (58 W, color 60), green (58 W, color 66), and blue (58 W, color 67). The tubes are dimmable and driven by an ADIODA-PCIF12 MDA PCI multifunction card from Wasco which runs with 12 bits. Light from the tubes is filtered by three sets of diffusion foil (diffusion filter 216 White Diffusion by Lee Filters). Depending on the intensity needed for the experimental setting, neutral density filters can be used to reduce the intensity of the tubes below their dimmable range. We use an i1 Basic Pro photometer to measure luminance and spectra of the setting.

4.2 Software

In order to conduct all measurements needed and to control the tubes and the photometer, we wrote a python package called `achrolab` (`www.uni-tuebingen.de/psychologie/meth/achrolab`) that includes all necessary functions and classes. The package incorporates two submodules called `eyeone` and `wasco` that are responsible for controlling the photometer

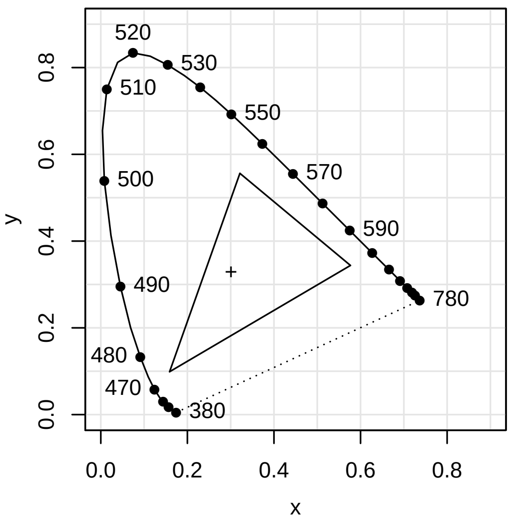


Figure 4.2: CIE 1931 plot with gamut of tubes (triangle). Cross shows coordinates for black-and-white monitor. Tubes are adjusted so that monitor and wall are perceptually indistinguishable.

and the multifunction card driving the tubes, respectively. The two submodules can be used independently to control an i1 Basic Pro photometer and a Wasco multifunction card in every setting. The functionality of these modules is somewhat limited to our needs, but could be expanded easily. The package `achrolab` is very specific for our experimental setting.

The python package `achrolab` contains all functions and classes to control the equipment of our color laboratory. Its main functionality is to match the color and luminance of the wall with the background color and luminance of the 10-bit black-and-white monitor. This matching process is conducted in several steps. First, the luminance functions of the tubes are determined. This is done by measuring the whole range for one color channel (e. g., all red tubes) from the lowest to its highest intensity. Then, we fit a nonlinear regression function

$$y(x) = \phi_1 + (\phi_2 - \phi_1) \exp\left(-\exp\left(\phi_3\right)x\right) \tag{4.1}$$

(Pinheiro & Bates, 2000), with ϕ_1 being an asymptote, $\phi_2 = y(0)$, and ϕ_3 the logarithm of the rate constant, to the data. This is repeated for all three color channels (red, green, and blue). The nonlinear regression function was

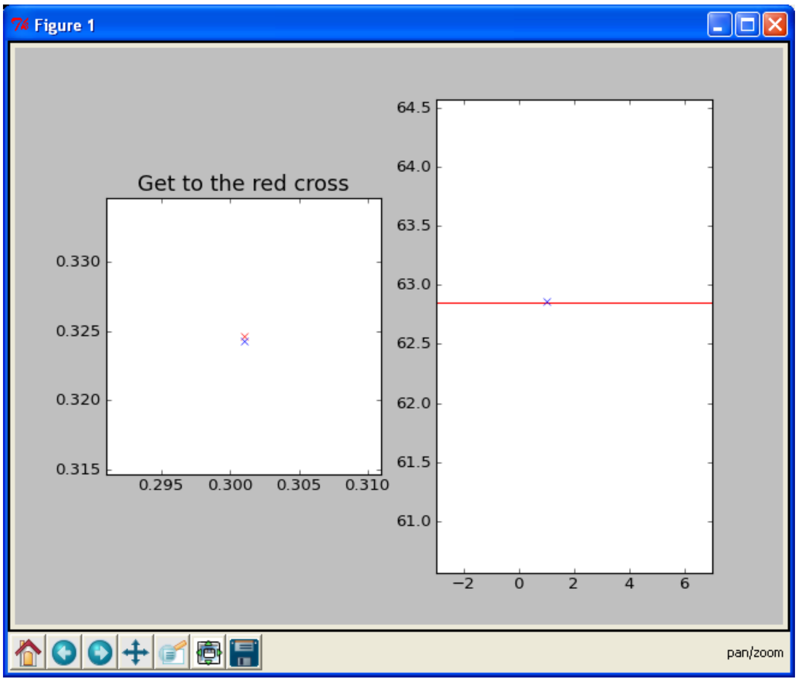


Figure 4.3: Screen shot of graphical interface for adjusting tubes.

used since it fitted the luminance curves best for all tubes. It should be kept in mind that this function does not have a theoretical interpretation. It was chosen after trying several different approaches; the most important criterion for it being that the function is invertible.

After fitting the nonlinear regression functions, we measure the color of the monitor in xyY coordinates of the CIE 1931 color space, where x and y determine the color of the display and Y gives the luminance in $\frac{\mathrm{cd}}{\mathrm{m}^2}$ (see Figure 4.2). The inverted luminance functions are then used to determine a rough estimation of the color of the wall that would best fit the color of the monitor. Next, we measure the color of the wall in xyY coordinates and compare these with the ones we obtained for the monitor on a graphical display shown in Figure 4.3. The red cross shows the xy coordinates for the monitor and the red line shows the luminance Y in $\frac{\mathrm{cd}}{\mathrm{m}^2}$. The blue crosses are the measured xyY coordinates for the wall. By adjusting the tubes in small steps manually, the coordinates can be matched as closely as possible (blue cross in Figure 4.3). After that, the adjusted color is shown which means that the monitor background is set to the measured color and the tubes are adjusted accordingly. Usually, this configuration does not look completely perceptually indistinguishable. Therefore, we apply the following perceptual adjustment procedure: The tubes are manually changed in small steps until the walls *look* exactly like the monitor background. This can be done via the key board or via knobs which control the luminance of the tubes.

Once adjusted, the setting is very robust. After switching on the tubes and the monitor and having both run for a couple of hours, the adjustment is effective for several weeks, i. e., walls and monitor background are perceptually indistinguishable. After a couple of months, a new adjustment might be in order due to changes in the color of the wall. The wall color is very susceptible to the extensive exposure to the light of the tubes (it 'bleaches out').

Chapter 5

Experiments

This chapter describes the experiments that were conducted in order to investigate the dimensionality of color space for achromatic surface colors. The experiments were all conducted in an illuminated room with a constant illumination (see Chapter 4). Illumination was identical for all experiments. Presentation of stimuli under controlled illumination conditions was supposed to help disentangle different concepts like illumination, background, and surround. Furthermore, presenting stimuli in an illuminated room was supposed to ensure that all stimuli were perceived as surface colors (even though presented on a computer screen). All stimuli were decrements, further ensuring that subjects perceived surface colors.

Investigating achromatic color space for surface colors for single gray patches and infield-surround configurations in a systematic way will contribute on a fundamental level to understanding the relevant concepts playing a role in lightness perception. In Section 2.2, we saw that the literature on lightness perception is full of inconsistently used concepts. In order to disentangle the different roles of context effects and illumination on the perception of lightness, we need experimental settings which allow us to clearly attribute results to certain stimulus conditions. Furthermore, it is important to clearly describe stimulus situations and to relate conclusions to these stimulus conditions. An achromatic color space for simple gray patches will look different than one for infield-surround configurations. It is advisable to first look at different stimulus situations separately and then try to integrate the results into a common theory. Stimuli should increase in complexity over the course of several experiments (Mausfeld, 1998) so that results can be related to each other.

Data were analyzed individually for all subjects. In all experiments, subjects performed same-different judgments. As mentioned before, same-different judgments are supposed to overcome many conceptual problems especially for color stimuli (Niederée, 1998). The intention was to overcome as many conceptual problems as possible by using same-different judgments and a constant illumination.

Fechnerian scaling provides a tool to derive subjective distances for individual subjects from discrimination probabilities obtained with same-different judgments. It is based on a solid theoretical foundation dating back to Fechner's original idea (Dzhafarov & Colonius, 2011). Fechnerian scaling is suited to deal with stimulus spaces of arbitrary dimensionality (Dzhafarov & Colonius, 1999). No assumptions about the physical properties of these dimensions have to be made (Dzhafarov & Colonius, 2005a).

In the first experiment, simple gray patches (without surrounds) under a constant illumination were presented to subjects. In the second and third experiment, subjects judged infield-surround configurations under the same illumination conditions. Fechnerian scaling was applied to data of the first two experiments in an attempt to uncover the underlying perceptual dimensions involved in the perception of achromatic surface colors. The third experiment focused on the underlying psychometric functions for same-different judgments for infield-surround configurations and what their shape reveals about the number of underlying perceptual dimensions.

5.1 Experiment I: Simple Gray Patches

In the first experiment, simple gray patches were presented to subjects in an illuminated room. Subjects had to judge whether two stimuli looked same or different. This experiment was conducted to show that not illumination alone or the presence of a background introduce a new dimension, but only more complex stimuli like infield-surround configurations used in Experiments II and III. In a stimulus configuration, where there is no other interpretation of the scene than a single object under a certain illumination, lightness is expected to be the only perceptual dimension needed to discriminate between objects with different luminance. Thus, a one-dimensional achromatic color space for a configuration with two gray patches presented next to each other under a constant illumination of the room should be found.

In a second condition, subjects carried out a greater-less task and judged which of two stimuli looked lighter. This condition was conducted to investigate if there was perceptual learning over the course of the experiment (see detailed explanation below).

5.1.1 Subjects

Four female subjects (aged 21 to 28) participated in the experiment. All subjects were naïve as to the purpose of the experiment and had normal or corrected to normal vision. Subjects were undergraduate psychology students at the University of Tübingen participating in the experiment as part of their psychology curriculum. They were tested for color inefficiencies with a German version of the Ishihara Test (Velhagen & Broschmann, 1985). All subjects had normal color vision. Subject 3 was left-handed.

5.1.2 Stimuli and Procedure

Table 5.1 shows luminance values of the nine stimuli presented in the first experiment. Stimuli were rendered on a computer screen using PsychoPy (Peirce, 2007, 2009) and presented in an illuminated room (see Figure 4.1) where the background filled the complete visual field of the subject. Stimuli had a size of 0.81 degrees of visual angle and were presented in pairs side by side on the monitor with a distance of 2.32 degrees of visual angle. During the whole experiment, a fixation cross of 0.05 degrees of visual angle and a luminance of

Table 5.1: Luminance Values of Stimuli Used in Experiment I

Stimulus	Luminance ($\frac{cd}{m^2}$)
0	37.35
1	38.69
2	40.05
3	41.43
4	**42.88**
5	44.34
6	45.89
7	47.45
8	49.07
Background	133.27

Note: Luminance values are evenly spaced on a log scale. Stimulus 4 (bold) was the standard for greater-less judgments.

$73.97\,\frac{cd}{m^2}$ was presented in the middle between the two stimuli. Stimuli were presented until subjects responded or for a maximum of 500 ms, with an inter stimulus interval lasting for 2 s in order to avoid afterimages. Subjects were seated with a distance of 140 cm to the monitor (90 cm to the wall) with their heads secured by a chin rest with forehead support.

The first experiment consisted of 20 sessions. In 5 sessions (sessions 1 and 17 to 20) subjects judged which of two stimuli presented on the screen was lighter. For the remaining 15 sessions subjects judged if the two stimuli were same or different. The stimulus presentation was identical in both conditions; the only difference being the instructions and that subjects performed judgments for all possible pairs in the same-different condition.

Greater-Less Judgments

Subjects performed greater-less judgments for sessions 1 and 17 to 20 (except for Subject 1 who only did 17 sessions in total of which the first and the last were greater-less judgments). The first session was meant to test whether subjects were able to distinguish between stimuli. Responses were analyzed after the first session, and a psychometric function was fitted to decide if subjects could discriminate between stimuli.

Stimulus 4 (see Table 5.1) was the standard stimulus for the greater less judgments. All other stimuli were presented in random order, paired with Stimulus 4. The pair $(i, 4)$ was considered to be different from pair $(4, i)$, with $i = 0, \ldots, 8$. All pairs were presented 25 times in each session. Trials were split into 25 blocks with 17 trials each (except for Subject 1 who did 30 blocks with 15 trials each and repetitions of pairs differed, see Appendix A.1).

Same-Different Judgments

For sessions 2 to 16, subjects performed same-different judgments on each stimulus pair by pressing the right or left button on a computer mouse. Half of the subjects were instructed to press the left mouse button when the stimuli looked the same and the other half was instructed to press the left mouse button when the stimuli looked different. Stimulus pairs were presented multiple times. The number of repetitions differed for different pairs (see Table 5.2). In total, subjects did 8,940 trials distributed over 15 sessions on 15 different days. The number of trials varied for different pairs, in order to minimize the response bias for pressing different. However, this response bias could not be eliminated completely. There were 6,240 trials (about 70 %) for which the stimuli were different, as opposed to only 2,700 trials in which stimuli were the same. Subjects did not get feedback on correct or incorrect trials and were also not instructed to *identify* which stimuli were the same or different, but to answer which stimulus pairs *looked* the same or different to them. One session consisted of 40 experimental blocks with 15 trials each. At the beginning of each session two training blocks were presented, so that subjects would adapt to the illumination and memorize which button was same and which was different. Responses from the training blocks were not included in data analysis.

Table 5.2: Number of Repetitions for Each Stimulus Pair in Experiment I

	0	1	2	3	4	5	6	7	8
0	300	160	120	80	60	40	20	0	0
1	160	300	160	120	80	60	40	20	0
2	120	160	300	160	120	80	60	40	20
3	80	120	160	300	160	120	80	60	40
4	60	80	120	160	300	160	120	80	60
5	40	60	80	120	160	300	160	120	80
6	20	40	60	80	120	160	300	160	120
7	0	20	40	60	80	120	160	300	160
8	0	0	20	40	60	80	120	160	300

5.1.3 Results

Greater-Less Judgments

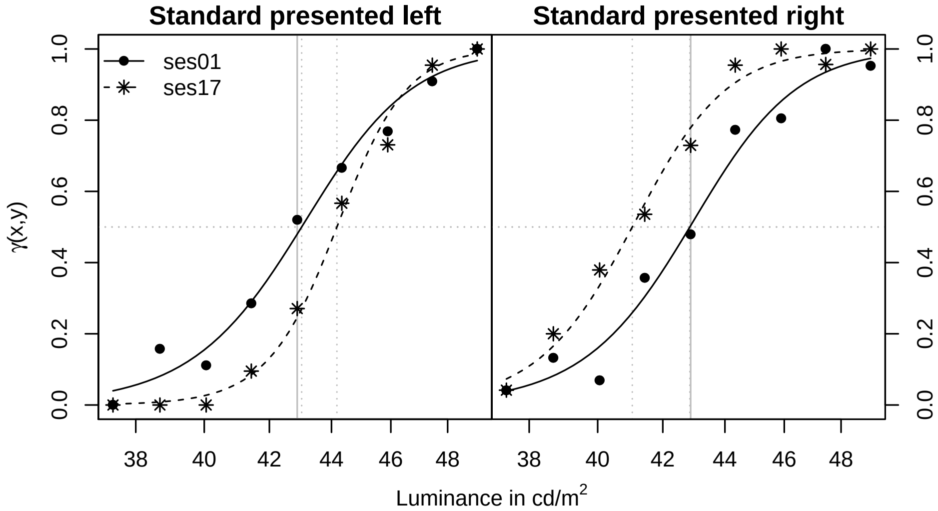
Figures 5.1a, 5.2a, 5.3a, and 5.4a show psychometric functions for all four subjects for greater-less judgments. Subjects responded by pressing the left button when the left stimulus seemed to be lighter and vice versa. Responses are therefore binary and can be represented with an indicator variable

$$Z = \begin{cases} 1 & \text{if } x \text{ is judged greater than } y, \\ 0 & \text{if } x \text{ is judged less than } y. \end{cases} \tag{5.1}$$

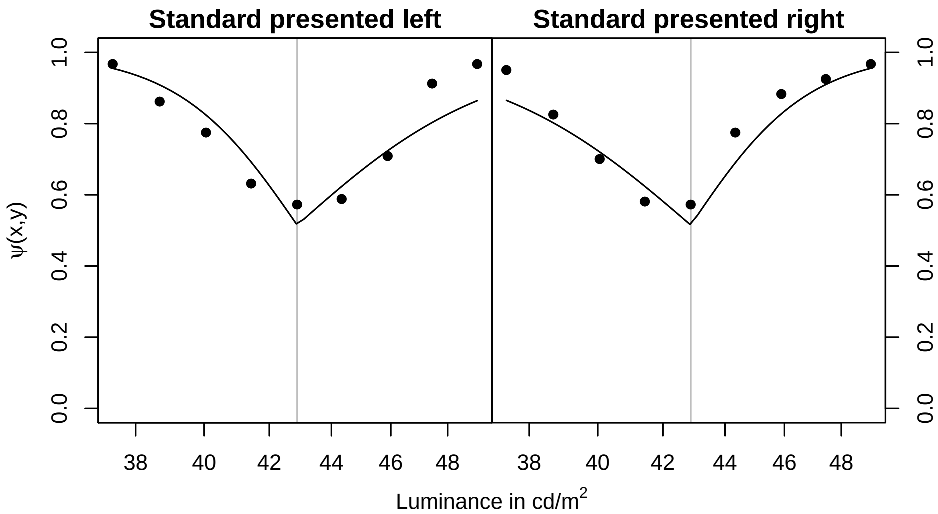
Logistic psychometric functions of the form

$$\log \frac{P(Z=1)}{1-P(Z=1)} = \beta_0 + \beta_1 \text{Luminance} \tag{5.2}$$

were fitted to the responses. This model showed a good fit for most of the data points (see Table 5.3) except for Subject 2 when the standard was presented

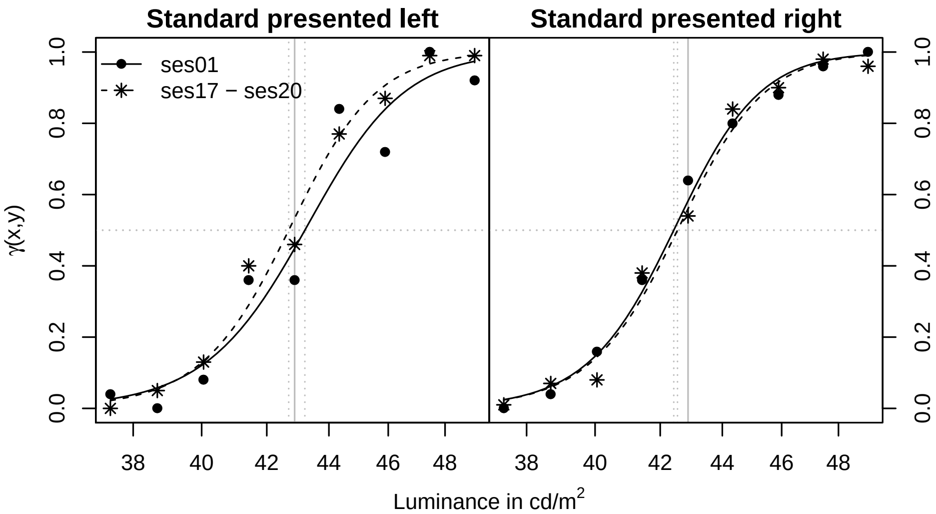


(a) Greater-less judgments. Goodness-of-fit tests see Table 5.3.

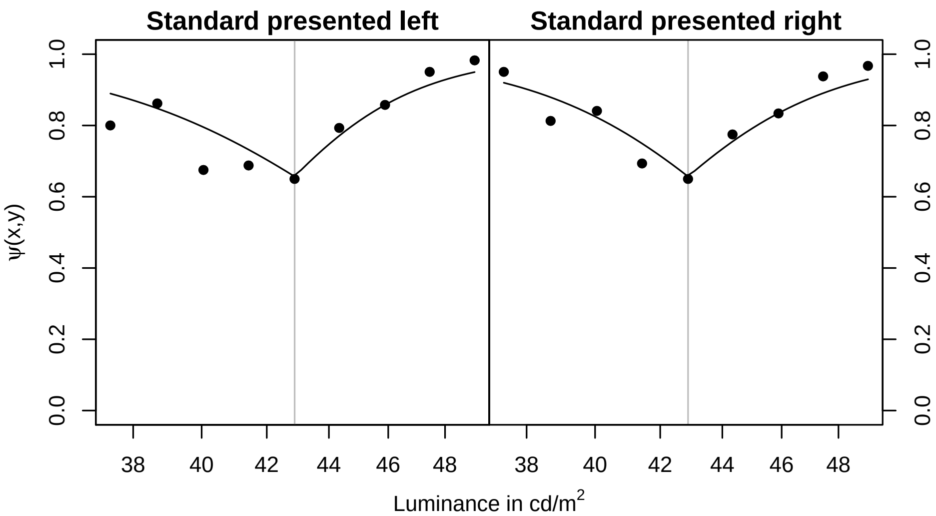


(b) Same-different judgments. Fitted with Quadrilateral Dissimilarity Model.

Figure 5.1: Psychometric functions for Subject 1.

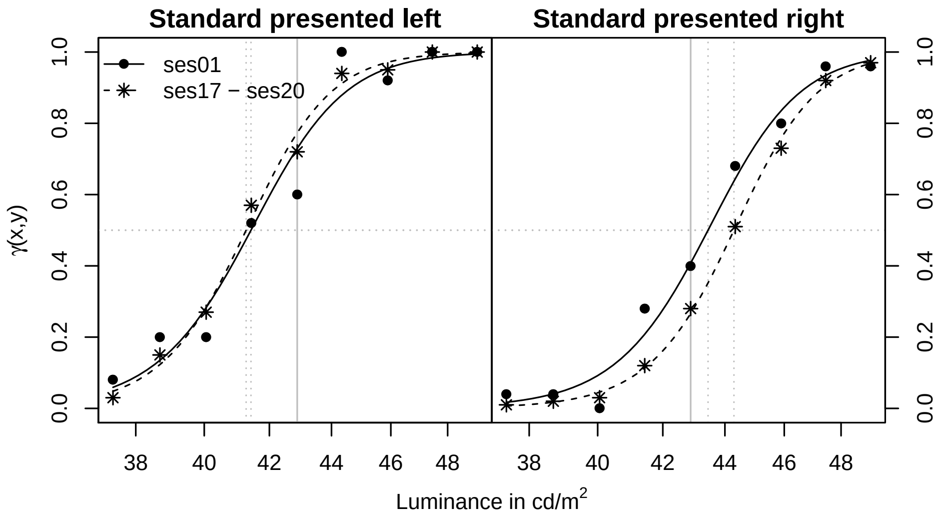


(a) Greater-less judgments. Goodness-of-fit tests see Table 5.3.

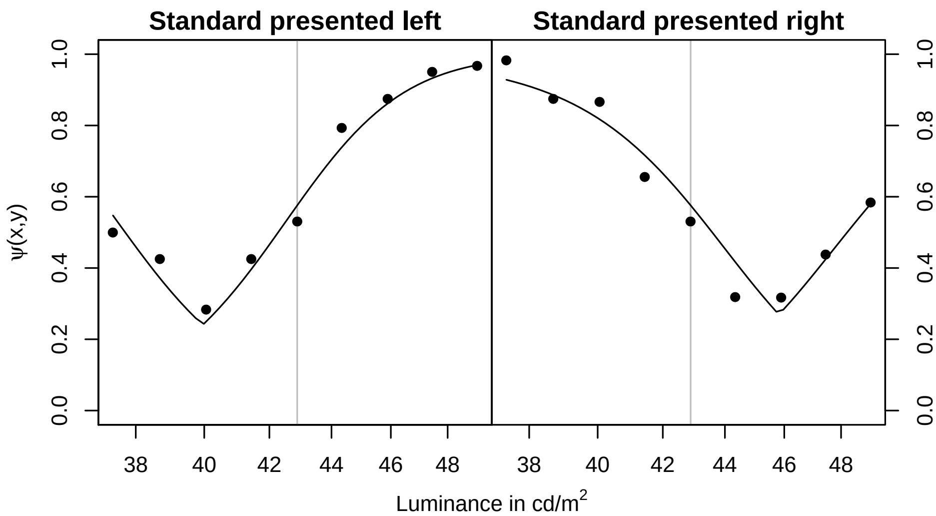


(b) Same-different judgments. Fitted with Quadrilateral Dissimilarity Model.

Figure 5.2: Psychometric functions for Subject 2.

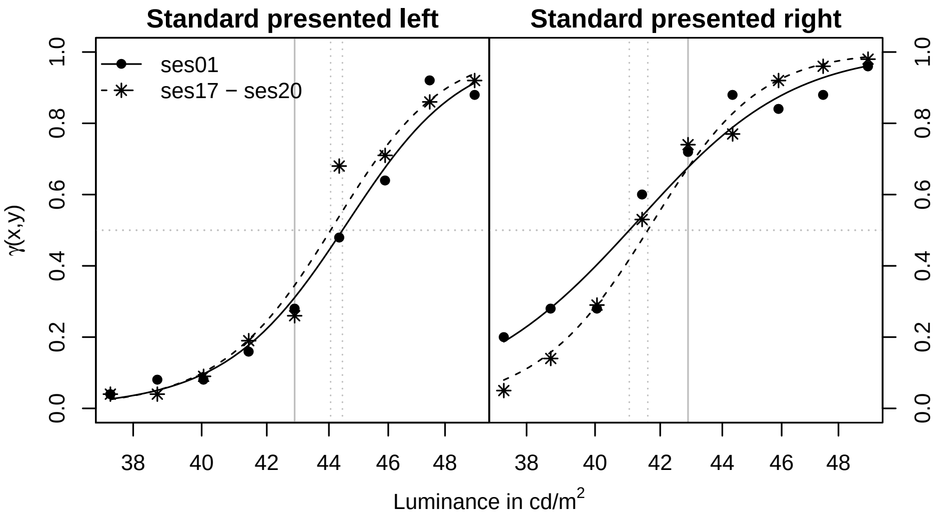


(a) Greater-less judgments. Goodness-of-fit tests see Table 5.3.

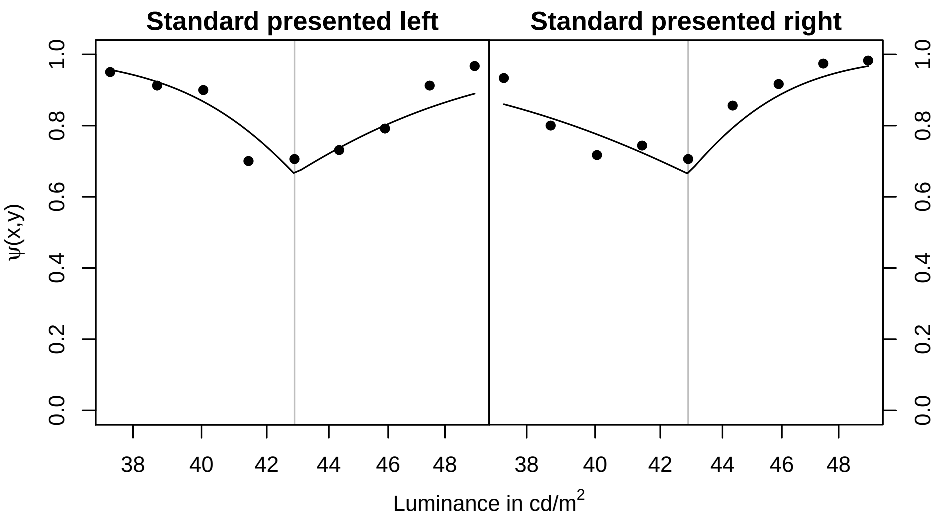


(b) Same-different judgments. Fitted with Quadrilateral Dissimilarity Model.

Figure 5.3: Psychometric functions for Subject 3.



(a) Greater-less judgments. Goodness-of-fit tests see Table 5.3.



(b) Same-different judgments. Fitted with Quadrilateral Dissimilarity Model.

Figure 5.4: Psychometric functions for Subject 4.

Table 5.3: Goodness-of-Fit Tests for Psychometric Functions for Greater-Less Judgments

		Session 1		Sessions 17–20	
	Standard	$G^2(7)$	p	$G^2(7)$	p
Subject 1	left	6.26	0.510	3.61	0.823
	right	8.29	0.309	6.26	0.510
Subject 2	left	17.42	**0.015**	15.92	**0.026**
	right	3.18	0.868	15.76	**0.027**
Subject 3	left	11.99	0.101	8.48	0.292
	right	7.22	0.407	1.62	0.978
Subject 4	left	3.24	0.862	13.66	0.057
	right	4.68	0.699	7.38	0.391

Note: See Figures 5.1a, 5.2a, 5.3a, and 5.4a. Bold numbers indicate lack of fit of the model in Equation 5.2.

on the left. In order to test for time effects, two more complex models

$$\log \frac{P(Z=1)}{1-P(Z=1)} = \beta_0 + \beta_1 \text{Luminance} + \beta_2 \text{Time}, \tag{5.3}$$

$$\log \frac{P(Z=1)}{1-P(Z=1)} = \beta_0 + \beta_1 \text{Luminance} + \beta_2 \text{Time} + \beta_3 (\text{Luminance} \times \text{Time}) \tag{5.4}$$

were fitted to the data. Time is a binary predictor with two levels: first session vs. last four sessions. The models were sequentially tested against each other in order to test if subjects' performance changed over the course of the experiment. Models in Equations 5.2 and 5.3 were tested against each other using a likelihood-ratio test in order to test for a shift of the psychometric function over time. Comparing the models in Equations 5.3 and 5.4 tested if there were additional differences in slope for the psychometric functions. Table 5.4 shows the results for likelihood-ratio tests for each subject. The results show that, overall, the time effects are small and reveal no systematic pattern. Thus, we conclude that subjects' performance did not change in a

Table 5.4: Likelihood-Ratio Tests of Time and Time×Luminance Effects

		Time		Time×Lum.	
	Standard	$\Delta G^2(1)$	p	$\Delta G^2(1)$	p
Subject 1	left	9.66	**0.002**	5.44	**0.019**
	right	21.92	<**0.001**	0.98	0.322
Subject 2	left	2.37	0.124	1.06	0.304
	right	0.14	0.713	0.01	0.909
Subject 3	left	0.26	0.608	0.63	0.426
	right	6.92	**0.009**	0.31	0.577
Subject 4	left	0.90	0.342	0.25	0.615
	right	1.12	0.290	6.45	**0.011**

Note: See Figures 5.1a, 5.2a, 5.3a, and 5.4a. Bold numbers indicate significant differences between models.

systematic way over the course of the experiment. The effect of perceptual learning, if any, seems to be small.

Same-Different Judgments

Estimated discrimination probabilities for the same-different judgments for all subjects are shown in Tables 5.5, 5.6, 5.7, and 5.8 and Figures 5.1b, 5.2b, 5.3b, and 5.4b show psychometric functions for same-different judgments when Stimulus 4 (see Table 5.1) is the standard. For same-different judgments, Dzhafarov and Colonius (2006a) define the Point of Subjective Equality (PSE) as the minimum for each row and column. Subject 3 showed a strong perceptual bias. She consistently judged stimuli on the left to be darker than stimuli on the right; e. g., Stimulus 2 on the left has its PSE with Stimulus 0 on the right; Stimulus 4 on the right has its PSE with Stimulus 6 on the left (compare Figure 5.3b). If this bias is taken into account, the stimulus space is reduced to seven stimuli, since Stimuli 0 and 1 on the left and Stimuli 7 and 8 on the right cannot have PSEs for this subject. The psychometric functions are

Table 5.5: Discrimination Probabilities for Subject 1

	0	1	2	3	4	5	6	7	8
0	0.503	0.581	0.558	0.850	0.950	0.975	1.000	1.000	1.000
1	0.725	0.503	**0.494**	0.625	0.825	0.850	0.975	1.000	1.000
2	0.858	0.637	**0.593**	0.613	0.700	0.887	0.950	1.000	1.000
3	0.988	0.850	0.762	0.543	0.581	0.717	0.800	0.917	1.000
4	0.967	0.863	0.775	0.631	0.573	0.588	0.708	0.912	0.967
5	1.000	0.967	0.963	0.783	0.775	0.550	0.562	0.700	0.812
6	1.000	1.000	0.983	0.838	0.883	0.631	0.553	0.662	0.717
7	1.000	1.000	1.000	0.967	0.925	0.875	0.794	0.490	**0.512**
8	1.000	1.000	1.000	0.975	0.967	0.900	0.767	0.625	**0.570**

Note: Bold numbers show violations of regular minimality for Stimuli 2 and 8.

cross sections of the complete discrimination probability functions $\psi(x, y)$ for unidimensional stimuli (see Figures 5.5, 5.6, 5.7, and 5.8). A discrimination probability function "assigns to every ordered pair of stimuli (x, y) the probability $\psi(x, y)$ with which they are judged to be different" (Dzhafarov, 2003a, p. 184). One can only talk about a discrimination probability function when this function fulfills the property of regular minimality (see Section 3.4).

Tables 5.5, 5.6, 5.7, and 5.8 show that there are violations of regular minimality for all four subjects. Regular minimality is the fundamental empirical property data must fulfill in order to apply Fechnerian scaling (see Section 3.5). In oder to apply Fechnerian scaling, two assumptions must hold: (1) Regular minimality holds for all stimuli, and (2) we have true discrimination probabilities from which Fechnerian distances can be computed. When collecting data in a psychophysical experiment, one does not obtain true probabilities, but estimates of these probabilities that are bound to be subject to measurement error. Matrices of relative frequencies might therefore not satisfy regular minimality, even though the true discrimination probabilities are regular minimality compliant. Trendtel, Ünlü, and Dzhafarov (2010) and Dzhafarov, Ünlü,

Table 5.6: Discrimination Probabilities for Subject 2

	0	1	2	3	4	5	6	7	8
0	0.540	0.688	0.767	0.938	0.950	0.975	1.000	1.000	1.000
1	0.681	0.607	0.675	0.833	0.812	0.933	0.975	1.000	1.000
2	0.708	0.600	0.647	0.694	0.842	0.950	0.950	1.000	1.000
3	0.887	0.700	**0.644**	**0.653**	0.694	0.842	0.863	0.933	0.975
4	0.800	0.863	0.675	0.688	0.650	0.794	0.858	0.950	0.983
5	1.000	0.933	0.912	0.808	0.775	0.717	**0.719**	0.892	0.925
6	0.950	0.950	0.950	0.875	0.833	0.819	**0.747**	0.794	0.925
7	1.000	0.950	1.000	0.933	0.938	0.892	0.844	0.730	**0.750**
8	1.000	1.000	1.000	0.975	0.967	0.975	0.917	0.825	**0.757**

Note: Bold numbers show violations of regular minimality for Stimuli 2, 6, and 8.

Trendtel, and Colonius (2011) propose a method to test whether a certain number of violations of regular minimality for a given discrimination matrix is statistically significant. A discrimination matrix for nine stimuli can have between zero and nine violations of regular minimality (one possible violation for each row or column). The procedure described by Dzhafarov et al. (2011) tests the null hypothesis 'the matrix has no structure' against the alternative hypothesis 'regular minimality holds.' One can therefore determine if a given number of violations is statistically significant.

For Subject 1, there are two violations of regular minimality ($p = 0.019$) and the alternative hypothesis that data satisfy regular minimality can be accepted with an α level of 0.05. Subject 2 has three violations of regular minimality. All three violations are small and do not seem systematic, but the null hypothesis cannot be rejected since this corresponds to $p = 0.128$. Looking at the complete data set of Subject 3, there are three violations again. But when the stimulus space is reduced to 7 stimuli, taking into account this subject's bias to judge stimuli presented on the left to be darker, there is one

Table 5.7: Discrimination Probabilities for Subject 3

	0	1	2	3	4	5	6	7	8
0	**0.413**	0.725	0.867	0.938	0.983	0.975	1.000	1.000	1.000
1	**0.312**	0.420	0.562	0.833	0.875	0.983	1.000	1.000	1.000
2	**0.275**	0.394	0.470	0.769	0.867	0.975	0.967	1.000	1.000
3	0.388	0.342	0.369	0.517	0.656	0.875	0.912	0.983	1.000
4	0.500	0.425	0.283	0.425	0.530	0.794	0.875	0.950	0.967
5	0.750	0.567	0.450	0.383	**0.319**	0.580	0.700	0.900	0.912
6	0.950	0.675	0.450	0.287	**0.317**	0.456	0.570	0.819	0.942
7	1.000	0.950	0.775	0.683	0.438	0.383	0.450	0.650	0.756
8	1.000	1.000	0.850	0.600	0.583	0.425	**0.283**	**0.481**	**0.663**

Note: Bold numbers show violations of regular minimality. Subject shows strong bias (see text for details) which reduces data to a 7×7 matrix with one violation of regular minimality.

violation with seven stimuli. This allows us to reject the null hypothesis with $p = 0.009$. Subject 4 showed four violations out of nine possible ones. This corresponds to $p = 0.414$. Again, all violations are small and do not seem to be systematic. It becomes apparent that it is difficult to obtain data that satisfy regular minimality, even with stimuli as simple as the ones used in this experiment. The procedure introduced by Dzhafarov et al. (2011) only considers the overall number of violations, not their magnitude. Since all violations are small and do not seem to be systematic, it is assumed that they are all due to measurement error.

How to deal with these violations? They cannot be ignored, since regular minimality has to be satisfied for all stimuli. Assuming that violations found in this experiment are due to measurement error, it would be helpful to have predictions for the data that are regular minimality compliant for all stimuli and then perform Fechnerian scaling on these predicted data. Fitting a Quadrilateral Dissimilarity Model as introduced by Dzhafarov and Colonius

Table 5.8: Discrimination Probabilities for Subject 4

	0	1	2	3	4	5	6	7	8
0	0.557	0.656	0.700	0.838	0.933	0.975	1.000	1.000	1.000
1	0.725	**0.640**	**0.631**	0.742	0.800	0.900	0.950	1.000	1.000
2	0.833	0.700	**0.640**	0.719	0.717	0.838	0.933	0.950	1.000
3	0.950	0.900	0.775	0.703	0.744	0.792	0.812	0.900	0.975
4	0.950	0.912	0.900	**0.700**	**0.707**	0.731	0.792	0.912	0.967
5	1.000	0.983	0.963	0.892	0.856	0.697	**0.719**	0.767	0.800
6	1.000	0.975	0.933	0.938	0.917	0.781	**0.737**	0.750	0.783
7	1.000	1.000	1.000	0.983	0.975	0.917	0.869	0.740	**0.731**
8	1.000	1.000	0.950	1.000	0.983	0.950	0.858	0.775	**0.727**

Note: Bold numbers show violations of regular minimality for Stimuli 2 (or 1), 4, 6, and 8.

(2006b, Section 7.2) to the data gives us these predictions. The Quadrilateral Dissimilarity Model is based on the 'Uncertainty Blob Model' introduced by Dzhafarov (2003b) and in Section 3.4.

For our experiment, we assume that mutual PSEs mean that $x = y$, except for Subject 3 who showed a strong bias that was taken into account by fitting a model that has its PSEs for $y = x - 3$. The mean step size between stimuli was $1.5\,\frac{\mathrm{cd}}{\mathrm{m}^2}$ (see Table 5.1) and Subject 3 matched stimuli on the left with stimuli two steps darker on the right; this corresponds to a bias of $3\,\frac{\mathrm{cd}}{\mathrm{m}^2}$. The increasing function $\beta(s)$ in Equation 3.6 was chosen as $(1 + \exp(-\theta s - \eta))^{-1}$. Furthermore, Dzhafarov and Colonius's definitions for $D(a, b)$, $R_1(a)$, and $R_2(b)$ were adapted to

$$D(a, b) = \gamma |\log a - \log b|, \tag{5.5}$$

$$R_1(a) = \beta_0 + \beta_1 \log a + \beta_2 (\log a)^2, \quad \text{and} \tag{5.6}$$

$$R_2(b) = \beta_3 + \beta_4 \log b + \beta_5 (\log b)^2. \tag{5.7}$$

This model is used as a data approximation here, in order to deal with few

Table 5.9: Parameter Estimates for Quadrilateral Dissimilarity Model

	Subject 1	Subject 2	Subject 3	Subject 4
$\beta_0 + \beta_3$	0.958	0.416	-8.071	0.650
β_1	0.083	-0.959	86.315	-0.452
β_2	0.358	0.199	-11.501	0.673
γ	5.272	5.501	7.420	5.461
β_4	-0.300	-1.206	-78.671	-0.869
β_5	-0.296	0.514	10.698	-0.251
θ	1.668	1.238	1.300	1.283
η	-1.624	-2.271	-13.094	-1.435
Minima	175.762	109.804	143.743	95.629

violations of regular minimality, and does not necessarily have a theoretical interpretation in this context.

Figures 5.5, 5.6, 5.7, and 5.8 show the response surfaces predicted by the Quadrilateral Dissimilarity Model as described above and the observed response surfaces. Parameters were estimated by minimizing Pearson's

$$X^2 = \sum \frac{(y_{\text{diff}} - n\psi_{\text{pre}}(x, y|p))^2}{n\psi_{\text{pre}}(x, y|p)(1 - \psi_{\text{pre}}(x, y|p))} \tag{5.8}$$

using `nlm()` in R (version 3.0.1, R Core Team, 2013), where y_{diff} denotes the frequency to answer different, n denotes how often the pair was presented and $\psi_{\text{pre}}(x, y|p)$ denotes the probability to say different as predicted by the Quadrilateral Dissimilarity Model with parameter vector p. Table 5.9 shows parameter estimates for p. The fit of the Quadrilateral Dissimilarity Model was only fair for all subjects with $X_1^2(75) = 175.762$, $p \leq 0.001$; $X_2^2(75) = 109.804$, $p = 0.005$; $X_3^2(75) = 143.743$, $p \leq 0.001$; $X_4^2(75) = 95.629$, $p = 0.054$. Bootstrapping a sampling distribution assuming that each cell is independently binomially distributed, showed that the assumption of an underlying χ^2 distribution holds for these data. Since Figures 5.5, 5.6, 5.7, and 5.8 show that the model captures the qualitative pattern of the data very well for all subjects, we

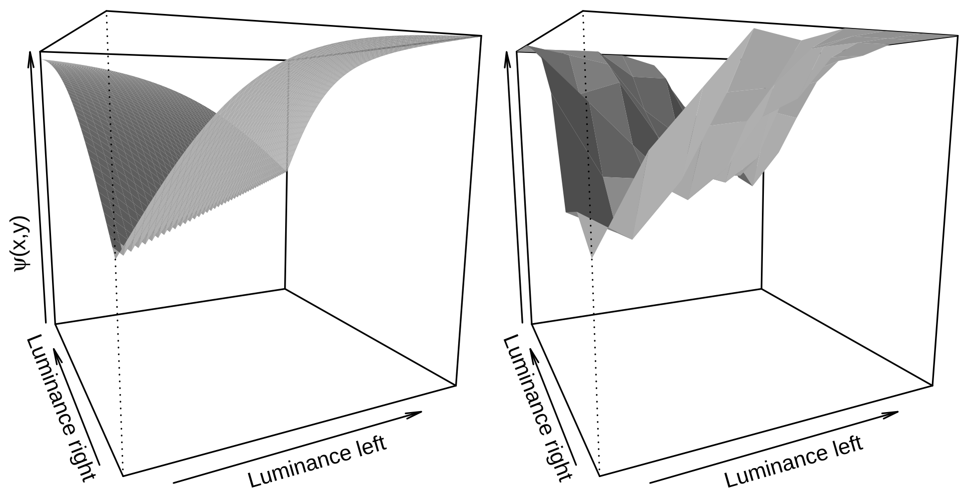


Figure 5.5: Discrimination probability function for Subject 1. Surface in the left plot shows fitted Quadrilateral Dissimilarity Model. The right plot shows the observed response surface.

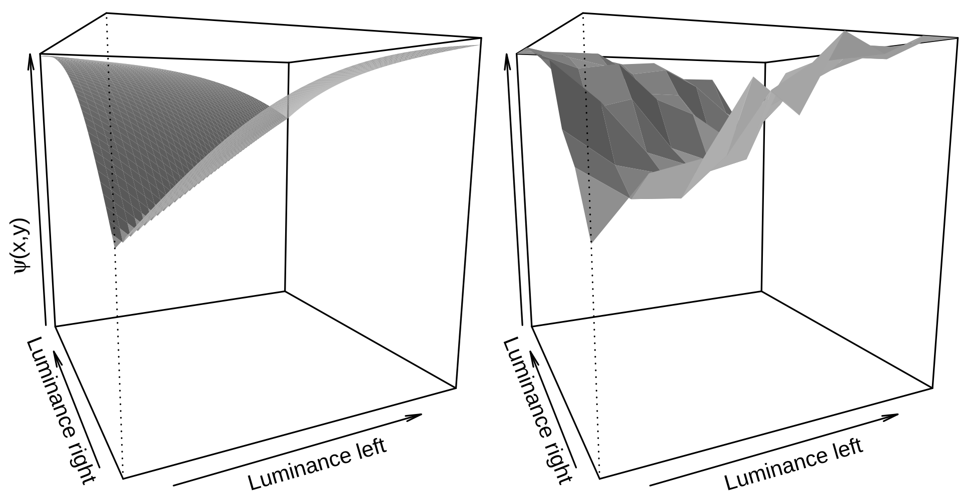


Figure 5.6: Discrimination probability function for Subject 2. See Figure 5.5 for details.

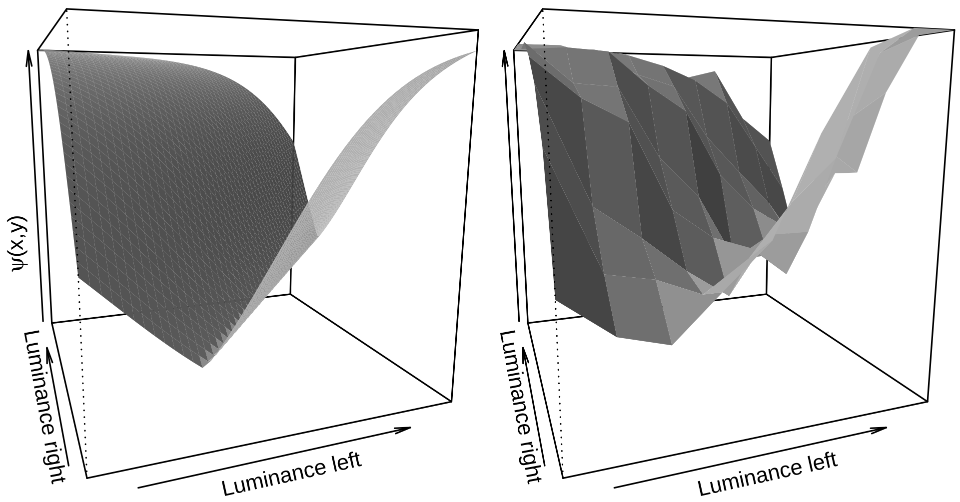


Figure 5.7: Discrimination probability function for Subject 3. See Figure 5.5 for details.

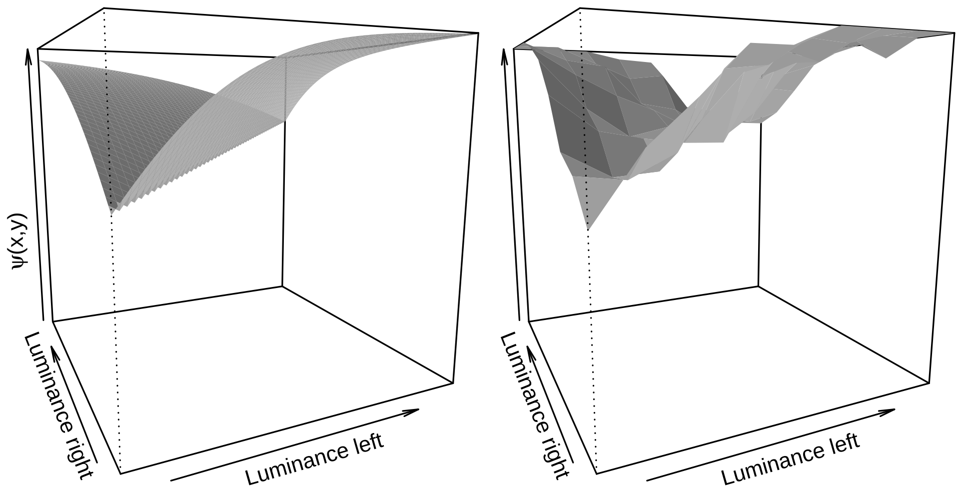
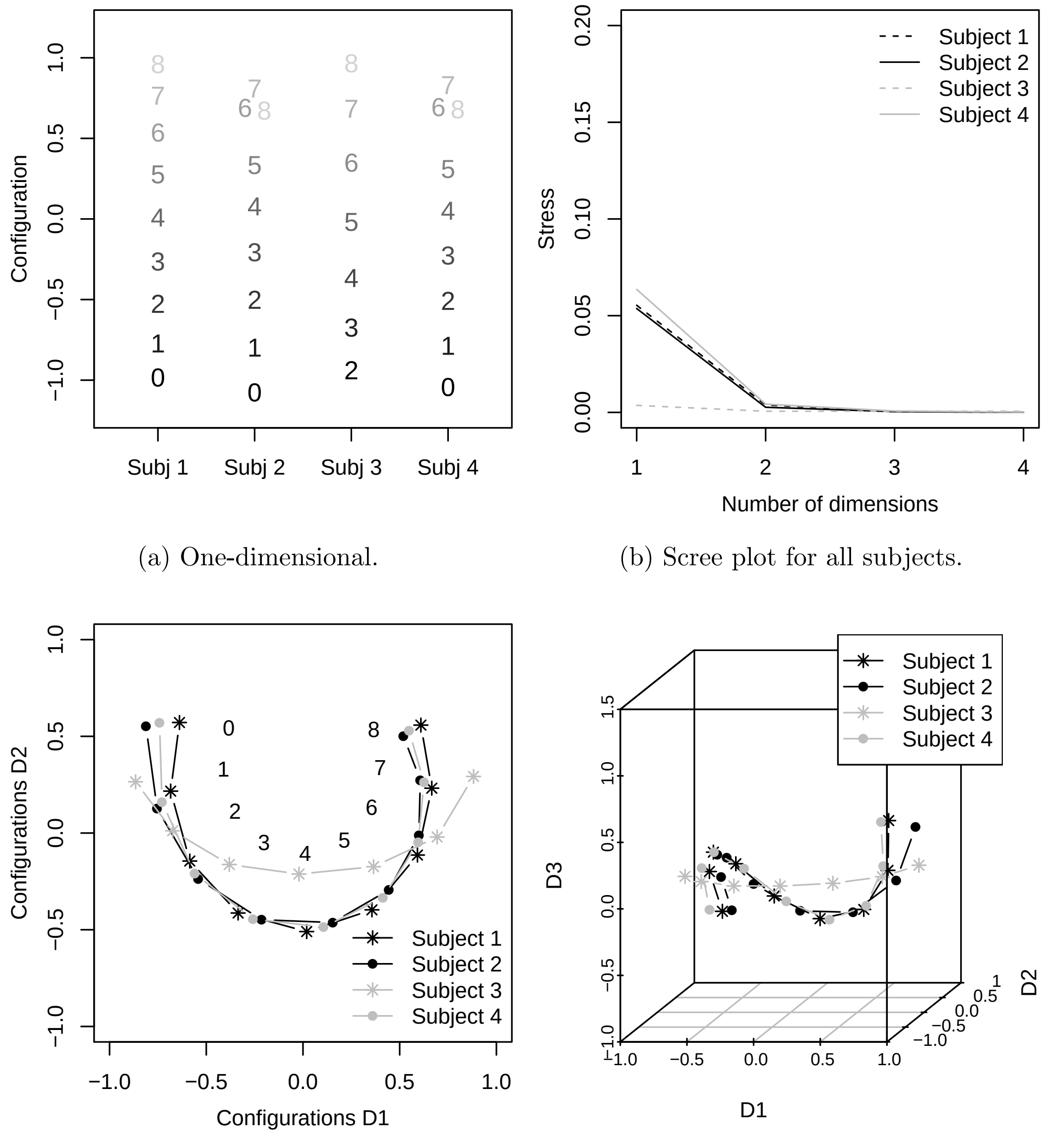


Figure 5.8: Discrimination probability function for Subject 4. See Figure 5.5 for details.

decided to use the predictions obtained with the model to deal with violations of regular minimality. Table 5.10 shows predicted discrimination probabilities and Table 5.11 the Fechnerian distances for these predictions.

Our research question was: How many dimensions are needed to represent Fechnerian distances found for these stimuli? It was hypothesized that subjects would need a single perceptual dimension to discriminate between gray patches presented under a constant illumination. Figure 5.9a shows one-dimensional multidimensional scaling (MDS) solutions for the Fechnerian distances for each subject. MDS solutions were calculated using the SMACOF algorithm (as explained by Borg & Groenen, 2005) to minimize the stress function (see Equation 3.17). Since Fechnerian distances are proper distances, a metric MDS approach (as implemented in the `SmacofSym()` function from the R package `smacof` by deLeeuw & Mair, 2009) was used.

Figure 5.9a shows that the stimuli can be easily arranged on one dimension from dark to light for all subjects. The stimuli were all very similar (compare Table 5.1). Especially the lighter stimuli (Stimuli 6, 7, and 8) lie close together for most subjects. Figure 5.9b shows scree plots for all subjects. Stress up to 0.2 is usually considered to be low. This is an arbitrary boundary and its interpretation is questionable. Nevertheless, for all subjects the stress is below 0.2 for a one-dimensional MDS solution. It stands out that Subject 3 has the lowest stress of all subjects. One should keep in mind, that this stress was calculated with two stimuli less than for the other subjects and this itself reduces stress (see Borg & Groenen, 2005, Chapter 3.5). However, calculating the stress for Subjects 1, 2, and 4 with a subset of 7 stimuli shows that the stress reduction that can be explained by having less stimuli is not as pronounced as the one for Subject 3 (see Figure A.1 in Appendix). In order to confirm that a one-dimensional solution represents the data best, the solutions for two- and three-dimensional MDS were considered. The results are depicted in Figures 5.9c and 5.9d and show a topologically one-dimensional structure, represented in two as well as in three dimensions.



(a) One-dimensional.

(b) Scree plot for all subjects.

(c) Two-dimensional.

(d) Three-dimensional.

Figure 5.9: MDS solutions for Fechnerian distances obtained from predictions of Quadrilateral Dissimilarity Model.

Table 5.10: Discrimination Probabilities Predicted with Quadrilateral Dissimilarity Model

		0	1	2	3	4	5	6	7	8
Subject 1	0	0.501	0.618	0.720	0.802	0.865	0.909	0.941	0.961	0.975
	1	0.686	0.505	0.619	0.719	0.802	0.864	0.909	0.940	0.961
	2	0.824	0.687	0.508	0.619	0.720	0.801	0.864	0.908	0.939
	3	0.909	0.822	0.686	0.511	0.624	0.721	0.804	0.865	0.909
	4	0.955	0.909	0.824	0.692	0.515	0.624	0.724	0.803	0.865
	5	0.978	0.954	0.908	0.825	0.690	0.518	0.630	0.726	0.805
	6	0.990	0.978	0.955	0.910	0.827	0.698	0.522	0.630	0.727
	7	0.995	0.990	0.978	0.955	0.910	0.830	0.697	0.525	0.634
	8	0.998	0.995	0.990	0.978	0.955	0.912	0.830	0.701	0.529
Subject 2	0	0.530	0.670	0.784	0.865	0.920	0.953	0.973	0.985	0.991
	1	0.650	0.562	0.697	0.803	0.879	0.927	0.958	0.976	0.986
	2	0.753	0.678	0.595	0.722	0.822	0.891	0.936	0.963	0.979
	3	0.831	0.773	0.704	0.626	0.749	0.840	0.904	0.943	0.967
	4	0.890	0.848	0.795	0.732	0.657	0.771	0.858	0.914	0.950
	5	0.928	0.900	0.862	0.815	0.755	0.686	0.797	0.874	0.925
	6	0.955	0.936	0.911	0.878	0.835	0.781	0.715	0.816	0.887
	7	0.972	0.960	0.943	0.921	0.891	0.852	0.802	0.742	0.836
	8	0.982	0.975	0.964	0.950	0.930	0.903	0.868	0.824	0.768
Subject 3	0	0.462	0.620	0.760	0.862	0.928	0.964	0.983	0.992	0.997
	1	0.340	0.494	0.655	0.789	0.885	0.942	0.973	0.987	0.995
	2	0.229	0.361	0.523	0.684	0.817	0.903	0.954	0.979	0.991
	3	0.351	0.242	0.383	0.550	0.716	0.840	0.921	0.963	0.983
	4	0.547	0.371	0.250	0.397	0.576	0.739	0.862	0.933	0.970
	5	0.717	0.553	0.387	0.257	0.417	0.599	0.767	0.880	0.944
	6	0.839	0.718	0.565	0.411	0.276	0.426	0.621	0.784	0.893
	7	0.910	0.831	0.716	0.574	0.425	0.298	0.444	0.640	0.804
	8	0.950	0.902	0.825	0.716	0.580	0.443	0.317	0.455	0.658
Subject 4	0	0.590	0.676	0.750	0.811	0.860	0.897	0.926	0.946	0.962
	1	0.742	0.609	0.692	0.762	0.821	0.867	0.903	0.930	0.949
	2	0.851	0.756	0.628	0.706	0.775	0.830	0.875	0.908	0.934
	3	0.918	0.858	0.767	0.646	0.724	0.788	0.842	0.883	0.914
	4	0.957	0.923	0.867	0.783	0.664	0.737	0.801	0.851	0.890
	5	0.977	0.959	0.927	0.876	0.794	0.682	0.754	0.813	0.861
	6	0.988	0.979	0.962	0.933	0.885	0.810	0.700	0.768	0.824
	7	0.994	0.989	0.980	0.965	0.937	0.893	0.820	0.717	0.782
	8	0.997	0.994	0.990	0.982	0.967	0.943	0.900	0.833	0.734

Note: Gray predictions for Subject 3 were excluded from analysis.

Table 5.11: Fechnerian Distances Obtained from Predicted Discrimination Probabilities

	1	2	3	4	5	6	7	8
0	0.298	0.536	0.698	0.804	0.868	0.907	0.930	0.942
1		0.293	0.525	0.691	0.795	0.861	0.899	0.922
2			0.286	0.522	0.683	0.789	0.853	0.892
3				0.290	0.517	0.681	0.783	0.847
4					0.281	0.515	0.673	0.776
5						0.288	0.512	0.670
6							0.280	0.506
7								0.281
0	0.228	0.413	0.541	0.623	0.665	0.683	0.684	0.676
1		0.218	0.388	0.508	0.579	0.617	0.631	0.630
2			0.205	0.366	0.472	0.537	0.569	0.580
3				0.199	0.343	0.441	0.496	0.523
4					0.183	0.320	0.406	0.454
5						0.177	0.298	0.374
6							0.161	0.272
7								0.149
2	0.241	0.503	0.780	1.074	1.285	1.357		
3		0.262	0.539	0.834	1.110	1.264		
4			0.277	0.571	0.848	1.110		
5				0.294	0.571	0.843		
6					0.276	0.548		
7						0.272		
0	0.219	0.383	0.493	0.562	0.602	0.624	0.634	0.635
1		0.210	0.365	0.471	0.535	0.573	0.593	0.601
2			0.200	0.351	0.448	0.510	0.544	0.562
3				0.197	0.336	0.429	0.485	0.517
4					0.185	0.321	0.407	0.459
5						0.183	0.307	0.388
6							0.171	0.290
7								0.165

Stimuli can be arranged meaningfully on one dimension. The original discrimination probabilities in Tables 5.5, 5.6, 5.7, and 5.8 show that regular minimality was violated for some of the lighter stimuli for all subjects. This seems to be mirrored by the Quadrilateral Dissimilarity Model and the Fechnerian distances.

5.1.4 Discussion

Greater-less judgments in this experiment were conducted for two reasons. First, it seemed necessary to check in a first session if subjects were able to discriminate between stimuli. Since all stimuli only differed by about 1.5 $\frac{\mathrm{cd}}{\mathrm{m}^2}$ (see Table 5.1), this seemed necessary in case some subjects would be less sensitive than others. All subjects showed good performance in this first session and stimuli did not have to be adjusted for individual subjects. Secondly, the last four sessions were conducted in order to check if any perceptual learning occurred over the course of the experiment. The results do not show any systematic shifts and we therefore conclude that perceptual learning was small or not present at all.

Our main focus was on the results from the same-different judgments. Data show that one perceptual dimension is needed to represent gray patches in a room with controlled illumination conditions. All subjects show a similar pattern and for all subjects the stimuli can be represented from dark to light. A one-dimensional representation of the Fechnerian distances appears to be an adequate representation for these data. In the two-dimensional MDS solutions for the data, the stimuli are arranged on circles for all subjects (see Figure 5.9c). These results are similar to the ones obtained by Izmailov and Sokolov (1991) and Logvinenko and Maloney (2006). Logvinenko and Maloney (2006) found three of these circles for different illuminations. Since illumination did not change here, it seems plausible to find only one of these circles for our data. Logvinenko and Maloney (2006, p. 80) note "[...] that for any fixed illuminant, the locus of achromatic colors for equi-illuminated surfaces forms the familiar one-dimensional continuum that we might think of as *lightness*."

Individual differences between subjects should not be neglected as can be seen in the same-different data of Subject 3. This subject has a strong bias to judge stimuli presented on the left to be darker than stimuli presented on the right. This bias also showed in her greater-less judgments (but not as pronounced). It is not possible to decide if this subject has a positional bias (Wickens, 2002, argues that two-alternative-forced choice tasks like the greater-less task are especially prone to positional bias) or indeed a perceptual bias (for an ongoing discussion of this topic, mostly focusing on separating decision and perceptual biases, see, Anton-Erxleben, Abrams, & Carrasco, 2010, 2011; García-Pérez & Alcalá-Quintana, 2013; Schneider, 2011; Schneider & Komlos, 2008). All subjects showed strong asymmetry in their discrimination probabilities, meaning that generally $\psi(x, y) \neq \psi(y, x)$, and nonconstant self-dissimilarity, $\psi(x_i, x_i) \neq \psi(x_j, x_j)$, for $i \neq j$. This is a common empirical result (Dzhafarov, 2002d) and emphasizes how important it is to introduce the notion of different observation areas. Aggregating results over different observation areas does not do the data justice in this case. Fechnerian scaling takes different observation areas into account and calculates distances within each observation area.

Figures 5.5, 5.6, 5.7, and 5.8 show that all subjects show strong response biases to answer different. This might in part be due to the fact that the paradigm itself forces subjects to answer different more often than same (see above and Table 5.2). But there are strong individual differences. Subject 4, e. g., tends to answer different in about 70 % of the trials where stimuli were physically identical. It seems therefore difficult to interpret these probabilities as the actual PSEs for these stimuli. A PSE in a greater-less paradigm is defined as the point where subjects answer greater with a probability of 0.5. For same-different judgments, the PSE is the minimum for each row and column (Dzhafarov & Colonius, 2006a). Two stimuli from different observation areas are judged to be different less often than when they are paired with all other stimuli in the corresponding observation areas. Dzhafarov and Colonius (2005a) show that the perceptual dissimilarities between stimuli are independent of any response bias (in the meaning of having a tendency to respond different) and that the Fechnerian distances are invariant up to multiplication of positive constants. This does not affect the dimensionality of the metric space these distances belong to. Thus, for our results it seems irrelevant if subjects showed strong response bias. The conclusions drawn from the results remain the same.

Same-different judgments are propagated to overcome many conceptual problems in psychophysics and especially in the investigation of color space (see, e. g., Niederée, 1998). But what is the best method to apply same-different judgments? Most experiments on color perception and color discrimination use matching as the task of choice. Problems with this approach have been addressed many times (see, e. g., Logvinenko & Maloney, 2006). Even though same-different judgments seem conceptually cleaner, they lead to many organizational and technical problems. Same-different judgments can only be applied to stimuli which are very similar in order to be meaningful. That implies subjects should have high uncertainty whether stimuli look same or different. But high uncertainty makes it necessary to collect a high number of trials, since it results in a large variance in the data. Having subjects judge gray patches for 15 hours has many disadvantages, and it is unclear if the conceptual advantages can really trump these more technical disadvantages. Motivational problems emerge, and it is very costly to collect 20 hours of data, then being unable to include them in the data analysis because subjects have stopped following instructions after some time. There do not seem to be systematic differences when analyzing our data separately for the first half (first 8 sessions) and the second half (sessions 10 to 16). Therefore, we assume that our subjects did indeed follow instructions throughout the course of the experiment. At the same time, this means that violations of regular minimality cannot simply be attributed to lack of motivation or fatigue.

Violations of regular minimality are a problem that proves to be unexpectedly pronounced for these kind of data. We did not expect to find as many violations as we did, especially since subjects' sensitivity was tested in the first session and proved satisfactory for all of them. Nonetheless, these data seem to support our hypothesis that the perceptual space for achromatic surface colors is one-dimensional when no context except for a uniform illumination is present.

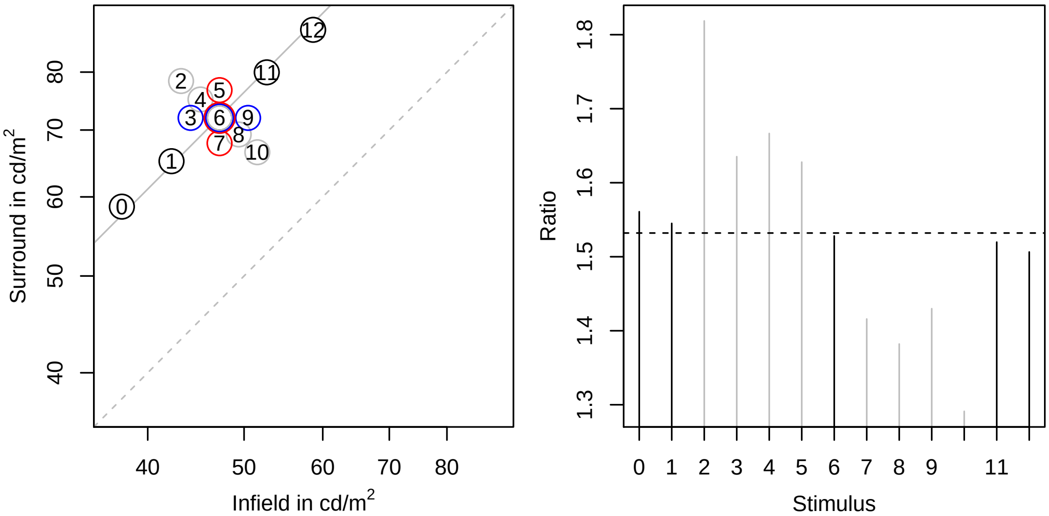
5.2 Experiment II: Infield-Surround Configurations

In the second experiment, stimuli were similar to the ones used in Experiment I except that surrounds were introduced to investigate if this kind of context influences the dimensionality of the perceptual color space of achromatic surface colors. Subjects performed same-different judgments on infield-surround stimuli presented in an illuminated room. Surrounds in this experiment are considered to be context effects since an interpretation of surrounds as illumination is not plausible in our experimental setting.

As explained in Section 2.2, many authors claim that colors should be investigated using infield-surround configurations since it is a comparatively natural setting. We usually perceive colors surrounded by other colors. In fact, without a surround or background, colors are usually not perceived as surface colors. Evans (1964) pointed this out early on and showed that this must be connected to the number of perceptual dimensions needed to discriminate between stimuli: "In general, in a normal environment, colored areas are seen surrounded by other colored areas. We need to inquire whether such a situation introduces any new perceptual variables" (p. 1468). Experiment II investigates whether surrounds introduce indeed extra perceptual dimensions.

5.2.1 Subjects

Four subjects participated in Experiment II. Three of these subjects were naïve as to the purpose of the experiment, the fourth subject was the author. Subjects were aged 20 to 30, three were female, Subject 4 was left-handed. All



(a) Infield-surround configurations. (b) Luminance ratios ($\frac{\text{Surround}}{\text{Infield}}$).

Figure 5.10: Stimuli used in Experiment II. Black circles have the same infield-surround ratio, blue circles have the same surround luminance, and red circles have the same infield luminance.

subjects had normal or corrected to normal vision and showed no color deficiency when tested with the test by Velhagen and Broschmann (1985).

5.2.2 Stimuli

Thirteen different infield-surround configurations were used. The layout of the different configurations can be seen in Figure 5.10a, and their luminance values are depicted in Table 5.12. All stimuli were decrements. For Stimuli 0, 1, 6, 11, and 12 (depicted in black), the ratio between infield and surround was (nearly) identical. The other stimuli differed in their ratios (see Figure 5.10b). Three stimuli (3, 6, and 9; blue) had the same surround with different infields, and three (5, 6, and 7; red) had the same infield with different surrounds.

Stimuli were rendered on a computer screen and presented through a cut-out in a white wall (see Figure 4.1). For Subjects 2 to 4, infields had a size of 0.81 degrees of visual angle and surrounds had a size of 4.58 degrees of visual angle. Two stimuli were presented side by side and separated by 0.4

Table 5.12: Luminance Values of Stimuli Used in Experiment II

Stimulus	Infields ($\frac{cd}{m^2}$)	Surrounds ($\frac{cd}{m^2}$)
0	37.79	58.98
1	42.41	65.52
2	43.37	78.88
3	44.34	72.51
4	45.39	75.64
5	47.45	77.24
6	**47.45**	**72.51**
7	47.45	67.19
8	49.62	68.57
9	50.71	72.51
10	51.82	66.90
11	52.98	80.51
12	58.98	88.85
Background	133.27	

Note: Stimulus 6 (bold) was used as standard in all graphs and was also the standard in Experiment III.

degrees of visual angle with a fixation cross of 0.05 degrees of visual angle and a luminance of 73.97 $\frac{\mathrm{cd}}{\mathrm{m}^2}$. Subjects were seated with a distance of 140 cm to the monitor (90 cm to the wall) with their heads secured by a chin rest with forehead support. For Subject 1 (the author) the stimuli were identical in luminance (see Table 5.12), but of different size. Infields had a size of about 1 degree of visual angle and surrounds were 2.96 degrees of visual angle.

5.2.3 Procedure

The procedure was similar to that in Experiment I. Subjects had to judge if two infields looked same or different. Stimuli were presented until subjects' response and for a maximum of 500 ms. Each stimulus pair was presented 60 times (except for Subject 1 for whom each pair was presented 60 times on average, see Appendix A.2 for details). Data was collected in 15 sessions on 15 different days. Each session consisted of 52 experimental blocks with 13 trials in each block. At the beginning of each session, two training blocks were presented, so that subjects would adapt to the illumination and memorize which button was same and which was different. Half of the subjects were instructed to press the left mouse button if infields looked the same and half were instructed to press the right mouse button if infields looked the same.

5.2.4 Results

Tables 5.13, 5.14, 5.15, and 5.16 show estimated discrimination probabilities for all subjects. Data for two subjects (1 and 4) had few violations of regular minimality. For Subjects 2 and 3 it is difficult to see a pattern in their data matrices (see Tables 5.14 and 5.15). Thus, there were severe violations of regular minimality for half of the subjects.

Table 5.13: Discrimination Probabilities for Subject 1

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.206	0.846	0.917	0.963	0.983	0.985	1.000	0.971	1.000	1.000	1.000	1.000	1.000
1	0.554	0.240	0.774	0.709	0.862	0.955	0.833	0.652	0.875	0.951	0.964	0.984	1.000
2	0.932	0.828	0.284	0.780	0.761	0.914	0.982	0.898	1.000	0.982	1.000	0.985	1.000
3	0.946	0.500	0.627	0.233	**0.188**	0.492	0.492	0.644	0.800	0.881	0.984	0.941	1.000
4	0.979	0.717	0.475	0.298	**0.291**	0.444	0.619	0.823	0.833	0.872	0.961	0.828	1.000
5	0.982	0.889	0.610	0.490	**0.226**	**0.295**	0.519	0.774	0.729	0.738	0.964	0.723	1.000
6	0.971	0.603	0.851	0.455	0.600	0.383	0.210	0.259	0.339	0.500	0.787	0.622	1.000
7	0.979	0.596	0.848	0.706	0.830	0.800	0.478	0.088	0.388	0.456	0.508	0.946	0.985
8	0.957	0.780	0.964	0.851	0.877	0.716	0.434	0.116	0.024	0.183	0.204	0.785	0.945
9	0.977	0.879	0.959	0.896	0.859	0.694	0.317	0.350	0.135	0.065	0.278	0.556	0.932
10	1.000	0.939	0.981	0.952	0.938	0.969	0.857	0.500	0.320	0.393	0.027	0.970	0.985
11	1.000	0.967	0.955	0.902	0.839	0.698	0.833	0.935	0.825	0.672	0.848	0.204	0.692
12	1.000	1.000	1.000	0.986	1.000	0.986	0.926	0.979	0.985	0.905	0.918	0.791	0.270

Note: Bold numbers show violations of regular minimality for Stimuli 4 and 5.

Table 5.14: Discrimination Probabilities for Subject 2

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.483	0.317	0.367	0.400	0.400	0.467	0.600	0.550	0.583	0.583	0.600	0.783	0.800
1	0.800	0.533	0.667	0.433	0.450	0.433	0.350	0.350	0.350	0.483	0.483	0.700	0.683
2	0.850	0.750	0.600	0.550	0.483	0.483	0.467	0.633	0.367	0.483	0.417	0.400	0.850
3	0.833	0.783	0.633	0.567	0.533	0.450	0.433	0.433	0.433	0.350	0.400	0.583	0.633
4	0.933	0.683	0.700	0.567	0.583	0.450	0.517	0.400	0.550	0.400	0.533	0.367	0.717
5	0.950	0.867	0.800	0.767	0.717	0.467	0.467	0.550	0.417	0.350	0.483	0.367	0.583
6	0.917	0.783	0.683	0.667	0.633	0.467	0.467	0.483	0.317	0.367	0.333	0.383	0.533
7	0.883	0.800	0.733	0.633	0.667	0.567	0.283	0.350	0.383	0.233	0.300	0.400	0.567
8	0.883	0.800	0.883	0.817	0.783	0.683	0.467	0.383	0.333	0.383	0.267	0.267	0.450
9	0.900	0.900	0.833	0.850	0.800	0.783	0.600	0.667	0.533	0.467	0.567	0.367	0.333
10	0.883	0.833	0.850	0.733	0.800	0.683	0.633	0.483	0.317	0.400	0.217	0.417	0.417
11	0.883	0.883	0.867	0.850	0.867	0.833	0.817	0.850	0.733	0.667	0.567	0.433	0.417
12	0.967	0.900	0.950	0.900	0.817	0.883	0.950	0.833	0.833	0.800	0.783	0.850	0.517

Note: Violations of regular minimality seem severe. There is no visible structure in the data.

Table 5.15: Discrimination Probabilities for Subject 3

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.483	0.633	0.700	0.817	0.867	0.900	0.883	0.883	0.900	0.933	0.767	1.000	1.000
1	0.667	0.667	0.717	0.750	0.867	0.900	0.883	0.717	0.783	0.900	0.683	0.983	0.900
2	0.733	0.667	0.500	0.717	0.600	0.700	0.667	0.783	0.683	0.800	0.717	0.933	0.983
3	0.867	0.700	0.600	0.667	0.617	0.700	0.750	0.733	0.683	0.800	0.767	0.883	0.967
4	0.767	0.733	0.717	0.700	0.617	0.767	0.683	0.817	0.717	0.783	0.650	0.917	0.967
5	0.900	0.833	0.700	0.800	0.650	0.733	0.750	0.700	0.683	0.750	0.717	0.917	0.933
6	0.783	0.683	0.833	0.650	0.750	0.800	0.600	0.717	0.683	0.767	0.800	0.917	0.917
7	0.883	0.633	0.767	0.800	0.683	0.850	0.667	0.667	0.650	0.733	0.633	0.917	0.983
8	0.767	0.883	0.800	0.783	0.683	0.833	0.767	0.533	0.700	0.783	0.733	0.883	0.983
9	0.983	0.883	0.850	0.767	0.700	0.750	0.633	0.750	0.783	0.617	0.767	0.883	0.967
10	0.850	0.650	0.767	0.783	0.733	0.833	0.733	0.633	0.650	0.717	0.550	0.800	0.967
11	0.933	0.900	0.850	0.833	0.883	0.750	0.800	0.917	0.750	0.900	0.800	0.833	0.933
12	0.900	0.917	0.883	0.900	0.883	0.850	0.850	0.900	0.933	0.883	0.867	0.783	0.817

Note: Violations of regular minimality are not bold, since there are too many (8) to see a consistent pattern.

Table 5.16: Discrimination Probabilities for Subject 4

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.117	0.267	0.567	0.633	0.750	0.883	0.950	0.850	0.967	0.983	0.900	1.000	1.000
1	0.450	0.100	0.550	0.333	0.567	0.650	0.433	0.400	0.483	0.750	0.400	1.000	0.983
2	0.433	0.350	0.150	0.367	0.283	0.467	0.800	0.767	0.817	0.933	0.883	0.933	1.000
3	0.700	0.217	0.483	**0.217**	0.183	0.300	0.183	0.250	0.550	0.650	0.717	0.900	1.000
4	0.683	0.467	0.317	**0.183**	0.133	0.367	0.283	0.383	0.567	0.633	0.700	0.833	0.983
5	0.900	0.650	0.633	0.283	0.333	0.133	**0.167**	0.500	0.433	0.550	0.650	0.500	0.967
6	0.967	0.650	0.883	0.567	0.467	0.383	**0.217**	0.200	0.200	0.183	0.200	0.550	0.967
7	0.950	0.567	0.917	0.650	0.617	0.583	0.450	0.083	0.133	0.183	0.200	0.850	0.967
8	1.000	0.750	0.967	0.783	0.783	0.700	0.400	0.150	**0.167**	**0.117**	0.100	0.767	0.933
9	1.000	0.933	0.917	0.800	0.833	0.700	0.467	0.367	**0.117**	**0.233**	0.200	0.517	0.850
10	0.933	0.783	0.917	0.850	0.933	0.800	0.683	0.267	0.233	0.300	0.100	0.683	0.850
11	0.983	0.983	0.950	0.933	0.833	0.733	0.700	0.667	0.550	0.533	0.517	0.217	0.667
12	1.000	0.967	1.000	1.000	1.000	0.983	0.967	0.950	0.883	0.867	0.767	0.683	0.217

Note: Bold numbers show violations of regular minimality for Stimuli 3, 6, 8, and 9.

For the more complex stimuli used in this experiment, it might not be as clear how to interpret regular minimality than it was for the simple gray patches of Experiment I. In Experiment II, subjects had to judge if *infields* were same or different. This might imply that regular minimality does not need to be satisfied in a canonical form (meaning that minima are on the diagonal, see Dzhafarov & Colonius, 2006b). However, it seems plausible to assume that the probability to judge two stimuli to be different reaches its minimum when infield *and* surround are both identical. More precisely, to judge stimuli same should be most probable when all physical properties (except presentation side) are identical. This leads again to the assumption that regular minimality should hold in canonical form.

Some violations of regular minimality that could be explained in terms of statistical fluctuation as was done in Experiment I were expected. For Subjects 1 and 4 this assumption holds for the current experiment. Subject 1 had two violations of regular minimality with thirteen stimuli which means the alternative hypothesis that data are regular minimality compliant can be accepted with $p \leq 0.001$. Subject 4 had four violations ($p = 0.041$) and Subject 3 had eight violations ($p = 0.913$). The violations for Subject 3 are severe and the alternative hypothesis that regular minimality is satisfied does not hold. For Subject 2 it is not possible to count the number of violations since there is no visible structure. According to Dzhafarov (2002d), one should conclude from such data as of Subject 2 that there is no underlying discrimination probability function.

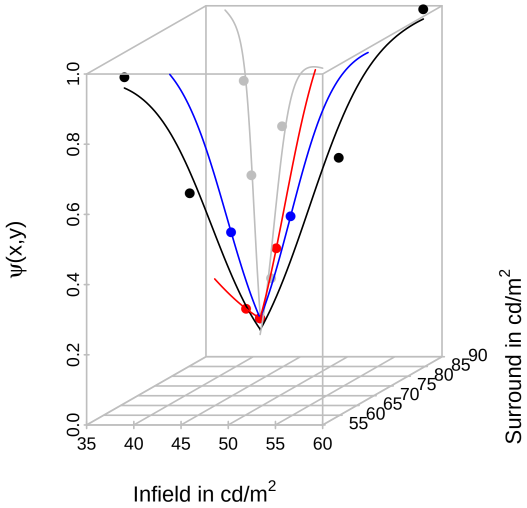
For the current data, we could not deal with violations of regular minimality in the same way as in Experiment I. Since there are two underlying physical dimensions (luminance of the infield and luminance of the surround), the discrimination probability function for these data needs to be at least five-dimensional. The Quadrilateral Dissimilarity Model can only be used to make predictions for one physical dimension varied in two observation areas. The approach adopted here is a simple statistical procedure and does not have an underlying theoretical foundation (like the predictions obtained with the

Quadrilateral Dissimilarity Model). It is assumed that matrices exist which satisfy regular minimality and then tested if these matrices significantly differ from the original data matrices.

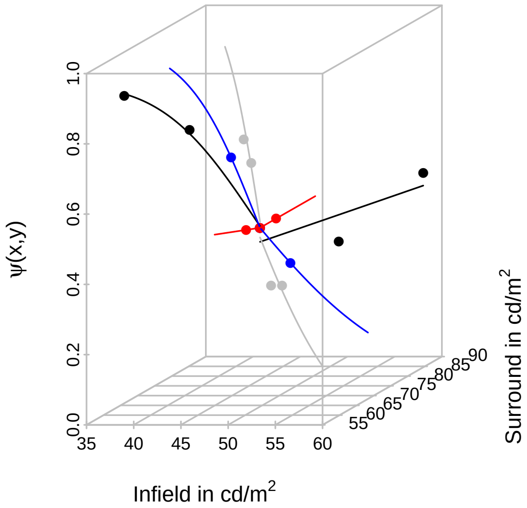
In order to find matrices that satisfy regular minimality, the violations for each row and column were considered and different responses were added or subtracted until regular minimality was satisfied for this row and column. Then the next cell was considered and so on. For example, if the violation was due to the fact that $a_{34} = 16$ and $a_{33} = 24$, the mean for both cells $\frac{a_{34}+a_{33}}{2} = 20$ was taken and then these cells were set to $a_{34} = 20$ and $a_{33} = 19$. Cells were adjusted so that regular minimality would hold in canonical form for reasons explained above.

The discrimination probabilities for the corrected matrices can be found in Tables A.3, A.4, and A.5 in the Appendix. Whether these tables differed significantly from the original data matrices was tested by calculating Pearson's X^2 (similar to the procedure in Experiment I, see Equation 5.8). For Subject 1, this resulted in $X_1^2(4) = 4.164,\ p = 0.384$. One degree of freedom was taken for each cell that was changed. For Subject 3, this test resulted in $X_3^2(17) = 3.917,\ p = 0.999$ and for Subject 4 in $X_4^2(11) = 8.710,\ p = 0.649$. This shows that the violations seem to be small and that merely counting the violations, like the approach by Dzhafarov et al. (2011) does, is not the best procedure to evaluate if regular minimality holds for a given matrix. For Subject 2, it was impossible to find a matrix that was still similar to the original data matrix. Data of this subject will not be interpreted any further. There could be a number of reasons why regular minimality does not hold for this subject, the most probable being motivational issues.

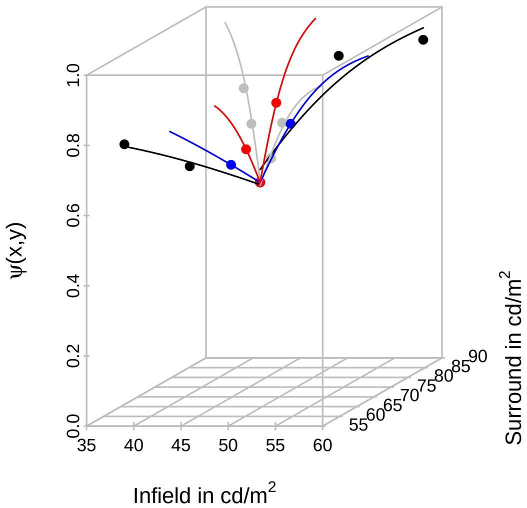
Figure 5.11 shows two-dimensional psychometric functions for all subjects when Stimulus 6 is taken as the standard and presented on the left side. Colors are the same as in Figure 5.10, i. e., the black line shows data for stimuli that had the same infield-surround ratio, the red line corresponds to stimuli which had the same infield luminance, and the blue line shows data for stimuli that had the same surround lumiance. For Subject 4, data for the corrected



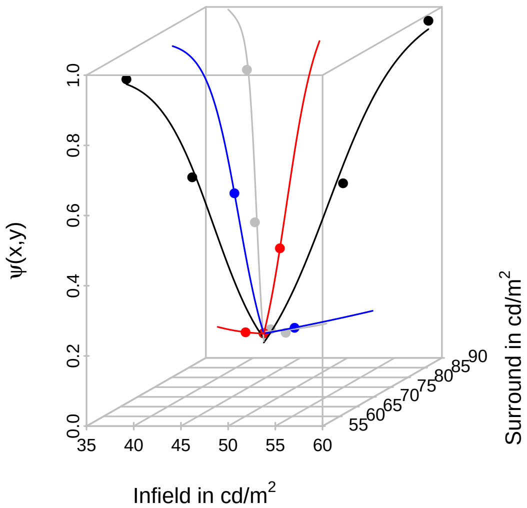
(a) Subject 1.



(b) Subject 2.



(c) Subject 3.



(d) Subject 4.

Figure 5.11: Two-dimensional psychometric functions. Standard is Stimulus 6 (see Table 5.12) presented on the left side. Colors are explained in Figure 5.10.

data matrix is shown (since this subject had one of its few regular minimality violations for Stimulus 6). Data of Subject 2 shows that regular minimality is strongly violated (see Figure 5.11b). The psychometric functions for the other subjects show how data look when regular minimality is satisfied in canonical form: With a global minimum for the standard. The shape of the psychometric functions obtained in this second experiment imply a two dimensional structure of the underlying perceptual space (Dzhafarov & Colonius, 1999) since the probability to answer different increases in every direction of the stimulus space (more details are discussed in Section 5.2.5).

The Fechnerian distances calculated from the corrected discrimination probabilities can be found in Tables A.6, A.7, and A.8 in the Appendix. A metric MDS was performed on the Fechnerian distances in order to find the dimensionality of these subjective (perceptual) distances. The MDS was performed with the distance smoothing optimization procedure as described in Groenen, Heiser, and Meulman (1999).

One-Dimensional MDS

Performing a one-dimensional metric MDS on the Fechnerian distances shown in Tables A.6, A.7, and A.8 results in the configurations shown in Figure 5.12. The shading of the stimulus names corresponds to the luminance of the infields. Infields of the Stimuli 5, 6, and 7 were identical (cf. Figure 5.10a). For Subjects 1 and 4, the one-dimensional MDS solution looks similar and seems in good accordance with the luminance of the infields. For Subject 3, this can still be assumed when considering that the number of violations for this subject was larger. The MDS solutions show high correlations with the log luminance of the infield (see Table 5.17) implying that this might be the relevant physical change in the stimuli that resulted in the lightness perceived by the subjects. The correlations of this dimension with log luminance of surround and the log ratio between luminance of infield and luminance of surround are not as pronounced.

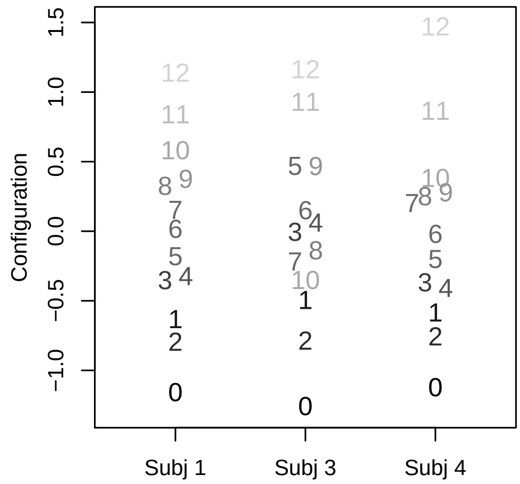


Figure 5.12: One-dimensional MDS solutions for Fechnerian distances for Subjects 1, 3, and 4.

Table 5.17: Correlations of One-Dimensional MDS Solution with log Luminance of Infield, log Luminance of Surround, and log Infield-Surround Ratio

		$\hat{\rho}$	95 % CI	
Subject 1	infield	**0.978**	0.926	0.994
	surround	0.543	-0.011	0.842
	$\frac{\text{infield}}{\text{surround}}$	0.594	0.063	0.863
Subject 3	infield	**0.843**	0.546	0.952
	surround	0.786	0.416	0.933
	$\frac{\text{infield}}{\text{surround}}$	0.136	-0.448	0.639
Subject 4	infield	**0.972**	0.905	0.992
	surround	0.601	0.075	0.866
	$\frac{\text{infield}}{\text{surround}}$	0.516	-0.048	0.831

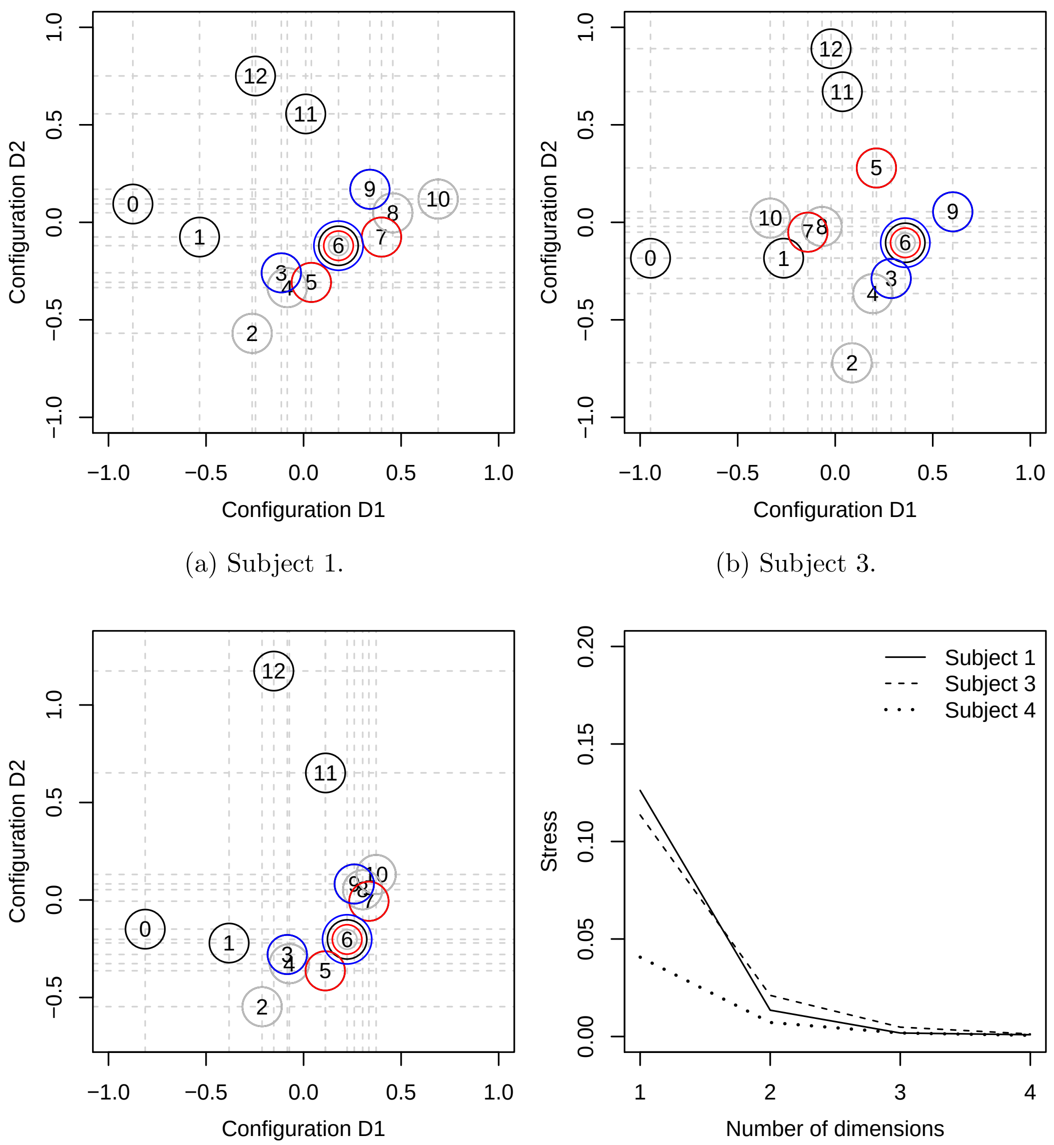
Two-Dimensional MDS

It was hypothesized that the surround of the stimuli in Experiment II would introduce a second perceptual dimension. Figure 5.13d shows scree plots for the three subjects. As discussed before, scree plots are of little help when deciding how many dimensions are needed to arrange the Fechnerian distances. Although the stress looks higher than in Experiment I, this is probably due to the fact that there are 13 stimuli instead of 9. Therefore, as in Experiment I, the arrangements of the distances in higher dimensional plots were considered. If the distances can be arranged in a one-dimensional perceptual space, it would be expected to find stimuli forming a line in two-dimensional space.

Logvinenko and Maloney (2006) found that distances obtained with their Maximum Likelihood Parametric Scaling (MLPS) could best be represented with a City-block metric (Minkowski metric with an exponent $p = 1$, see Equation 3.16). They used aggregated data to estimate their model parameters. Ronacher and Bautz (1985) found individual differences ranging from a City-block to a Euclidean distance with stimuli differing in size and lightness (see Section 3.6). Our data do not allow us to draw conclusions about the underlying metric of the distances by comparing the stress calculated for different metrics. Computing stress for different values of p showed artifactual patterns as described by Shepard (1974).[1] Therefore, the results will not be shown here. Applying MLPS to our data is also not possible since the stimulus situation used by Logvinenko and Maloney (2006) differs too much from our situation. Even though their stimuli were presented side by side as in our experiment, they changed the illumination above one half of the display what results in a stimulus situation completely different from ours. Adjusting their model to fit our needs did not prove to be fruitful either.

In order to obtain an MDS solution, we will assume a City-block metric here, leaning on evidence found in the literature. In any case, the resulting con-

[1] "[. . .] while the finding that the lowest stress is attainable for $r = 2$ may be evidence that the underlying metric is Euclidean, the finding that a lower stress is attainable for a value of r that is much smaller or larger may be artifactual." (Shepard, 1974, p. 404)



(a) Subject 1.

(b) Subject 3.

(c) Subject 4.

(d) Scree plot.

Figure 5.13: Two-dimensional MDS solutions for Fechnerian distances with City-block distances and scree plot. Colors are explained in Figure 5.10.

Table 5.18: Correlations of Two-Dimensional MDS Solutions with log Luminance of Infield, log Luminance of Surround, and log Infield-Surround Ratio

		Dimension 1			Dimension 2		
		$\hat{\rho}$	95 % CI		$\hat{\rho}$	95 % CI	
Subject 1	infield	0.596	0.067	0.864	**0.656**	0.164	0.887
	surround	0.121	-0.460	0.630	0.246	-0.352	0.702
	$\frac{\text{infield}}{\text{surround}}$	**0.628**	0.118	0.876	0.557	-0.008	0.848
Subject 3	infield	0.417	-0.174	0.787	**0.761**	0.362	0.924
	surround	**0.600**	0.073	0.865	0.477	-0.099	0.814
	$\frac{\text{infield}}{\text{surround}}$	0.177	-0.414	0.663	0.417	-0.175	0.787
Subject 4	infield	**0.652**	0.158	0.885	**0.781**	0.404	0.931
	surround	0.236	-0.337	0.711	0.460	-0.121	0.807
	$\frac{\text{infield}}{\text{surround}}$	0.531	-0.028	0.837	0.462	-0.118	0.808

figurations do not change much for different metrics (Euclidean, Dominance, and City-block).

Figure 5.13 shows the two-dimensional MDS solutions for Subjects 1, 3, and 4. MDS dimensions and their signs were adjusted so that result patterns looked qualitatively similar. The two-dimensional MDS solutions for Experiment II do not appear to be one-dimensional (like they did for Experiment I). In order to interpret the two dimensions it was assumed that they are associated with luminance of infield, luminance of surround or the ratio between luminance of infield and luminance of surround. The ratio between infield and surround has long been hypothesized to influence the perception of lightness (see, e. g., Gilchrist, 2006). It is usually referred to as contrast. Table 5.18 shows how these physical dimensions correlate with the MDS configurations. Apparently, it is difficult to interpret the pattern of these correlations. For the second dimension, the log luminance of the infield is the only significant correlation, but the first dimension does not show a consistent pattern.

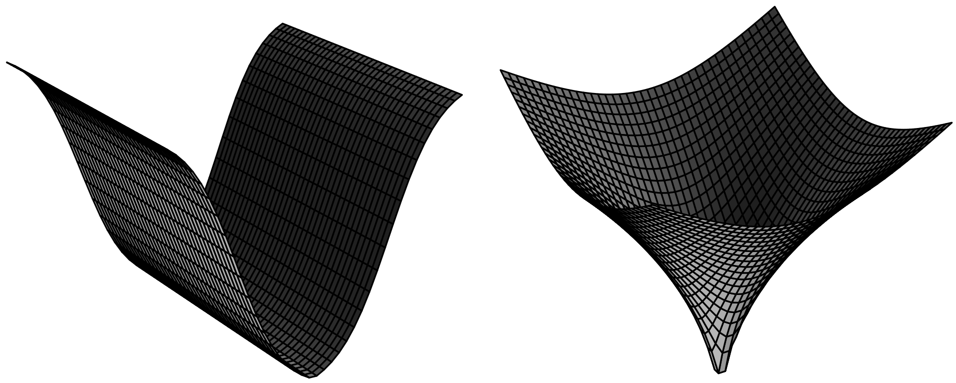
The relevant dimension does not seem to be infield-surround ratio here. This would have been expected according to the literature on infield-surround configurations (see, e. g., Gilchrist, 2006; Jacobsen & Gilchrist, 1988; Wallach, 1948). The relevant dimension influencing lightness of the infield seems to be luminance of the infield. This is termed luminance matching in the literature and is usually associated with increments (Gilchrist, 2006). What role the surround plays is unclear. Subjects do not seem to discount the background (Walraven, 1976) or apply the ratio principle (Wallach, 1948). Stimuli can best be arranged when taking into account the lightness of their infields and not the ratio between infield and surround (cf. Figure 5.10b). The correlations with the second dimension in Table 5.18 seem to correspond to the correlations for the one-dimensional MDS shown in Table 5.17.

5.2.5 Discussion

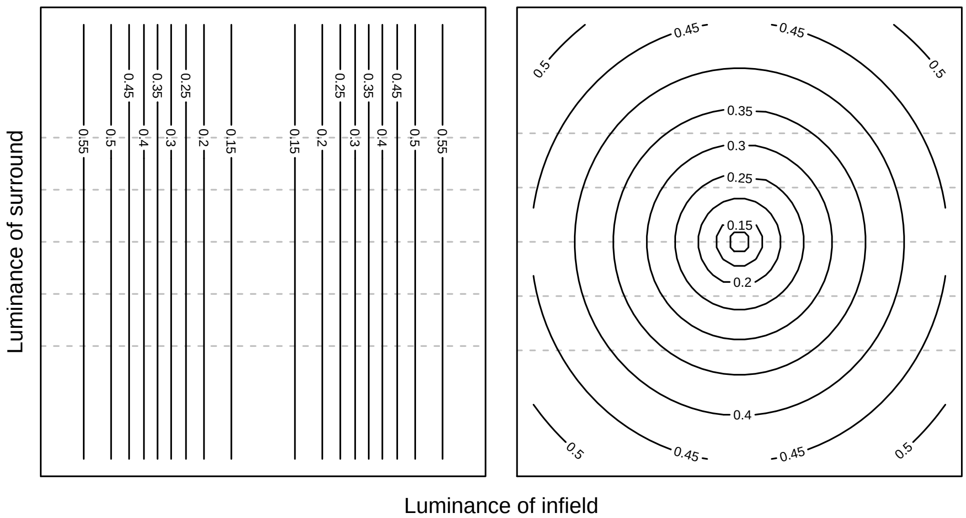
The results of Experiment II suggest that there might be a second relevant perceptual dimension for discriminating infield-surround configurations. However, the results are difficult to interpret. The scree plot for the MDS solutions does not suggest that a second perceptual dimension is needed (see Figure 5.13d), and the one-dimensional MDS solutions can be readily interpreted. Lightness of the infield seems to be the relevant perceptual dimension. However, the two-dimensional MDS solutions do not look like there is only one underlying perceptual dimension. Logvinenko and Maloney (2006, p. 83) "[...] found that perceived dissimilarity judgment is, effectively, the weighted combination of light and surface *cues to dissimilarity.*" We would have expected something along these lines for our data as well. The lightness of the infields should have been affected by the luminance of the infield as well as the luminance of the surround or at least the ratio between these two. But for three subjects, there are three different result patterns (see Table 5.18). The only consistent result is that the luminance of the infield influences the lightness of the infield. Logvinenko and Maloney (2006) changed the perceived illumination of their displays. It seems as if that influences lightness perception in a different way than changing local context effects like surrounds. They found that illumination and lightness of the stimuli both influence how similar they are judged. The surround does not seem to influence the lightness of the infield in a similar way, further supporting that it is misleading to consider surrounds as illumination.

The shape of the two-dimensional psychometric functions obtained in this experiment is closely related to the underlying perceptual dimensionality (Dzhafarov & Colonius, 1999). When a single standard is compared to stimuli varying on two dimensions (e. g., luminance of infield and luminance of surround), we obtain global minima of $\psi(x, y)$ that form a $(n-1)$-dimensional hypersurface (i. e., a line for $n = 2$, Dzhafarov & Colonius, 1999, p. 260). When only a single perceptual dimension is needed to discriminate between stimuli in a two-dimensional stimulus space, these minima are constant (and the psychometric function forms a valley). When we need two perceptual dimensions the minima for $\psi(x, y)$ are not constant (the psychometric function looks tulip-shaped), i. e., the line consists of minima of different magnitude for different points in the stimulus space. Figure 5.14 shows theoretical psychometric functions for a two-dimensional stimulus space for both of these cases.

Let us consider the case of a two-dimensional psychometric function which has only one underlying perceptual dimension. It needs to have a constant minimum that forms a line in the stimulus space. If one considers, for example, a situation where the surround does not influence the perception of the infield, it would be expected that the two-dimensional psychometric function has a constant minimum forming a line independent of different luminance values of the surround. The left side of Figure 5.14b shows equal probability curves when discrimination probabilities depend on the luminance of the infield only. If the lightness of the infield only depends on the luminance of the infield, lines of the contour plot have the same height when luminance of the infield is the same. On the right side, a contour plot is shown where lightness of the infield



(a) Theoretical psychometric functions.



(b) Equal probability curves for functions in upper panels.

Figure 5.14: Theoretical psychometric functions for a two-dimensional stimulus space with one (left) or two (right) underlying perceptual dimensions. Dotted lines in lower panel indicate different surrounds.

is influenced by the surround. The probability to say different increases in every direction from the (global) minimum. Figure 5.14 shows the two most extreme cases. The global form of the psychometric function could also lie between these two extremes.

In the current experiment, psychometric functions that look tulip shaped were found which would suggest that more than one perceptual dimension is needed. However, our stimuli were spaced rather widely, and it might be possible that data for the relevant stimuli that form a constant minima line were not collected. The form of the two-dimensional psychometric functions is crucial when determining how many perceptual dimensions are needed to discriminate between infields of infield-surround configurations. Therefore, a third experiment was conducted focusing on size and location of the minima of psychometric functions for infield-surround configurations obtained with same-different judgments.

5.3 Experiment III: Psychometric Functions

The question how many perceptual dimensions a perceptual space for achromatic infield-surround stimuli has, could not be answered satisfactorily in Experiment II. The picture given by the spatial arrangement of the Fechnerian distances by the three subjects is ambiguous at best. Therefore, it seemed necessary to take a step back and look at the shape of the underlying two-dimensional psychometric functions. Psychometric functions obtained in Experiment II suggest that the psychometric functions are tulip shaped. But the stimuli were rather widely spaced (compare Figure 5.15, gray data points). It is possible that we did not 'hit' the right stimuli and that the constant minima lie between stimuli that were chosen for Experiment II. Therefore, Experiment III investigates the shape of psychometric functions obtained with same-different judgments for infield-surround stimuli in a finer grid of luminance values.

An adaptive procedure to determine minima for the psychometric functions was developed. Figure 5.15 shows the starting configuration for the third ex-

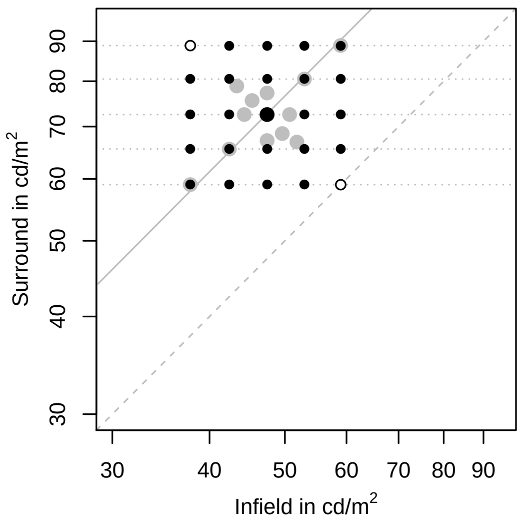


Figure 5.15: Stimulus configurations for pre-session of Experiment III. Gray dots show stimuli used in Experiment II (see also Figure 5.10). Big black dot is the standard with an infield of 47.45 $\frac{\mathrm{cd}}{\mathrm{m}^2}$ and a surround of 72.51 $\frac{\mathrm{cd}}{\mathrm{m}^2}$. No data was collected for open circles.

periment. The luminance for five different surrounds was fixed and subjects had to compare different infields for those surrounds to one standard configuration. It was expected that the global minimum for the psychometric function would be obtained when comparing the standard to an infield which is placed within a surround with the same luminance as the standard. When only one perceptual dimension is needed, a line with constant minima for all surrounds should exist (Dzhafarov & Colonius, 1999) and the minima for $\psi(x, y_j)$, with $j = 1 \ldots 5$, should all have the same magnitude. In contrast, when the minima for the other surrounds are higher than the one for the standard surround, it would support that two perceptual dimensions are needed to discriminate between stimuli with different surrounds (see Figure 5.14).

Since Experiment III aims at characterizing the form of the psychometric function and not at finding distances between stimuli, each subject was presented with the standard fixed in one observation area (left or right). It is not assumed that the psychometric functions for both observation areas are identical. However, it is assumed that the global (overall) shape of these functions should be similar. Therefore, the standard was presented in only one observation area for each subject in order to get more repetitions for each trial and have a more accurate picture of what the psychometric function for one observation area looks like.

5.3.1 Subjects

Three subjects participated in the third experiment. Two of the subjects were female, all were 20 years old, and naïve as to the purpose of the experiment. They all had normal or corrected to normal vision and were tested for color deficiencies with the Ishihara Test (Ishihara, 2012) which showed no abnormalities. Subject 3 was left handed.

5.3.2 Stimuli and Procedure

The procedure was identical to that in Experiment II. Stimuli had the same size and were presented in the same way. Stimulus 6 in Experiment II was taken as the standard in the current experiment. In Experiment II, all pairwise comparisons were presented to subjects. In this experiment, the standard configuration was fixed in one observation area and presented on every trial. Other specifications like monitor distance, fixation cross, and head rest were the same. All subjects pressed the left mouse button for same and the right one for different. Experiment III consisted of three different parts: First, subjects performed a pre-session, then three or four sessions in which the exact minima for $\psi(x, y_j)$, $j = 1 \ldots 5$, were determined with an adaptive procedure. Thirdly, subjects did thirty matchings for each surround.

Pre-Session

In the pre-session, subjects performed same-different judgments comparing the 23 stimuli shown in Figure 5.15 to the standard. Each stimulus was presented 30 times in 30 blocks with 23 trials each. These data were used to approximately determine where the minima for the five surrounds (see Table 5.19)

Table 5.19: Luminance Values of Surrounds Used in Adaptive Procedure

Surround	Luminance ($\frac{\text{cd}}{\text{m}^2}$)
1	58.98
2	65.52
3	**72.51**
4	80.51
5	88.85
Background	133.27

Note: Surrounds correspond to surrounds of Stimuli 0, 1, 6, 11, and 12 in Experiment II. Surround 3 (bold) was the surround of the standard configuration.

were placed for each subject. The standard (big black dot in Figure 5.15) used in Experiment III was identical to the standard shown in Figure 5.11 (see also Table 5.12, Stimulus 6). Five different surrounds were used. How they were arranged around the standard can be seen in Figure 5.15. For each surround, a quadratic model (see Equation 5.9) was fitted to the proportion of different responses. The data were plotted together with the model predictions to estimate the location of the minimum for this surround. The model was then used to determine the starting borders for the adaptive procedure. Six stimuli were chosen so that the lowest and the highest stimulus values predicted a 'different' response in about 80 %.

Adaptive Procedure

In order to determine the exact minima of $\psi(x, y_j)$ for each of the five different surrounds, subjects had to perform same-different judgments for six stimuli equally spaced between the borders determined in the pre-session. Stimuli were presented in random order and interwoven for all surrounds. Each session consisted of 900 trials total presented in 36 blocks (unless the adaptive procedure was finished beforehand). Figure 5.16 shows the procedure for one surround. As in Experiment I and II, stimuli were presented on a 10-bit black-

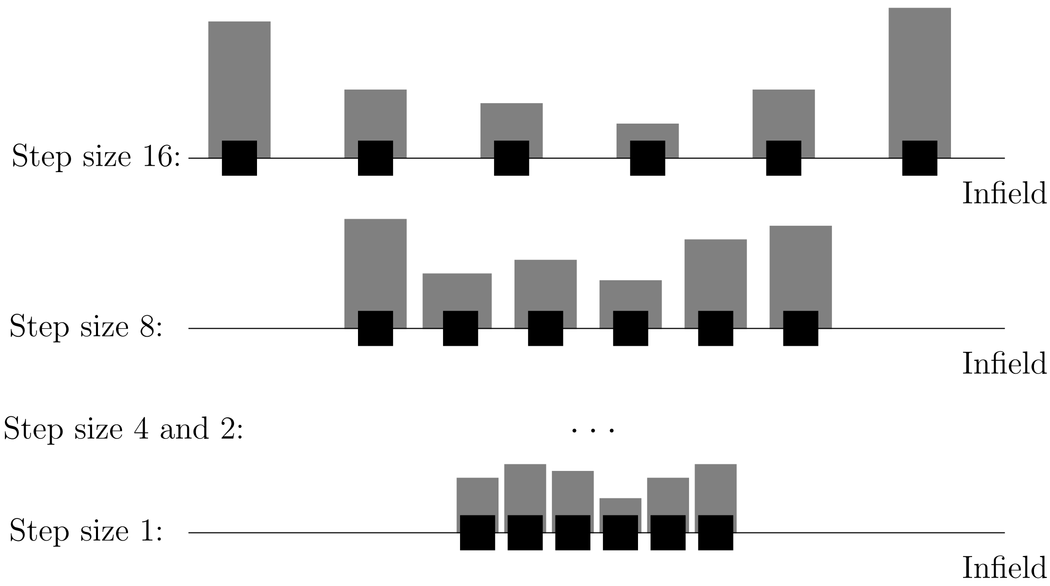


Figure 5.16: Schematic overview of adaptive procedure for one surround. Black squares are luminance values of infields and gray bars symbolize relative frequencies to respond different for that stimulus.

and-white monitor with 1024 intensity steps. The procedure used these 'color names' to evenly space the stimuli. The first borders were approximated on behalf of the data collected in the pre-session, for example gray 375 and 455 for the darkest surround. The spacing for the first stimulus configurations was 16 steps, i. e., the colors 375, 391, 407, 423, 439, and 455 were taken as infields. Thirty same-different judgments were collected for each stimulus. After collecting these 180 data points, a quadratic model

$$y_{ij} = \beta_0 + \beta_1 \text{Infield}_i + \beta_2 \text{Infield}_i^2 + \varepsilon_{ij} \tag{5.9}$$

was fitted to the proportion of different responses and compared to a model that predicted the mean

$$y_{ij} = \beta_0 + \varepsilon_{ij}. \tag{5.10}$$

In order to increase the power to detect curvature, an α level of 0.25 was chosen for these comparisons. If data from the pre-session showed very low curvature this α level was adjusted up to 0.5. If the quadratic model fitted the data better—meaning there was still curvature in the data—the borders were adjusted and the step size was halved to 8 steps. In order to adjust the borders, the minimum of the quadratic function was taken and the new borders were chosen in such a way that this minimum lay half way between the two new borders. With the new step size, the new interval was now half the size of the one before. Then the process started over. The adaptive procedure terminated if the quadratic model did not describe the data better than the simpler model, therefore, showing that there did not seem to be any curvature left in the data, or when the minimal step size of one was reached. For the responses collected, the procedure always terminated before the minimal step size was reached. Subjects were asked if they noticed that there was a standard and that it was always presented on the same side (left or right) of the monitor. All subjects answered with no.

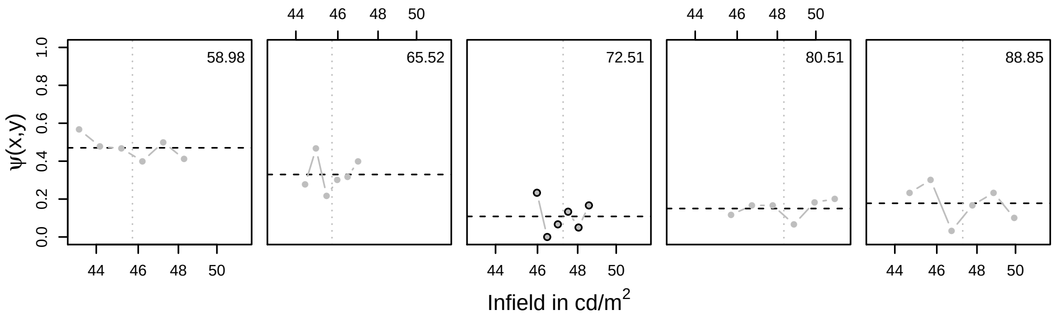
Matching

After the adaptive procedure, subjects did one more session where they adjusted the infields of the left or right stimulus, so that they would look the same as the infield of the standard. For each of the five surrounds, subjects performed 30 matchings in random order. Infields of the test stimulus could be adjusted by scrolling the mouse wheel. It was not possible to set a luminance for the infield that was higher than the luminance of the surround, i. e., the highest possible luminance was that of the surround. Afterwards they were asked if they thought it was hard to follow instructions and adjust the infield to a perfect match.

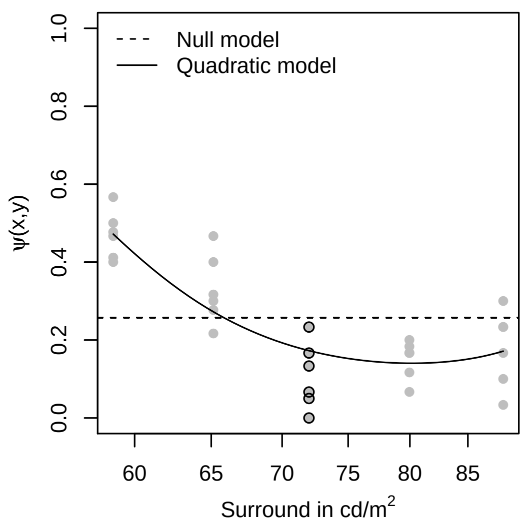
5.3.3 Results

Adaptive Procedure

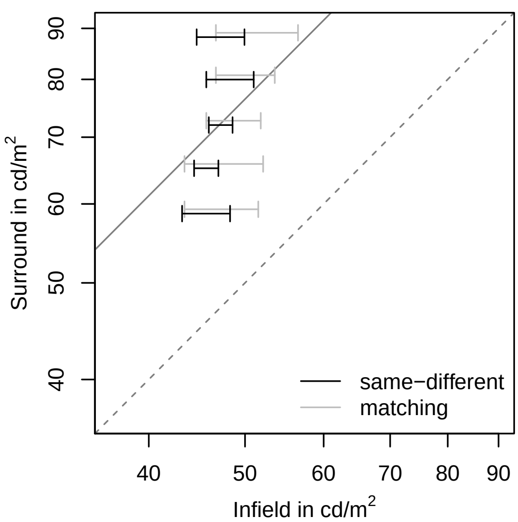
The results of the adaptive procedure are shown in Figures 5.17, 5.18, and 5.19. For Subject 3, only data for four surrounds could be collected. The results show that the minimum probability to say different, $\psi(x, y_j)$, differs for the different surrounds for all subjects. If the perceptual space for infield-surround configurations is one-dimensional, $\psi(x, y_j)$ should be the same for all



(a) Minima of $\psi(x, y_j)$ for five different surrounds.

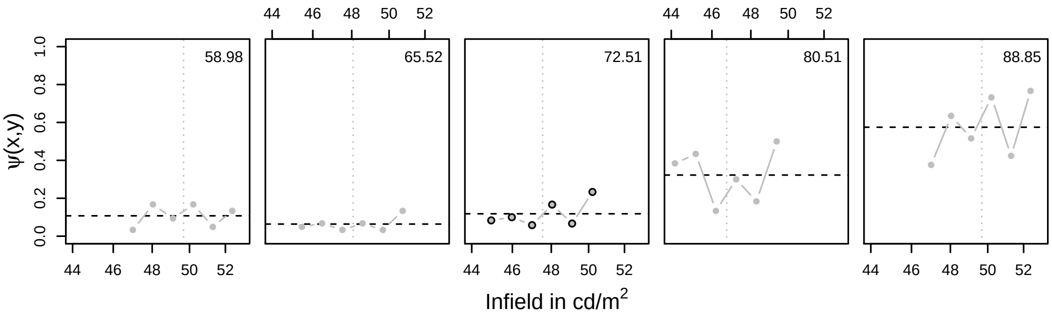


(b) Data points as above.

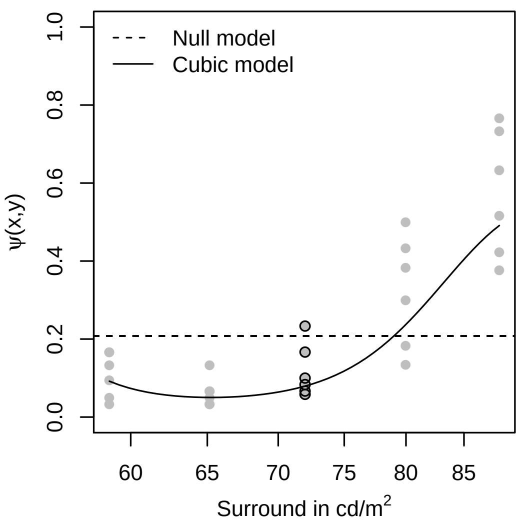


(c) 'Same intervals' for different surrounds.

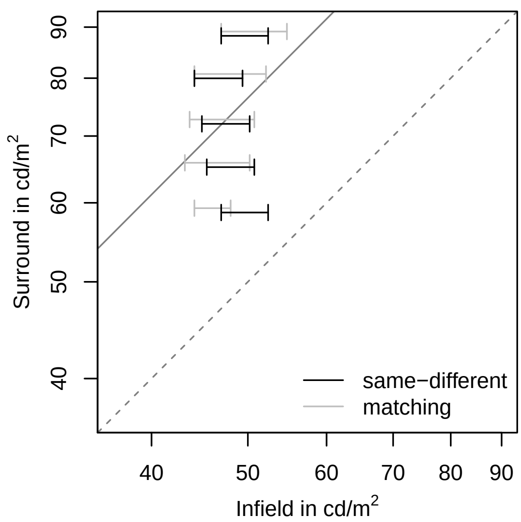
Figure 5.17: Estimated minimum discrimination probabilities for different surrounds for Subject 1. Gray solid line in Subfigure (c) shows which stimuli should be judged to be same according to the ratio principle. Black circled data points in Subfigures (a) and (b) indicate standard surround. In each trial, standard was constant at 47.45 $\frac{\text{cd}}{\text{m}^2}$ (infield) and 72.51 $\frac{\text{cd}}{\text{m}^2}$ (surround).



(a) Minima of $\psi(x, y_j)$ for five different surrounds.

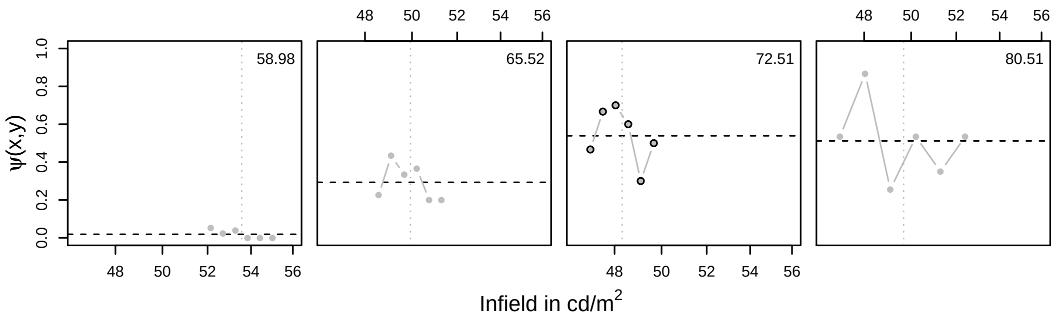


(b) Data points as above.

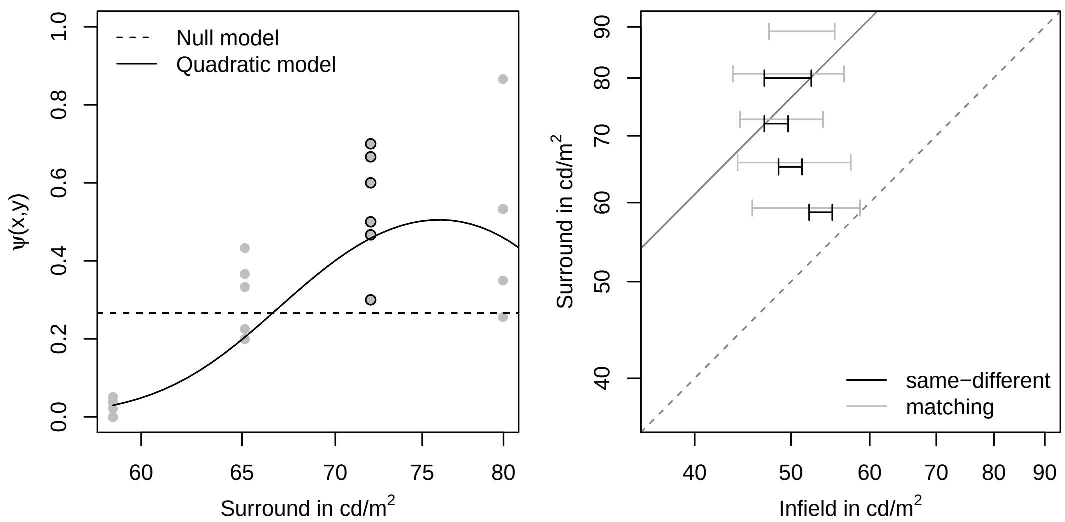


(c) ‘Same intervals’ for different surrounds.

Figure 5.18: Estimated minimum discrimination probabilities for different surrounds for Subject 2. See Figure 5.17 for details.



(a) Minima of $\psi(x, y_j)$ for four different surrounds.



(b) Data points as above.

(c) ‘Same intervals’ for different surrounds.

Figure 5.19: Estimated minimum discrimination probabilities for different surrounds for Subject 3. See Figure 5.17 for details.

surrounds. For none of the subjects, the minima for $\psi(x, y_j)$ were located on the ratio principle line (see Figures 5.17c, 5.18c, and 5.19c). As in Experiment II, subjects seemed to judge infields less often different which were physically nearly identical (they performed luminance matching). There is a certain range that subjects still perceive as same for the infields (as was to be expected). This range does not seem to be different when comparing infields with the same or with different surrounds. But $\psi(x, y_j)$ seems to be higher when infields in different surrounds are compared, at least for Subjects 1 and 2. Subject 3 seems to show the opposite. Overall, the pattern of the psychometric functions looks different for different subjects; a finding that already showed in the results of Experiment II (see Figure 5.11. But for all subjects the size of the minima for $\psi(x, y_j)$ varied with surround when comparing infields with surrounds different from the surround of the standard.

Figures 5.17a, 5.18a, and 5.19a show discrimination probabilities when the adaptive procedure found no curvature left in the data and ended data collection for that surround. It can be seen that the variance in the data was sometimes high, which might be due to the high uncertainty for these stimuli (even though repetitions for each data point varied between 30 and 120). The p-values for comparing the null model to the quadratic model from Equation 5.9 are shown in Table 5.20. The β_0 parameters for the null model show the mean $\psi(x, y_j)$ for that surround which could be considered an approximation of the minimum of the psychometric function for that surround. The estimated β_2's of the quadratic model show that the model could not find curvature in the data for any of the surrounds. This is also shown by the p-values which show that the null model did not significantly differ from the quadratic model. At this point the adaptive procedure aborted the data collection for the respective surround.

In order to analyze if the magnitude of the minimum $\psi(x, y_j)$ differed for the different surrounds, logistic regression models were fitted to the data points

Table 5.20: Parameter Estimates and p-Values for Model Comparison when Adaptive Procedure Stopped

	Surround	β_0 (Eq 5.10)	β_2 (Eq. 5.9)	p (H_0: $\beta_2 = 0$)
Subject 1	1	0.470	≤ 0.001	0.264
	2	0.321	0.003	0.719
	3	0.098	0.006	0.276
	4	0.133	0.001	0.398
	5	0.144	0.001	0.481
Subject 2	1	0.103	$\leq \lvert -0.001 \rvert$	0.831
	2	0.066	≤ 0.001	0.332
	3	0.121	≤ 0.001	0.384
	4	0.322	0.002	0.361
	5	0.601	$\leq \lvert -0.001 \rvert$	0.425
Subject 3	1	0.018	≤ 0.001	0.113
	2	0.293	-0.006	0.250
	3	0.539	-0.006	0.546
	4	0.512	0.001	0.776

Note: See Figures 5.17a, 5.18a, and 5.19a for data points models were fitted to.

that were collected within the last interval. The null model

$$\operatorname{logit}(\psi(x, y_{ij})) := \log \frac{\psi(x, y_{ij})}{1 - \psi(x, y_{ij})} = \beta_0 \tag{5.11}$$

was compared to a quadratic, cubic and saturated model

$$\operatorname{logit}(\psi(x, y_{ij})) = \beta_0 + \beta_1 \text{Surround}_{ij} + \beta_2 \text{Surround}_{ij}^2, \tag{5.12}$$

$$\operatorname{logit}(\psi(x, y_{ij})) = \beta_0 + \beta_1 \text{Surround}_{ij} + \beta_2 \text{Surround}_{ij}^2 + \beta_3 \text{Surround}_{ij}^3, \tag{5.13}$$

$$\operatorname{logit}(\psi(x, y_{ij})) = \beta_0 + \beta_j, \tag{5.14}$$

with $i = 1 \ldots 6$, for the data points and $j = 1 \ldots 5$, for the different surrounds ($j = 1 \ldots 4$, for Subject 3). The observation areas were fixed for each subject, meaning that the standard was always on the left or the right side over the course of the experiment. For Subjects 1 and 3, the standard was presented on the left side. Comparing these models showed that it can be ruled out that there are constant minima for $\psi(x, y_j)$ for different surrounds (see Table 5.21). For Subject 1, the quadratic model does not fit the data very well and the cubic term does not improve the fit. Nonetheless, the qualitative pattern of the data seems to be captured by the quadratic model and it can be ruled out that the null model can describe the data satisfactorily. For Subjects 1 and 2, comparing the infield of the standard to an infield in another surround increased the probability to say different. According to the probability-distance hypothesis or "the old, famous psychological rule of thumb: equally often noticed differences are equal, unless always or never noticed" (Luce & Edwards, 1958, p. 232), this means that the perceptual distances for these infields vary with surround. None of the probabilities observed here were 1 or 0. It can also be assumed that the actual minima for $\psi(x, y_j)$ were found when the psychometric functions are smooth increasing functions. For Subject 3, it must therefore be assumed that the result pattern would look clearer if data for more surrounds had been collected. Like Subject 3 in Experiment I, this subject might show a bias to judge the stimulus on the left (this subjects standard) to be darker than the stimulus on the right.

Table 5.21: Model Comparisons for Null Model, Quadratic, Cubic, and Saturated Model

	Model	Resid. df	Resid. Dev	df	Deviance	p
Subject 1	null	29	166.97			
	quadratic	27	55.69	2	111.27	≤ 0.001
	cubic	26	52.81	1	2.89	0.089
	saturated	25	45.17	1	7.63	0.006
Subject 2	null	29	365.53			
	quadratic	27	88.58	2	276.95	≤ 0.001
	cubic	26	80.19	1	8.39	0.004
	saturated	25	80.18	1	0.01	0.909
Subject 3	null	23	394.53			
	quadratic	21	97.02	2	297.510	≤ 0.001
	saturated	20	94.73	1	2.29	0.130

Matching

Figures 5.17c, 5.18c, and 5.19c also show the results for the matching task subjects performed after the same-different judgments of the adaptive procedure. Gray intervals show the range (minimum to maximum luminance) to which subjects set the luminance of the infield. The results for the matching task show the same overall pattern as the same-different judgments. In both tasks, subjects seem to perceive the smallest difference between stimuli when luminance is the same (or similar) for the infields. The bigger range of the matching task in comparison to the same-different judgments is not necessarily interpretable. The range of the same-different judgments was restrained by the procedure used. It might in fact be bigger. One might assume that the same-different intervals are as big as the ones for the matching task unless shown differently in an experiment designed to address this question.

None of the subjects reported any difficulties in performing the matching task nor that matches seemed unsatisfactory. When asked if they found the same-different judgments difficult they all answered yes. All subjects reported that they did not follow any response strategies throughout the experiment, but answered spontaneously (two of them used the word intuitively).

5.3.4 Discussion

The results of Experiment III suggest that a second perceptual dimension is needed in the perceptual space of infield-surround configurations presented under a constant illumination. The psychometric functions show that it cannot be assumed that the perceptual space for achromatic decrements under constant illumination is one-dimensional, since that would mean constant minima for $\psi(x, y_j)$ for all five surrounds, what was not found. Although subjects match infields according to their luminance, the perceptual distance between infields with different surrounds seems to be bigger than for infields with the same surround. Our results also show that the size of this effect might interact with the observation area. When the standard was presented on the left side (Subject 1) infields with lighter surrounds were judged equally often different than infields with the standard surround. When the standard was presented on the right side (Subject 2) this pattern was reversed. This interpretation deserves some caution since it is based on the results of only two subjects. Collection of more data and subjects seems necessary to get a clearer picture. Future studies might want to repeat this experiment with more subjects and for more surrounds. Additionally, it might be necessary to start with a broader range of infields (maybe ten) to cover a bigger range from the start to minimize the possibility that the adaptive procedure 'runs astray' as happened during pre-data collection when still testing the adaptive procedure.

That subjects tend to perceive lightness of infields as being minimally different when luminance is similar was already a result of Experiment II. However, the number of violations of regular minimality and the puzzling results of the MDS for different dimensions did not strengthen our confidence in those data. That infield-surround ratio does not seem to play a role here, supports results found in Experiment II, but raises questions about results found in other studies. This might be a unique finding for infield-surround configurations presented on a bright background in an illuminated room. Future experiments could look at different illuminations for different sessions and see how that influences the form of the psychometric functions and the position of the minimum discrimination probabilities.

That the matching task showed results very similar to those obtained with the same-different judgments is a reassuring result. It shows that the adaptive procedure did not go astray for the current subjects. However, it was surprising that subjects reported that the same-different task was more difficult than the matching task. It is often argued that subjects experience difficulties while doing matching tasks. They report that they cannot adjust infields to obtain a satisfactory match (Logvinenko & Maloney, 2006). This was not the case for our subjects. It might be due to the fact that the surrounds did not differ very much from each other. The biggest difference was 29.87 $\frac{\mathrm{cd}}{\mathrm{m}^2}$. Furthermore, subjects did not have to compare increments to decrements. Since it has been established that increments and decrements are processed in different ways (Gilchrist, 2006; Whittle & Challands, 1969) this might be an impossible task. We deliberately avoided this in all our experiments. However, when considering that subjects did not report any difficulties performing the matching task, one should not forget that the results from the same-different judgments show that the minimum discrimination probabilities of these 'matches' is around 0.5 for some of the surrounds.

Chapter 6

General Discussion

6.1 Empirical Results

The goal of this thesis was to systematically investigate the dimensionality of the perceptual space of achromatic surface colors for infield-surround configurations presented under constant illumination. In Experiment I, simple gray patches were presented to the subjects. The results showed that under constant illumination conditions one dimension seemed sufficient to represent perceptual distances between stimuli.

In Experiment II, subjects were presented with thirteen different infield-surround configurations under the same illumination conditions as in Experiment I. Subjects performed same-different judgments. The two-dimensional psychometric functions obtained with the same-different task suggest that two perceptual dimensions are needed to discriminate between two infields embedded in different surrounds. However, dimensionality results obtained with multidimensional scaling (MDS) did not convey a consistent picture. Although stress was low and a one-dimensional MDS solution was readily interpretable in terms of lightness of the infield as the relevant dimension, psychometric functions and the two-dimensional MDS solutions suggested a second dimension.

In order to get a clearer picture of these results, a third experiment was conducted. In this experiment, the shape of the two-dimensional psychometric functions was focused on. A tulip shaped psychometric function indicates the need for two perceptual dimensions (as the shape found in Experiment II suggested). A psychometric function with constant minimum values for different points in the stimulus space indicates that subjects need only a single perceptual dimension to discriminate between stimuli (Dzhafarov & Colonius, 1999). The results of all three subjects supported the hypothesis that psychometric functions do not have constant minima for different positions in the stimulus space (as a shape forming a valley would suggest). Thus, the results argue for two dimensions.

6.2 Conclusions for Dimensionality

The results found in Experiment II do not convey a clear picture when we try to interpret them in the light of dimensionality. The psychometric functions and the layout of the two-dimensional MDS solutions suggest two perceptual dimensions. Izmailov and Sokolov (1991) did a similar experiment where they presented infield-surround configurations with varying surrounds. Using nonmetric MDS, they found a two-dimensional solution where stimuli could be arranged on a circle. This supports a one-dimensional mental representation since a circle is a topologically one-dimensional structure embedded in a two-dimensional space. This result seemed surprising under theoretical considerations which imply that this stimulus situation should lead to two perceptual dimensions (Niederée, 1998). Our results differ from the ones by Izmailov and Sokolov (1991) with respect to the two-dimensional MDS solutions. We had a different stimulus situation since we used eleven different surrounds (as compared to three) and presented stimuli in an illuminated room (they used an apparatus with Maxwellian-viewing conditions). However, results of Experiment III showed that subjects matched infields according to their luminance. This is in good agreement with the results from Izmailov and Sokolov (1991) who found that subjects “perceived achromatic differences only between disc fields, independently from ring fields” (p. 255). This is also in accordance with the one-dimensional MDS solutions found in our second experiment. Thus, these results seem to hold for dark and illuminated viewing conditions. However, this alone does not allow us to draw conclusions about the existence of a second perceptual dimension, even though the shape of the psychometric functions suggests the need for two perceptual dimensions.

The third experiment showed that surrounds interpreted as local context effects indeed introduce a second perceptual dimension. According to the probability-distance hypothesis, an increased probability to say different implies an increased perceptual distance for these stimuli. The shape of the psychometric functions obtained in Experiment III shows that when infields are presented with different surrounds their perceptual distance increases. Even though subjects judge infields less often different when they have the same luminance, they are still not perceived as being same, because this would require that minima of the psychometric functions for different surrounds are constant. This shows that same-different judgments and their two-dimensional psychometric functions are well suited to investigate questions about dimensionality of perceptual color space for achromatic colors. Procedures like the ones used by Izmailov and Sokolov (1991) might not always capture the whole picture. The results by Izmailov and Sokolov (1991) imply a one-dimensional perceptual space, but the MDS could only pick up the luminance ‘matching’ and could not give any insight on the second underlying perceptual dimension. That subjects only seem to take luminance of the infield into account when judging if infields are same or different in the second and third experiment, is therefore not at odds with a two-dimensional perceptual space.

The results of the psychometric functions found in Experiment III support a two-dimensional interpretation. Most studies reporting two-dimensional results for achromatic color space argue that these two dimensions are associated with perceived surface reflectance (lightness) and illumination (Logvinenko, 2005; Logvinenko & Ross, 2005; Logvinenko & Maloney, 2006). This fits the theoretical distinction between surface colors and illumination colors made by Mausfeld (1998) and others (Evans, 1964; Heggelund, 1992; Niederée, 1998).

Heggelund (1992) shows that his two perceptual processes w (white) and b (black-luminous) are associated with luminance of infield (w) and contrast (b). The w process corresponds to object color and the b process to illumination (what he calls light) color. Results reported so far suggest that for achromatic stimuli object color (lightness) strongly correlates with luminance and infield-surround ratio influences lightness when surrounds are perceived as illumination. Since illumination was constant for all our experiments, it appears that the w process influenced the perception of infields more than the b process (when trying to interpret results in terms of Heggelund's theory).

Gilchrist (2006) associates his two types of lightness constancy with luminance and infield-surround ratio as well: "Empirical results lie closest to a luminance ratio match for illumination-independent constancy but closest to a luminance match for background-independent constancy" (p. 290). Seen as the distinction between object and illumination colors, object color is constant for different surrounds and illuminations, and illumination color changes with different illuminations. Subjects in our experiments showed background-independent constancy. However, this constancy does not imply that they use only one perceptual dimension to obtain this constancy.

The concept of two independent color codes that are not incommensurable is in opposition to theories like discounting the background (Walraven, 1976). Most computational theories on color perception incorporate some kind of discounting the background in their models in order to explain color constancy (usually, illumination-independent color constancy, Mausfeld, 1998). These approaches focus on the discovery of reflectance properties as the main goal of color perception. Mausfeld (1998) argues that these approaches show "systematic drawbacks from a psychological point of view" (p. 240). The perception of illumination is neglected in these approaches (see Gilchrist, 2006, Chapter 8, for an overview) and they went as far as to declare that illumination is not perceived at all. That this is obviously wrong does not have to be explicitly explained as it is evident to everybody that we can discriminate between night and day. The distinction between object and illumination colors, on the other hand, seems like a fruitful approach and theoretically less flawed. Logvinenko and Ross (2005, p. 63) summarize this as "[t]herefore, the question is not how the visual system discounts the illumination changes, but how it encodes them, and takes them into account when calculating lightness."

However, there has been much confusion in the literature on what really constitutes illumination due to the interpretation of surrounds as illumination. As mentioned before, this is a critical assumption. This thesis aimed at disentangling these concepts by presenting stimuli under constant illumination conditions. Surrounds were, therefore, perceived as context effects only. Niederée (1998, 2010) postulates that we need two perceptual dimensions to discriminate between infield-surround configurations presented in the dark. This implies that two perceptual dimensions are needed to take context effects into account.

The results found in Experiment III suggest that this is indeed the case. The two-dimensional psychometric functions do not form constant minima over the two-dimensional stimulus space which implies two perceptual dimensions. Our results are in opposition to results found for the ratio principle by Wallach (1948). The main difference between experiments is that Wallach's stimuli were presented in the dark (he used decrements as well). Apparently, lightness perception of infield-surround configurations differs for different illumination conditions. This itself supports the hypothesis of a two-dimensional perceptual color space for achromatic infield-surround configurations. But the relevant dimension here seems to be illumination and not context.

In the literature, authors often distinguish between *luminance matching* and *ratio matching*. Luminance matching means that infields are matched according to their luminance and ratio matching means that infields are matched when infield-surround ratio is identical for both stimuli. Usually, ratio matching is associated with decrements and luminance matching is associated with increments when presented under dark room conditions (Gilchrist, 2006). Whittle (1986), however, showed that high contrast decrements with small infields produce luminance matching, whereas low contrast decrements tend to produce ratio matching. He concludes: “This difference between the ways stimuli are most simply produced suggests a characterisation of the ΔL [infield-surround ratio] and L_a [luminance] ranges as corresponding to the ‘world of shadows’ and the ‘world of objects’ respectively” (p. 1686). This is again a distinction between illumination and object color and Whittle’s results suggest that subjects interpreted infield-surround configurations with higher contrast as surrounds and those with lower contrast as illumination.

The matching task in Experiment III showed that subjects performed luminance matching for decrements under a constant illumination and that the location of the minima for the same-different judgments could also be explained by different luminance values of the infields. This implies that ratio matching might only be applied when surrounds are interpreted as illumination (as might be the case in dark room conditions). When stimuli are clearly perceived as surface colors, subjects do not have any problems to match infields according to their luminance. This result should not be surprising, since we are confronted with colored objects on different backgrounds every day and usually these objects do not seem to change their color when we move them around from background to background. On the contrary, color constancy is a very stable perception in every day viewing conditions (Foster, 2003).

Our results show that context effects introduce an extra perceptual dimension. Most studies concerned with dimensionality of achromatic color space found that an extra perceptual dimension is needed to discriminate between stimuli under different illumination conditions. Apparently, illumination is not the only relevant factor introducing a second perceptual dimension. Surround and illumination were often confounded in traditional dark room experiments investigating the perception of lightness. The results found here show that an additional perceptual dimension is introduced for local context effects. It is an open question how many perceptual dimensions are needed for stimulus situations with different local contexts as well as different illumination conditions.

Based on the results found in this thesis, it is difficult to draw conclusions on the nature of the second dimension found in Experiment III. Recent evidence suggests that the second dimension constitutes a higher-level process like the perception of transparency (Ekroll & Faul, 2012a, 2012b, 2013). Experiments manipulating transparency of stimuli or other possible interpretations of the scene need to be conducted in order to investigate if higher-level interpretations play a critical role here. It could also be argued that surround does not introduce a second *perceptual* dimension, but rather a dimension of uncertainty. Subjects do not perceive infields in different surrounds as being different, but they have a higher uncertainty if infields really are the same. In every day life, we experience this by the fact that we often move objects onto the same background when we are not sure if they really have the same color (for example putting a top and a sweater right next to each other on the uniformly colored sheets of the bed). However, this uncertainty clearly implies that we perceive stimuli with different surrounds not always as same.

6.3 Methodological Criticism

Niederée (1998) criticizes that the stimulus duration of 0.5 s and the low range of surrounds was responsible for the failure to find a two-dimensional solution for the achromatic data collected by Izmailov and Sokolov (1991). In our experiments, stimuli were presented for 0.5 s as well. This was necessary in order to be able to collect a sufficient number of trials. Subjects already conducted 15 to 20 sessions of about one hour. Presenting stimuli for less than one second is common practice in psychophysical experiments (e. g., Whittle, 1986). The interpretations of the scene subjects are supposed to perform here should be highly automatized, and there is no reason to assume that 0.5 s are not enough time to discriminate between stimuli. Furthermore, pre-experiments were conducted with presentation times of 1 s and results looked comparable. Moreover, subjects reported that the experiment was already tedious and that the long presentation time was adding to this feeling.

The luminance range used in our experiments was smaller than the one used by Izmailov and Sokolov (1991). Since we were interested in local changes of the psychometric functions, this was a necessary restriction. If the surround indeed influences the perception of the infield, there is no reason to assume that small changes (as long as they are actually perceivable) do not already make a difference. Furthermore, our data clearly show that the different surrounds made a difference in the perception.

A problem that questions our approach in general were the violations of regular minimality. While planning the experiments, there was no reason to expect that subjects would have trouble to judge stimuli which are physically identical less often different than other stimuli. Since this critical empirical assumption was violated for some subjects, application of Fechnerian scaling was difficult. Therefore, the distances computed with Fechnerian scaling have to be interpreted with care. However, violations of regular minimality did not seem to be systematic for any of the subjects. The difficulty of the task and the accompanying uncertainty was clearly underestimated. Hopefully, the number of violations of regular minimality was only caused by measurement error. The statistical test evaluating if the violations are significant (Dzhafarov et al., 2011) only counts the number of violations. It would be helpful to develop statistical tools which also take the magnitude of the violations into account. The violations found for our data were small in magnitude, and a statistical procedure that takes this size into account might give a clearer picture of the severity of the violations found here.

A straightforward approach could be to develop an algorithm which finds a matrix that is minimally different from the original data matrix but is regular minimality compliant. Both, the number and size of changes necessary to create a matrix like this could be taken as a measure to evaluate the severity of the violations. Assuming that regular minimality violations are due to statistical fluctuation, another approach to deal with violations could be to use an adaptive procedure that ensures that repetitions for critical cells are increased. A combination of both approaches might be beneficial.

Problems with violations of regular minimality already reveal some of the problems with the same-different judgments paradigm. In Experiment I, subjects performed one session of greater-less judgments to determine their sensitivity and to evaluate if they were able to discriminate between stimuli, but still subjects had difficulties discriminating stimuli with same-different judgments. They also reported that they found the same-different task harder than both the greater-less task (Experiment I) and the matching task (Experiment III).

The same-different task differs from the greater-less task in the type of decision involved (Schneider & Komlos, 2008), but why the same-different task appears to be so much more difficult has not been discussed often. Dzhafarov and Colonius (1999) discriminate between both tasks and their underlying theoretical and empirical assumptions, but they do not consider practical implications or underlying decision tasks. To make things worse, the literature often does not discriminate very well between same-different and greater-less judgments (e. g., Kingdom & Prins, 2010). Often, authors talk about same-different tasks when they actually mean greater-less judgments. This is unfortunate and complicates the understanding of the problems with the task.

Often, same-different judgments are compared to the so-called yes-no task, just for discrimination instead of detection. In a yes-no task subjects have to say if a stimulus was present or not. In a same-different task subjects have to state if a 'difference was present or not.' The yes-no task (and therefore the same-different task as well) has often been criticized as being prone to response bias. The reponse depends on some mental criterion subjects apply (in our case this might be a tendency to answer different, when they are unsure; Wickens, 2002). Our results showed that the same-different judgments were prone to response bias. It is surprising that subjects judge stimuli presented next to each other on a monitor which are physically identical to be different in more than 50 % of the cases (even if the design induces some kind of response bias, see Section 5.1.2 for more details). These response biases could have a number of origins: First, subjects show a tendency to answer different when they are unsure (as mentioned above). Secondly, the different observation areas could influence the perception. Subjects might perceive stimuli presented on the right as being different from stimuli presented on the left (as was the case for Subject 3 in Experiment I). It has been shown in a temporal greater-less task that subjects are more sensitive when the standard is presented first (Dyjas, Bausenhart, & Ulrich, 2012). In addition, our results showed that the probability to say different depends on observation area as well ($\psi(x, y) \neq \psi(y, x)$). It has been discussed before (see Section 5.1.4) that response bias should not influence the dimensionality of the precpetual space. Nonetheless, it might be interesting to address these questions in more detail in future experiments.

That the matching was perceived as being easier than the same-different task is surprising. Many authors report that infields with different surrounds are impossible to match (Logvinenko & Maloney, 2006; Niederée, 1998). However, subjects in Experiment III had no difficulty to match infields in different surrounds. As mentioned above, the luminance range of the surrounds was limited and this might be one reason. Additionally, subjects only had to match decrements. When matching increments and decrements the task might become increasingly more difficult. Logvinenko and Maloney (2006) argue that matching infields with different surrounds is difficult because of the two-dimensional nature of the perceptual space. Subjects are only able to manipulate one of the variables (luminance of the infield) and not the other (illumination in their case, luminance of the surrounds for our experiments). Therefore, matches can be unsatisfactory. Logvinenko and Maloney (2006) conclude that in matching tasks under different illuminations "observers are facing a problem that does not have a solution" (p. 81). In our experiment, illumination was constant and that might be the reason why subjects had less difficulties matching infields with different surrounds. Our perceptual system might compensate for luminance of surrounds when interpreted as context effects better than for different illuminations. Subjectively, most people would probably state that the color of an object is influenced stronger by different illumination conditions than by different surrounds.

6.4 Future Directions

The second perceptual dimension introduced by local context effects in the form of uniform surrounds found here is of a different nature than the second perceptual dimension found for different illumination conditions (as found by, e. g., Logvinenko & Maloney, 2006). It is therefore difficult to determine what this second dimension constitutes in qualitative terms. In order to investigate the nature of the second dimension, one needs to come up with stimulus situations that keep the distinction between illumination and surrounds and at the same time keep the illumination constant, so that context effects can be studied individually. Stimuli should increase gradually in their complexity. It is important to follow a continuous path here and relate results to each other (Mausfeld, 1998). Possible candidates for investigation are perceived transparency (Ekroll & Faul, 2013) or three-dimensional interpretations (Logvinenko, 2005).

Most theoretical work on the dimensionality of perceptual space for achromatic surface colors concludes that the relevant dimensions are lightness and perceived illumination (Heggelund, 1992; Logvinenko & Maloney, 2006; Logvinenko & Ross, 2005; Mausfeld, 1998; Niederée, 1998). The results found here suggest that context effects add an extra perceptual dimension as well and demonstrate how important it is to use experimental settings that allow us to distinguish between context and illumination.

Future research in this area should focus on clearly separating these different concepts. In order to further investigate how illumination influences the perception of achromatic surface colors, experiments reported here could be replicated under different illumination conditions. For example, Fechnerian distances for single gray patches as obtained in Experiment I could be compared for different illumination conditions. It would be interesting to investigate how Fechnerian distances relate to each other under different illuminations. Leaning on the results found by Logvinenko and Maloney (2006), it would be expected that "[w]hen illumination decreases, the lightness continuum shrinks" (p. 80), i. e., the Fechnerian distances should be smaller for darker illumination and vice versa.

Understanding how the visual system incorporates different contexts and situations can best be investigated when following a path from the simplest possible stimuli over more complex stimuli to the point of natural scenes (Mausfeld, 1998). Most research in visual perception focuses either on simple (low-level) stimuli or on natural scenes. Often, investigations define themselves more by the stimuli they use than by the psychological phenomena involved. Focusing on lightness perception and increasing the complexity of stimuli seems like a fruitful approach to overcome many of the contradictions in the literature. Brainard and Maloney (2011) introduce a class of models they call Equivalent Illumination Models that are well suited for use on simple scenes (what they call flat-matte-diffuse conditions) as well as on more generalized stimulus conditions (like three-dimensional scenes). Their computational approach tries to find the illuminant the visual system deduces for a specific stimulus situation in order to achieve color constancy. They do not make assumptions how or why the visual system misjudges the illuminant, but point out that this is the logical next step. This might be a worthwhile starting point to include cognitive influences like interpretation of the scene into the model. They emphasize the importance of constructing models that can make predictions for simple as well as complex scenes: "Without some principled way to generalize the understanding we gain from simple scenes, the task of measuring and characterizing the interaction of all the relevant scene variables in terms of how they affect surface color perception seems so daunting as to be hopeless" (p. 12).

In research on color perception, it is common practice to use very simple stimulus conditions, and the results found with these stimuli have deepened our understanding of color perception fundamentally. However, switching to more complex and therefore less artificial stimulus situations seems to be the next logical step. The results of this thesis underline how using conceptually clear tasks as well as more elaborate experimental settings help understand what we perceive.

References

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences.* Cambridge, MA: MIT Press.

Anton-Erxleben, K., Abrams, J., & Carrasco, M. (2010). Evaluating comparative and equality judgments in contrast perception: Attention alters appearance. *Journal of Vision*, *10*(11), 1–22.

Anton-Erxleben, K., Abrams, J., & Carrasco, M. (2011). Equality judgments cannot distinguish between attention effects on appearance and criterion: A reply to Schneider (2011). *Journal of Vision*, *11*(13), 1–8.

Arabie, P. (1991). Was Euclid an unnecessarily sophisticated psychologist? *Psychometrika*, *56*, 567–587.

Borg, I., & Groenen, P. (2005). *Modern multidimensional scaling: Theory and applications.* New York: Springer.

Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision*, *11*(5), 1–18.

deLeeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*(3), 1–30.

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2012). Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Attention, Perception, & Psychophysics*, *74*, 1819–1841.

Dzhafarov, E. N. (2002a). Multidimensional Fechnerian scaling: Pairwise comparisons, regular minimality, and nonconstant self-similarity. *Journal of Mathematical Psychology*, *46*, 583–602.

Dzhafarov, E. N. (2002b). Multidimensional Fechnerian scaling: Perceptual separability. *Journal of Mathematical Psychology*, *46*, 564–582.

Dzhafarov, E. N. (2002c). Multidimensional Fechnerian scaling: Probability-distance hypothesis. *Journal of Mathematical Psychology*, *46*, 352–374.

Dzhafarov, E. N. (2002d). Multidimensional Fechnerian scaling: Regular variation version. *Journal of Mathematical Psychology*, *46*, 226–244.

Dzhafarov, E. N. (2003a). Thurstonian-type representations for "same-different" discriminations: Deterministic decisions and interdependent images. *Journal of Mathematical Psychology*, *47*, 184–204.

Dzhafarov, E. N. (2003b). Thurstonian-type representations for "same-different" discriminations: Probabilistic decisions and interdependent images. *Journal of Mathematical Psychology*, *47*, 205–219.

Dzhafarov, E. N. (2008a). Dissimilarity cumulation theory in arc-connected spaces. *Journal of Mathematical Psychology*, *52*, 73–92.

Dzhafarov, E. N. (2008b). Dissimilarity cumulation theory in smoothly connected spaces. *Journal of Mathematical Psychology*, *52*, 93–115.

Dzhafarov, E. N. (2010a). Dissimilarity cumulation as a procedure correcting for violations of triangle inequality. *Journal of Mathematical Psychology*, *54*, 284–287.

Dzhafarov, E. N. (2010b). Dissimilarity, quasidistance, distance. *Journal of Mathematical Psychology*, *54*, 334–337.

Dzhafarov, E. N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. *Psychonomic Bulletin & Review*, *6*, 239–268.

Dzhafarov, E. N., & Colonius, H. (2001). Multidimensional Fechnerian scaling: Basics. *Journal of Mathematical Psychology*, *45*, 670–719.

Dzhafarov, E. N., & Colonius, H. (2005a). Psychophysics without physics: a purely psychological theory of Fechnerian scaling in continuous stimulus spaces. *Journal of Mathematical Psychology*, *49*, 1–50.

Dzhafarov, E. N., & Colonius, H. (2005b). Psychophysics without physics: extension of Fechnerian scaling from continuous to discrete and discrete-continuous stimulus spaces. *Journal of Mathematical Psychology*, *49*, 125–141.

Dzhafarov, E. N., & Colonius, H. (2006a). Reconstructing distances among objects from their discriminability. *Psychometrika*, *71*, 365–386.

Dzhafarov, E. N., & Colonius, H. (2006b). Regular Minimality: A fundamental law of discrimination. In H. Colonius & E. N. Dzhafarov (Eds.), *Measurement and representation of sensations* (pp. 1–46). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dzhafarov, E. N., & Colonius, H. (2007). Dissimilarity cumulation theory and subjective metrics. *Journal of Mathematical Psychology*, *51*, 290–304.

Dzhafarov, E. N., & Colonius, H. (2011). The Fechnerian idea. *American Journal of Psychology*, *124*, 127–140.

Dzhafarov, E. N., & Paramei, G. V. (2010). Space of facial expressions: Cumulated versus transformed dissimilarities. In A. Bastianelle & G. Vidotto (Eds.), *Fechner Day 2010. Proceedings of the 26th annual meeting of the International Society for Psychophysics* (pp. 605–610). Padua, Italy: The International Society for Psychophysics.

Dzhafarov, E. N., Ünlü, A., Trendtel, M., & Colonius, H. (2011). Matrices with a given number of violations of regular minimality. *Journal of Mathematical Psychology*, *55*, 240–250.

Ekman, G. (1954). Dimensions of color vision. *The Journal of Psychology*, *38*, 467–474.

Ekroll, V., & Faul, F. (2009). A simple model describes large individual differences in simultaneous colour contrast. *Vision Research*, *49*, 2261–2272.

Ekroll, V., & Faul, F. (2012a). Basic characteristics of simultaneous color contrast revisited. *Psychological Science*, *23*, 1246–1255.

Ekroll, V., & Faul, F. (2012b). New laws of simultaneous contrast? *Seeing and Perceiving*, *25*, 107–141.

Ekroll, V., & Faul, F. (2013). Transparency perception: the key to understanding simultaneous color contrast. *Journal of the Optical Society of*

America A, *30*, 342–352.

Evans, R. M. (1964). Variables of perceived color. *Journal of the Optical Society of America*, *54*, 1467–1474.

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

García-Pérez, M. A., & Alcalá-Quintana, R. (2013). Shifts of the psychometric function: Distinguishing bias from perceptual effects. *The Quaterly Journal of Experimental Psychology*, *66*, 319–337.

Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology*, *39*, 169–200.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals.* Hillsdale, NJ: Erlbaum.

Gilchrist, A. (1979). The perception of surface blacks and whites. *Scientific American*, *240*, 112–123.

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

Groenen, P. J. F., Heiser, W. J., & Meulman, J. J. (1999). Global optimization in least-squares multidimensional scaling by distance smoothing. *Journal of Classification*, *16*, 225–254.

Heggelund, P. (1992). A bidimensional theory of achromatic color vision. *Vision Research*, *32*, 2107–2119.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing- and test-field luminances. *Journal of Experimental Psychology*, *50*, 89–96.

Heller, J. (2001). Mittenbildung bei achromatischen Farben: Das klassische Experiment von Plateau. *Zeitschrift für Experimentelle Psychologie*, *48*(4), 259–271.

Helmholtz, H. v. (1911). *Handbuch der Physiologischen Optik: Die Lehre von den Gesichtsempfindungen.* Leipzig: Leopold Voss.

Hess, C., & Pretori, H. (1970/1894). Quantitative investigation of the lawfulness of simultaneous brightness contrast. *Perceptual & Motor Skills*, *31*, 947–969.

Hoffmann, D. D. (1998). *Visual intelligence.* New York: Norton & Company.

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

Irtel, H. (1991). *Psychophysische Invarianzen in der Farb- und Helligkeitswahrnehmung.* Berlin: Springer.

Irtel, H. (2005). Psychophysical scaling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1628–1632). Chichester, UK: John Wiley & Sons.

Ishihara, S. (2012). *Ishihara's tests for color deficiency.* Tokyo: Kanehara Trading.

Izmailov, C. A., & Sokolov, E. N. (1991). Spherical model of color and brightness discrimination. *Psychological Science*, *2*, 249–259.

Izmailov, C. A., & Sokolov, E. N. (2004). Subjective and objective scaling of large color differences. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 27–43). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Jameson, D., & Hurvich, L. M. (1964). Theory of brightness and color contrast in human vision. *Vision Research*, *4*, 135–154.

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction.* London: Elsevier Ltd.

Krantz, D. H. (1975). Color measurement and color theory: I. Representation theorem for Grassmann structures. *Journal of Mathematical Psychology*, *12*, 283–303.

Logvinenko, A. D. (2005). Does luminance contrast determine lightness? *Spatial Vision*, *18*, 337–345.

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

Logvinenko, A. D., & Ross, D. A. (2005). Adelson's tile and snake illusions: A Helmholtzian type of simultaneous lightness contrast. *Spatial Vision*, *18*, 25–72.

Luce, R. D. (2004). Guest editorial. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 3–7). Mahwah, NJ: Lawrence Erlbaum Associates.

Luce, R. D., & Edwards, W. (1958). The derivation of subjective scales from just noticeable differences. *Psychological Review*, *65*, 222–237.

Mausfeld, R. (1998). Color perception: From Grassman codes to a dual code for object and illumination colors. In W. Backhaus, R. Kliegl, & J. S. Werner (Eds.), *Color vision: Perspectives from different disciplines* (pp. 219–250). Berlin: de Gruyter.

Mausfeld, R. (2002). The physicalistic trap in perception theory. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world.* Retrieved from `http://www.uni-kiel.de/psychologie/psychophysik/mausfeld/`

Mausfeld, R. (2003). 'Colour' as part of the format of two different perceptual primitives: The dual coding of colour. In R. Mausfeld & D. Heyer (Eds.), *Colour perception: Mind and the physical world* (pp. 381–436). New York: Oxford University Press.

Mausfeld, R. (2010). Colour within an internalistic framework: The role of 'colour' in the structure of the perceptual system. In J. Cohen & M. Matthen (Eds.), *Color Ontology and Color Science.* Retrieved from `http://www.uni-kiel.de/psychologie/psychophysik/mausfeld/`

Mausfeld, R., & Niederée, R. (1993). An inquiry into relational concepts on incremental principles of colour relational stimuli. *Perception*, *22*, 427–462.

Moore, B. C. J. (2008). *An introduction to the psychology of hearing.* Bingley, UK: Emerald.

Niederée, R. (1998). *Die Erscheinungsweisen der Farben und ihre stetigen Übergangsformen* (Habilitationsschrift). Kiel: Christian-Albrechts-Universität.

Niederée, R. (2010). More than three dimensions: What continuity considerations can tell us about perceived color. In J. Cohen & M. Matthen (Eds.), *Color Ontology and Color Science* (pp. 91–122). Cambridge, MA: MIT Press.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in python. *Journal of Neuroscience Methods*, *162*, 8–13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 1–8.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Purves, D., & Lotto, R. B. (2011). *Why we see what we do redux: A wholly empirical theory of vision*. Sunderland, MA: Sinauer Associates.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org`

Rach, S., & Colonius, H. (2008). Fechnerian scaling toolbox for MATLAB [Computer software manual]. Retrieved from `http://www.staff.uni-oldenburg.de/s.rach`

Ronacher, B., & Bautz, W. (1985). Human pattern recognition: Individually different strategies in analyzing complex stimuli. *Biological Cybernetics*, *51*, 249–261.

Sawayama, M., & Kimura, E. (2012). Local computation of lightness on articulated surrounds. *i-Perception*, *3*, 505–514.

Schneider, K. A. (2011). Attention alters decision criteria but not appearance: A reanalysis of Anton-Erxleben, Abrams, and Carrasco (2010). *Journal of Vision*, *11*(13), 1–8.

Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of Vision*, *8*(15), 1–10.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unkonwn distance function. I. *Psychometrika*, *27*, 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unkonwn distance function. II. *Psychometrika*, *27*, 219–246.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.

Shepard, R. N. (1974). Representaion of structure in similarity data: Problems and prospects. *Psychometrika*, *39*, 373–421.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Soranzo, A., & Agostini, T. (2006). Photometric, geometric, and perceptual factors in illumination-independent lightness constancy. *Perception & Psychophysics*, *68*, 102–113.

Trendtel, M., Ünlü, A., & Dzhafarov, E. N. (2010). Matrices satisfying Regular Minimality. *Frontiers in Quantitative Psychology and Measurement*, *1*, 1–6.

Ünlü, A., Kiefer, T., & Dzhafarov, E. N. (2009). Fechnerian scaling in R: The package fechner. *Journal of Statistical Software*, *31*(6), 1–24.

Velhagen, K., & Broschmann, D. (Eds.). (1985). *Tafeln zur Prüfung des Farbensinnes* (27th ed.). Stuttgart: Thieme.

Vladusich, T., Lucassen, M. P., & Cornelissen, W. (2007). Brightness and darkness as perceptual dimensions. *Computational Biology*, *3*, 1849–1858.

Volbrecht, V. J., & Kliegl, R. (1998). The perception of blackness: An historical and contemporary review. In W. Backhaus, R. Kliegl, & J. S. Werner (Eds.), *Color vision: Perspectives from different disciplines* (pp. 187–206). Berlin: de Gruyter.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, *38*, 310–324.

Wallach, H. (1963). The perception of neutral colors. *Scientific American*, *208*, 278–285.

Walraven, J. (1976). Discounting the background—the missing link in the explanation of chromatic induction. *Vision Research*, *16*, 289–295.

Webster, M. A. (2003). Light adaptation, contrast adaptation, and human colour vision. In R. Mausfeld & D. Heyer (Eds.), *Colour perception: Mind and the physical world* (pp. 67–114). New York: Oxford University Press.

Whittle, P. (1986). Increments and decrements: Luminance discrimination. *Vision Research*, *26*, 1677–1691.

Whittle, P. (1994). The psychophysics of contrast brightness. In A. Gilchrist (Ed.), *Lightness, brightness, and transparency* (pp. 35–110). Hillsday, NJ: Lawrence Erlbaum Associates.

Whittle, P., & Challands, P. D. C. (1969). The effect of background luminance on the brightness of flashes. *Vision Research*, *9*, 1095–1110.

Wickens, T. D. (2002). *Elementary signal detection theory.* New York: Oxford University Press.

Würger, S. M., Maloney, L. T., & Krauskopf, J. (1995). Proximity judgments in color space: Tests of a Euclidean color geometry. *Vision Research*, *35*, 827–835.

Wyszecki, G., & Stiles, W. S. (2000). *Color science: Concepts, methods, quantitative data and formulae.* New York: John Wiley & Sons.

Appendix A

Details on Experiments

A.1 Experiment I

In Experiment I, subjects performed five sessions of greater-less judgments. Stimulus 4 was presented as the standard and paired with all stimuli in random order. In each session, each of these pairs was presented 25 times resulting in 125 trials for each pair in total. Subject 1 participated in only two of the five greater-less sessions. The number of repetitions for each of these sessions can be seen in Table A.1.

Table A.1: Repetitions for Greater-Less Sessions for Subject 1

	0	1	2	3	4	5	6	7	8
0	0	0	0	0	24	0	0	0	0
1	0	0	0	0	30	0	0	0	0
2	0	0	0	0	29	0	0	0	0
3	0	0	0	0	28	0	0	0	0
4	25	19	18	21	48	30	26	22	24
5	0	0	0	0	22	0	0	0	0
6	0	0	0	0	36	0	0	0	0
7	0	0	0	0	23	0	0	0	0
8	0	0	0	0	21	0	0	0	0

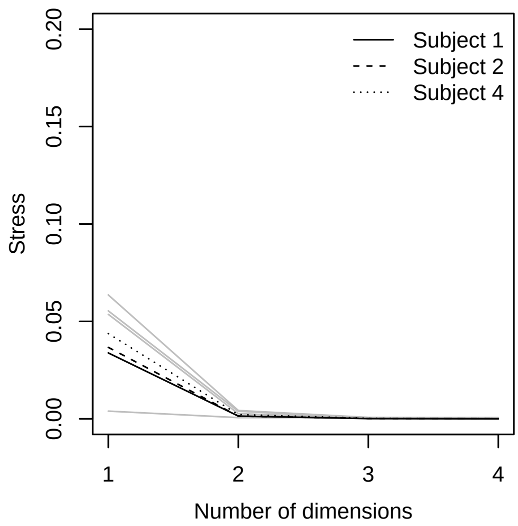


Figure A.1: Scree plot for multidimensional scaling on subset of 7 stimuli. Gray lines show stress for all subjects as seen and explained in Figure 5.9. Black lines show stress for a subset of 7 stimuli for Subjects 1, 2, and 4.

Data for Subjects 1, 2, and 4 were reanalyzed with a subset of 7 stimuli in order to evaluate the stress reduction found for Subject 3 (Figure A.1). It is obvious that the lower stress for Subject 3 cannot be attributed solely to the fact that stress was calculated for 7 instead of 9 stimuli.

A.2 Experiment II

In Experiment II, all stimulus pairs were presented 60 times over a course of 15 sessions. For Subject 1, stimulus pairs were presented 60 times on average. The exact number of repetitions for all pairs is depicted in Table A.2.

Tables A.3 to A.8 show discrimination probabilities after removing violations of regular minimality as explained in Section 5.2 and Fechnerian distances obtained for these discrimination probabilities.

Table A.2: Repetitions for Each Pair Presented in Experiment II for Subject 1

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	68	65	60	54	59	67	56	68	64	51	63	50	45
1	65	75	53	55	65	66	60	66	72	61	55	64	53
2	59	58	74	50	67	58	56	59	67	55	61	66	48
3	56	58	59	43	48	61	65	59	60	59	63	68	64
4	48	60	59	57	55	45	63	62	66	78	77	64	61
5	57	45	59	51	53	61	52	62	59	65	56	47	68
6	68	63	67	66	55	47	62	54	62	64	75	45	68
7	48	52	66	68	47	50	67	57	67	68	61	56	65
8	46	50	56	47	73	74	53	69	83	60	54	65	55
9	44	58	73	48	64	62	63	60	52	46	54	63	59
10	58	49	54	63	64	64	49	60	50	61	73	66	66
11	75	61	67	61	56	63	54	62	57	67	66	49	52
12	61	70	62	70	70	70	54	47	65	63	61	67	63

Table A.3: Discrimination Probabilities for Subject 1 without Violations of Regular Minimality

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.206	0.846	0.917	0.963	0.983	0.985	1.000	0.971	1.000	1.000	1.000	1.000	1.000
1	0.554	0.240	0.774	0.709	0.862	0.955	0.833	0.652	0.875	0.951	0.964	0.984	1.000
2	0.932	0.828	0.284	0.780	0.761	0.914	0.982	0.898	1.000	0.982	1.000	0.985	1.000
3	0.946	0.500	0.627	0.233	0.271	0.492	0.492	0.644	0.800	0.881	0.984	0.941	1.000
4	0.979	0.717	0.475	0.298	0.218	0.444	0.619	0.823	0.833	0.872	0.961	0.828	1.000
5	0.982	0.889	0.610	0.490	0.302	0.246	0.519	0.774	0.729	0.738	0.964	0.723	1.000
6	0.971	0.603	0.851	0.455	0.600	0.383	0.210	0.259	0.339	0.500	0.787	0.622	1.000
7	0.979	0.596	0.848	0.706	0.830	0.800	0.478	0.088	0.388	0.456	0.508	0.946	0.985
8	0.957	0.780	0.964	0.851	0.877	0.716	0.434	0.116	0.024	0.183	0.204	0.785	0.945
9	0.977	0.879	0.959	0.896	0.859	0.694	0.317	0.350	0.135	0.065	0.278	0.556	0.932
10	1.000	0.939	0.981	0.952	0.938	0.969	0.857	0.500	0.320	0.393	0.027	0.970	0.985
11	1.000	0.967	0.955	0.902	0.839	0.698	0.833	0.935	0.825	0.672	0.848	0.204	0.692
12	1.000	1.000	1.000	0.986	1.000	0.986	0.926	0.979	0.985	0.905	0.918	0.791	0.270

Table A.4: Discrimination Probabilities for Subject 3 without Violations of Regular Minimality

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.483	0.650	0.700	0.817	0.867	0.900	0.883	0.883	0.900	0.933	0.767	1.000	1.000
1	0.667	0.633	0.717	0.750	0.867	0.900	0.883	0.717	0.783	0.900	0.683	0.983	0.900
2	0.733	0.667	0.500	0.717	0.617	0.700	0.667	0.783	0.683	0.800	0.717	0.933	0.983
3	0.867	0.700	0.633	0.600	0.617	0.700	0.750	0.733	0.683	0.800	0.767	0.883	0.967
4	0.767	0.733	0.717	0.700	0.600	0.767	0.683	0.817	0.717	0.783	0.650	0.917	0.967
5	0.900	0.833	0.700	0.800	0.700	0.667	0.750	0.700	0.683	0.750	0.717	0.917	0.933
6	0.783	0.683	0.833	0.650	0.750	0.800	0.600	0.717	0.683	0.767	0.800	0.917	0.917
7	0.883	0.650	0.767	0.800	0.683	0.850	0.667	0.617	0.650	0.733	0.633	0.917	0.983
8	0.767	0.883	0.800	0.783	0.683	0.833	0.767	0.650	0.633	0.783	0.733	0.883	0.983
9	0.983	0.883	0.850	0.767	0.700	0.750	0.633	0.750	0.783	0.617	0.767	0.883	0.967
10	0.850	0.650	0.767	0.783	0.733	0.833	0.733	0.633	0.650	0.717	0.550	0.800	0.967
11	0.933	0.900	0.850	0.833	0.883	0.783	0.800	0.917	0.783	0.900	0.800	0.767	0.933
12	0.900	0.917	0.883	0.900	0.883	0.850	0.850	0.900	0.933	0.883	0.867	0.817	0.800

Table A.5: Discrimination Probabilities for Subject 4 without Violations of Regular Minimality

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.117	0.267	0.567	0.633	0.750	0.883	0.950	0.850	0.967	0.983	0.900	1.000	1.000
1	0.450	0.100	0.550	0.333	0.567	0.650	0.433	0.400	0.483	0.750	0.400	1.000	0.983
2	0.433	0.350	0.150	0.367	0.283	0.467	0.800	0.767	0.817	0.933	0.883	0.933	1.000
3	0.700	0.217	0.483	0.183	0.200	0.300	0.200	0.250	0.550	0.650	0.717	0.900	1.000
4	0.683	0.467	0.317	0.200	0.133	0.367	0.283	0.383	0.567	0.633	0.700	0.833	0.983
5	0.900	0.650	0.633	0.283	0.333	0.133	0.200	0.500	0.433	0.550	0.650	0.500	0.967
6	0.967	0.650	0.883	0.567	0.467	0.383	0.167	0.200	0.200	0.183	0.200	0.550	0.967
7	0.950	0.567	0.917	0.650	0.617	0.583	0.450	0.083	0.133	0.183	0.200	0.850	0.967
8	1.000	0.750	0.967	0.783	0.783	0.700	0.400	0.150	0.117	0.167	0.150	0.767	0.933
9	1.000	0.933	0.917	0.800	0.833	0.700	0.467	0.367	0.167	0.150	0.200	0.517	0.850
10	0.933	0.783	0.917	0.850	0.933	0.800	0.683	0.267	0.233	0.300	0.100	0.683	0.850
11	0.983	0.983	0.950	0.933	0.833	0.733	0.700	0.667	0.550	0.533	0.517	0.217	0.667
12	1.000	0.967	1.000	1.000	1.000	0.983	0.967	0.950	0.883	0.867	0.767	0.683	0.217

Table A.6: Fechnerian Distances for Subject 1 after Removing Violations of Regular Minimality

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0	0.954	1.359	1.338	1.439	1.497	1.501	1.587	1.708	1.706	1.767	1.590	1.524
1		0	1.077	0.737	0.855	1.118	0.987	0.920	1.235	1.357	1.617	1.508	1.490
2			0	0.791	0.734	0.942	1.184	1.375	1.569	1.436	1.670	1.452	1.446
3				0	0.118	0.395	0.505	0.927	1.044	1.045	1.432	1.279	1.483
4					0	0.282	0.569	1.009	1.108	1.110	1.496	1.160	1.512
5						0	0.447	0.886	0.986	0.918	1.374	0.972	1.470
6							0	0.439	0.539	0.541	0.927	1.042	1.446
7								0	0.392	0.529	0.805	1.431	1.603
8									0	0.229	0.472	1.187	1.626
9										0	0.579	0.958	1.502
10											0	1.501	1.606
11												0	1.009
12													0

Table A.7: Fechnerian Distances for Subject 3 after Removing Violations of Regular Minimality

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0	0.200	0.417	0.417	0.467	0.483	0.417	0.317	0.367	0.500	0.350	0.550	0.533
1		0	0.250	0.217	0.267	0.283	0.217	0.117	0.167	0.300	0.150	0.350	0.350
2			0	0.250	0.233	0.233	0.250	0.317	0.317	0.383	0.350	0.467	0.500
3				0	0.117	0.233	0.150	0.200	0.200	0.250	0.250	0.350	0.450
4					0	0.200	0.150	0.200	0.167	0.267	0.200	0.350	0.433
5						0	0.233	0.233	0.217	0.217	0.317	0.267	0.300
6							0	0.150	0.150	0.183	0.250	0.350	0.367
7								0	0.050	0.233	0.100	0.300	0.350
8									0	0.233	0.133	0.267	0.350
9										0	0.300	0.367	0.417
10											0	0.283	0.417
11												0	0.183
12													0

Table A.8: Fechnerian Distances for Subject 4 after Removing Violations of Regular Minimality

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0	0.500	0.733	0.767	0.850	1.033	1.150	1.267	1.300	1.267	1.267	1.550	1.667
1		0	0.633	0.267	0.350	0.533	0.650	0.767	0.800	0.800	0.933	1.333	1.633
2			0	0.400	0.317	0.667	0.700	0.967	1.000	1.000	1.067	1.333	1.633
3				0	0.083	0.267	0.383	0.617	0.650	0.650	0.783	1.067	1.517
4					0	0.350	0.383	0.650	0.683	0.683	0.817	1.083	1.583
5						0	0.283	0.600	0.600	0.600	0.733	0.883	1.550
6							0	0.367	0.317	0.317	0.450	0.867	1.467
7								0	0.083	0.150	0.250	0.833	1.500
8									0	0.067	0.167	0.750	1.417
9										0	0.233	0.683	1.350
10											0	0.850	1.300
11												0	0.917
12													0