

**UC Berkeley**

**UC Berkeley Electronic Theses and Dissertations**

**Title**

Binocular Contrast Perception and Its Application to Display Technologies

**Permalink**

<https://escholarship.org/uc/item/5554v5p3>

**Author**

Wang, Minqi

**Publication Date**

2023

Peer reviewed|Thesis/dissertation

Binocular Contrast Perception and Its Application to Display Technologies

By

Minqi Wang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Vision Science

in

the Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Emily Cooper, Chair

Professor Dennis Levi

Professor Martin Banks

Spring 2023

© Copyright 2023, Minqi Wang

# Abstract

Binocular Contrast Perception and Its Application to Display Technologies

by

Minqi Wang

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor Emily Cooper, Chair

Due to the lateral separation of our eyes, the human visual system takes in two slightly different views of the world. These two views must be combined by the visual system into a single percept. Importantly, binocular combination can be challenged when the left and right eye views differ from each other, for example, in luminance, contrast or color. Failures of binocular combination can result in double vision, losses in-depth perception, and troublesome visual appearances like binocular rivalry. With advances in wearable binocular display technologies for virtual and augmented reality, the subjective appearances associated with artificial binocular differences introduced by these displays warrant in depth evaluation. This dissertation contains a body of work characterizing the subjective appearance of images in which different contrasts are presented to the two eyes. The work herein builds on seminal studies in vision science that paved the way for our understanding of binocular appearance, but primarily studied simplistic visual stimuli such as grating patterns that differ substantially from the structured and complex visual inputs that our visual system evolved for. In Chapter 1, I extend on prior work and examine percepts associated with a range of stimulus patterns with various degrees of complexity to deepen our understanding of binocular combination. In the remaining chapters, I leverage these insights to examine the implications of binocular appearance for three key applications in computer graphics and binocular display design: display imperfections, tonemapping, and field of view coverage. Wearable display technologies will advance the way that we interact with digital content and with the world at large, but their utility and impact are contingent on whether or not they can integrate well with our natural vision. Thus, in each chapter I discuss design guidelines based on the perceptual results to aid the development of better binocular displays that prevents undesirable visual artifacts, and exploit binocular vision to loosen technical specifications.

# Acknowledgments

First, I would like to thank my parents: to my dad who has fostered my curiosity since I was a kid, which has led to my interest in science and research to this day; to my mom who has created the opportunity for me to envision a different future.

My PhD journey could not be accomplished alone without support from others. I would like thank the Vision Science community, including faculties, administrative staffs, post-docs, and students. I would like to thank my committee, Dennis and Marty, for providing feedback throughout my PhD research and offering their time and knowledge.

Special thanks to Iona McLean and other lab members who have created a lively lab environment that I feel a strong social belonging to. I would also like to thank the joint lab lunch crew on Minor Beach who I have shared engaging and humorous in-person conversations with, which are much needed after the pandemic.

Lastly, I would like to thank the tremendous support that I have received in the past years from my mentor Emily, who has helped me grow both professionally and personally. Emily has taught me not just about vision and the brain, but also organization skills, troubleshooting, decision making, communication, and interacting with others.

I know I will miss my time at Berkeley, a place where I have broadened my scientific knowledge and views of the world through interactions with the people here and learning about different perspectives. This period of time in my life will always be an important part of my memory.

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 The effect of spatial structure on binocular contrast perception</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 Methods . . . . .	6
1.2.1 Participants . . . . .	6
1.2.2 Setup . . . . .	6
1.2.3 Procedure . . . . .	7
1.2.4 Contextual Modulation . . . . .	8
1.2.5 Stimulus Types . . . . .	9
1.3 Data Analysis . . . . .	11
1.3.1 Outlier Criteria . . . . .	11
1.3.2 Data Transformation . . . . .	11
1.3.3 Statistical Tests . . . . .	12
1.4 Results . . . . .	13
1.4.1 Experiment 1 . . . . .	13
1.4.2 Experiment 2 . . . . .	14
1.4.3 Stimulus Orientation and Edge Blur . . . . .	16
1.4.4 Eye Dominance . . . . .	16
1.5 Discussion . . . . .	18
1.6 Conclusion . . . . .	22
1.7 Supplemental Material . . . . .	22
<b>2 The effect of interocular contrast differences on the appearance of augmented reality imagery</b>	<b>25</b>
2.1 Introduction . . . . .	26
2.2 Related Work . . . . .	27
2.2.1 Brightness . . . . .	27
2.2.2 Contrast . . . . .	28
2.2.3 Luster . . . . .	29
2.2.4 Binocular Rivalry . . . . .	29
2.2.5 Depth . . . . .	29
2.2.6 Modelling Dichoptic Percepts . . . . .	30
2.3 Perceptual Experiments . . . . .	30
2.3.1 Participants . . . . .	31
2.3.2 Experimental Setup . . . . .	31
2.3.3 Task . . . . .	31

2.3.4	Stimuli . . . . .	32
2.4	Results . . . . .	35
2.4.1	Experiment 1 . . . . .	35
2.4.2	Experiment 2 . . . . .	39
2.5	Discussion . . . . .	42
2.5.1	Predicting Perceptual Artifacts in Dichoptic AR Stimuli . . . . .	42
2.5.2	Modeling Interocular Differences and Their Effect on Perception . . . . .	43
2.5.3	Potential Benefits of Dichoptic Contrast in AR . . . . .	43
2.5.4	Potential Effects of Stimulus Motion . . . . .	44
2.6	Conclusion . . . . .	44
<b>3</b>	<b>A re-examination of dichoptic tone mapping</b>	<b>45</b>
3.1	Introduction . . . . .	45
3.2	Related Work . . . . .	47
3.2.1	Binocular Combination . . . . .	47
3.2.2	Dichoptic Tone Mapping Methods . . . . .	48
3.2.3	Current Study Motivation . . . . .	50
3.3	Exposure-Based Dichoptic Method . . . . .	50
3.4	User Studies . . . . .	51
3.4.1	Stimulus Generation . . . . .	52
3.4.2	Experiments 1, 2, and 3: Preference, Subjective Detail Visibility, and 3D Impression Ratings . . . . .	54
3.4.3	Experiments 4 and 5: Preference and Subjective Detail Visibility Forced Choice Comparisons . . . . .	58
3.4.4	Experiment 6: Objective Detail Visibility . . . . .	62
3.5	Discussion . . . . .	67
3.5.1	Relationship to Prior Work . . . . .	67
3.5.2	Binocular Contrast Perception in Natural Images . . . . .	68
3.5.3	3D Imagery . . . . .	69
3.5.4	Objective Visual Performance-Based Metric . . . . .	69
3.6	Conclusion . . . . .	70
3.7	Supplemental Material . . . . .	70
3.7.1	Binocular Rivalry and Eye Dominance (Experiments 1, 2, & 3) . . . . .	70
3.7.2	Scene Sample Analysis (Experiments 4 & 5) . . . . .	72
3.7.3	Effect of Contrast on Objective Task Performance (Experiment 6) . . . . .	73
<b>4</b>	<b>Perceptual guidelines for optimizing field of view in stereoscopic augmented reality displays</b>	<b>75</b>
4.1	Introduction . . . . .	75
4.2	Related Work . . . . .	77
4.2.1	The Natural Human Field of View . . . . .	77
4.2.2	Field of View in Stereoscopic Near-Eye Displays . . . . .	78
4.2.3	Nonuniformity Artifacts from Partial Field of View Overlap . . . . .	79
4.3	Perceptual Studies . . . . .	81
4.3.1	Methods. . . . .	81
4.3.2	Experiment 1. . . . .	83
4.3.3	Experiment 2. . . . .	86
4.4	3D Field of View in Augmented Reality . . . . .	89

4.4.1	Viewing Geometry and Camera Frusta . . . . .	89
4.4.2	Qualitative Horizontal Field-of-View Analysis . . . . .	90
4.4.3	Quantitative Horizontal Field-of-View Analysis . . . . .	91
4.5	Discussion . . . . .	94
4.5.1	Convergent and Divergent Configurations . . . . .	94
4.5.2	Predicting Fading in 3D AR Experiences . . . . .	95
4.5.3	Monocular Regions in Natural Vision . . . . .	96
4.5.4	Vertical Partial Overlap . . . . .	97
4.5.5	Limitations and Future Work . . . . .	97
4.6	Conclusion . . . . .	98
<b>Discussion</b>		<b>99</b>
4.6.1	Binocular Contrast Combination Depends on Spatial Structure . .	99
4.6.2	Predicting Secondary Perceptual Effects . . . . .	101
4.6.3	Eye Dominance . . . . .	101
4.6.4	Binocular Combination at Large Eccentricities . . . . .	103
4.6.5	Conclusion . . . . .	105

# INTRODUCTION

Humans, like many other animals, have two eyes that serve as sensors to gather light information. The signals from the two eyes travel down separate optic nerves to the brain, where they are integrated to generate a single visual percept of the world. According to several converging lines of vision research, this binocular combination leverages two types of interactions between the eyes: suppression and facilitation. The dynamics of these interactions, however, are dependent on the luminance and contrast of the visual input. If we close one eye in a typical setting, for example, we do not see the world suddenly change in brightness. However, for a stimulus at threshold, detection is better with binocular viewing than monocular viewing [17, 21, 27]. This binocular enhancement in detection (sometimes called binocular summation) serves as evidence that the two eyes' inputs facilitate each other when signals are weak [43, 54, 102]. At supra-threshold luminance levels, however, it appears that suppression dominates [17, 43, 54]. In the example of closing one eye, the open eye suppresses the closed eye.

These interactions illustrate how the binocular visual system has evolved clever adaptation mechanisms to serve the needs of vision in different circumstances. Investigations of binocular interactions in which the luminance, contrast, or visual quality differs between the two eyes form the foundation of our understanding of binocular combination in the brain [75, 99, 170]. Here, I use the term *dichoptic* to refer to a stimulus that has differences between the left and right eye images, including when the two eyes receive different supra-threshold inputs or monocular input. When the two eyes receive images with different contrast levels (dichoptic contrast), the general rule found in prior perceptual studies is that the eye seeing higher contrast will dominate and be the primary determinant of the binocular appearance [39, 103, 104]. This phenomenon is commonly referred to as 'winner-take-all' because the eye seeing lower contrast (the 'loser') is suppressed and does not affect the binocular appearance. This established finding had been extensively tested using simple dichoptic contrast stimuli such as sine wave gratings and geometric shapes [10, 39, 42, 103, 104].

However, much less is known about the the binocular appearance of more naturalistic dichoptic stimuli, for example, with added pattern complexity and potential contextual interactions. This remains an important question to study because our visual system operates in a complex but structured visual world. The very definition of 'winner-take-all' gets complicated when we leave the realm of gratings and shapes viewed in isolation. In this dissertation, I present a body of work characterizing the validity of the 'winner-take-all' rule for determining binocular appearance in naturalistic stimuli using human visual psychophysics. I first examine this issue using semi-natural stimuli and techniques that are closely related to conventional psychophysical studies (Chapter 1). With the prevalence of binocular display technologies for virtual and augmented reality (VR/AR), however,

understanding binocular appearance in fully natural imagery is particularly pertinent at this moment in time. Thus, in the remaining chapters I expand on this work and adopt customized stimuli derived from natural images to examine targeted applications in VR/AR technology (Chapters 2, 3, 4). Below I summarize the focus and findings of each chapter.

In Chapter 1, I characterize binocular contrast perception when people view stimuli with varying spatial pattern complexity embedded within different visual contexts. I provide evidence that spatial structure plays an important role in modulating interocular suppression and binocular appearance. In particular, I show that the surrounding context is a key modulator of the 'winner-take-all' rule. This modulation affects spatially irregular patterns, such as noise patterns, less than simple grating stimuli. Since natural images contain a rich repertoire of spatial structures that are highly variable, this result motivates the importance of using different stimulus patterns in basic binocular vision research, and specific stimuli that emulate the intended use case when generating insights for display applications. In addition to these stimulus-based observations, it is important to point out that the perceptual matching methodology used to probe binocular appearance in many studies [12, 42, 74, 103, 104], including the ones in Chapter 1, is limited. This methodology relies on appearance matching to determine the subjective appearance of the visual stimulus. That is, the experimenter cannot directly measure what the dichoptic stimulus looks like to the observer so instead they ask people to try to match the appearance with another stimulus, usually a binocular non-dichoptic or monocular one. Matching paradigms have shortcomings because experimenters have to predetermine the stimuli that might match with the dichoptic one, but dichoptic images are known to create unique appearances that are challenging to match non-dichoptically. Some examples of such appearances are stereopsis (sense of depth), binocular luster (shiny appearance), and binocular rivalry (change in appearance with fixed visual inputs). These aspects of binocular appearance can be difficult to capture with existing matching paradigms, but are nonetheless important to understand for designing good binocular display systems that have a challenging design space. In the following chapters, I therefore adopt both new stimulus types and new response paradigms to investigate several dichoptic scenarios relevant to modern stereoscopic systems and use perceptual data to determine new display guidelines in each case.

In Chapter 2, I analyze the qualitative appearance of dichoptic stimuli captured by a set of descriptive measures. I introduce customized stimuli to explicitly emulate augmented reality (AR) experiences with a see-through display superimposed over a natural background and manipulate these stimuli to have different luminance and contrast levels between the two eyes. I modify the conventional matching response paradigm to simultaneously measure multiple aspects of binocular appearance within a single stimulus. The results from this new paradigm demonstrate how a 'winner-take-all' contrast perception model is inadequate to capture the multi-faceted appearance of dichoptic AR imagery, and provide guidelines for predicting the appearance of dichoptic artifacts in AR stimuli.

In Chapter 3, I report an in-depth comparison of several dichoptic tonemapping methods, in which the contrast of natural imagery differ in the two eyes [174, 182, 183]. Such methods have been proposed to increase the perceived contrast and visual quality in binocular displays such as AR/VR systems. Using both subjective and performance-based measures, I present the comparative impact of these complex, spatially-varying binocular

contrast differences on image quality.

In Chapter 4, I address whether we can reduce the binocular suppression of AR content and expand the field of view in binocular AR display systems. I first characterize the trade-off between the field of view and image quality in a binocular head-mounted display system. I then propose a simple display guideline based on the fixation distance geometry, display configuration, and the perceptual results to reduce suppression artifacts in binocular AR displays.

Each chapter presents a standalone manuscript that is either published [161, 162, 163] or under peer review (Chapter 2).

# CHAPTER 1

## The effect of spatial structure on binocular contrast perception

To obtain a single percept of the world, the visual system must combine the inputs from the two eyes. Understanding the principles that govern this binocular combination process has important real-world clinical and technological applications. However, most research examining binocular combination has relied on simple visual stimuli and it is unclear how well the findings apply to real-world scenarios. For example, it is well known that when the two eyes view sine wave gratings with differing contrast (dichoptic stimuli), the binocular percept often matches the higher contrast grating. Does this “winner-take-all” property of binocular contrast combination apply to more naturalistic imagery, which include broadband structure and spatially varying contrast? To better understand binocular combination during naturalistic viewing, we conducted psychophysical experiments characterizing binocular contrast perception for a range of simple and complex visual stimuli. In two experiments, we measured binocular contrast perception of dichoptic sine wave gratings and naturalistic stimuli and asked how the contrast of the surrounding context affected percepts. Binocular contrast percepts were close to winner-take-all across many of the stimuli when the surrounding context was the average contrast of the two eyes. However, we found that changing the surrounding context modulated the binocular percept of some patterns and not others. We show that this contextual effect may be due to the spatial orientation structure of the stimuli. These findings provide a step towards understanding binocular combination in the natural world and highlight the importance of considering the effect of the spatial interactions in complex stimuli.

### 1.1 Introduction

Our binocular perception of the world is not just a simple average of what the two eyes see. For example, closing one eye does not reduce the apparent brightness of the world by half. A complex, hierarchical network of neural circuits is involved in performing binocular combination, and many properties of binocular interactions develop through our visual experience of the natural world (see [15] for review).

Laboratory studies in which the inputs to the two eyes are made to be binocularly discrepant (i.e., dichoptic) have been highly influential for characterizing both the general principles by which the visual system integrates information from the two eyes and the

importance of experience with natural visual stimulation for normal binocular function [99, 170]. At the same time, understanding binocular combination of dichoptic imagery has important clinical implications in areas such as amblyopia, in which there are disruptions in the balance of suppression and facilitation between the eyes [38]. Beyond amblyopia, there is a range of real-world situations in which image quality is different in the two eyes. For example, a difference in refractive error, or a unilateral cataract or scotoma, can result in different levels of contrast between the two eyes. How do these clinical interocular image differences influence binocular percepts during daily life? There are also emerging stereoscopic display methods that intentionally introduce luminance and contrast differences between the two eyes to create novel graphical effects. However, assessments of the efficacy of these techniques have produced mixed results, suggesting that our existing understanding of binocular combination may be insufficient to account for binocular percepts of this natural dichoptic imagery [161, 174, 181, 182, 183].

Common stimuli used to study binocular combination of dichoptic imagery include simple shapes or gratings with different luminance, colors, or contrast between the two eyes. Under most conditions, the existing literature tends to find a “winner-take-all” pattern, in which the binocular percept of the dichoptic stimulus is dominated by eye seeing the “stronger” stimulus – that is, the brighter, more saturated, or higher contrast stimulus [39, 42, 74, 81, 103, 104]. However, there are situations in which the “weaker” stimulus contributes more to the binocular percept, resulting in a percept that is closer to the average of the two eyes or even “loser-take-all”. For example, binocular luminance combination depends on the luminance level being tested, and some luminance levels result in averaging rather than winner-take-all [12, 42]. In addition, adding a contour to the lower luminance eye shifts the binocular percept towards that eye [42, 104]. In addition, studies of dichoptic color perception and dichoptic masking have shown that adding luminance contrast to one or both eyes can shift color perception towards averaging and can reduce masking effects [76, 79, 81].

While these studies have greatly advanced our understanding of binocular combination, laboratory stimuli often have limited visual complexity and do not capture the rich spectral and contextual information available in the natural environment. For example, natural images are a specific subset of all possible visual stimuli and contain statistical regularities that the brain exploits during visual encoding [146]. One common property of natural images is that they tend to have a broadband spatial frequency composition with amplitude spectra that fall off predictably as a function of frequency [52]. This property is important to consider for binocular combination of natural imagery because there is evidence that binocular interactions are different across different spatial frequencies [7]. However, it is unknown whether and how these spatial frequency regularities influence how the visual system combines naturalistic stimuli to form a coherent binocular percept.

When compared to natural viewing, a second key element that is missing in many prior studies of binocular combination is the influence of surrounding context. Contextual effects are ubiquitous in perception and are thus ultimately essential for understanding binocular percepts during natural vision. For example, the same patch of gray can be perceived to be a different shade depending on whether it is on a dark or light background [3] and the same grating can appear to have different contrast depending on its surrounding context [128, 172]. An effect known as surround suppression, in which a pattern appears to have higher or lower contrast depending on the visual similarity of the surrounding area, can be

observed both in early cortical neurons and psychophysically [172]. Importantly, human behavioral data suggests that surround suppression can also occur when the target is in one eye while the context is shown to the other eye, suggesting a binocular interaction [142]. It has recently been demonstrated that surrounding context can also reduce the “winner-take-all” effect in binocular contrast perception, as well as related phenomena such as dichoptic masking [62, 76]. For example, if there is a monocular contour around a low contrast sine wave grating (induced by a phase shift with its surround), the lower contrast sine wave grating was shown to dominate the dichoptic contrast percept in a “loser-take-all” manner [62]. However, our understanding of contextual effects in dichoptic perception remains incomplete.

In this study, we investigated the effect of surrounding context on binocular combination with noise and natural texture patterns that have rich spatial content and compared the results to those obtained with simple grating patterns. We present the findings from two experiments. In Experiment 1, we examine how different surrounding contexts influence binocular contrast perception of gratings, 1/f noise, and natural textures. In Experiment 2, we further investigate which properties of the stimulus patterns may contribute to differences in binocular contrast perception, and we find that contextual effects are strongest when the spatial orientation structure is continuous between the stimulus and its surrounding context. Our results suggest that binocular contrast combination during natural viewing depends on both the spatial structure of the imagery, as well as the visual similarity with the surrounding context.

## 1.2 Methods

### 1.2.1 Participants

Experiment 1 had ten participants (ages 22 - 26 years, all female), and Experiment 2 had 34 participants (ages 19 - 32 years, 25 female). Two participants from Experiment 2 were later excluded from the analysis (see outlier criteria in the data analysis section). All participants had normal or corrected-to-normal visual acuity and stereo vision. The targeted participant sample size in Experiment 2 was increased by a factor of 3 from Experiment 1 to increase statistical power, but no formal power analysis was used. The experimental procedures were approved by the University of California, Berkeley Institutional Review Board and were consistent with the tenets of the Declaration of Helsinki. Participants were compensated for their time.

### 1.2.2 Setup

All stimuli were presented on two linearized liquid crystal displays (LCDs) (LG 32UD99-W, 3840 x 2160 pixels, maximum luminance of 138cd/m<sup>2</sup>) and viewed through mirror haploscope as shown in Figure 1.1. The stimuli were generated using Matlab (Mathworks, Inc.) and Psychtoolbox [83, 132? ]. During the experiment, participants sat in a dark room and viewed the displays at a distance of 63cm with their head stabilized by a chin rest.

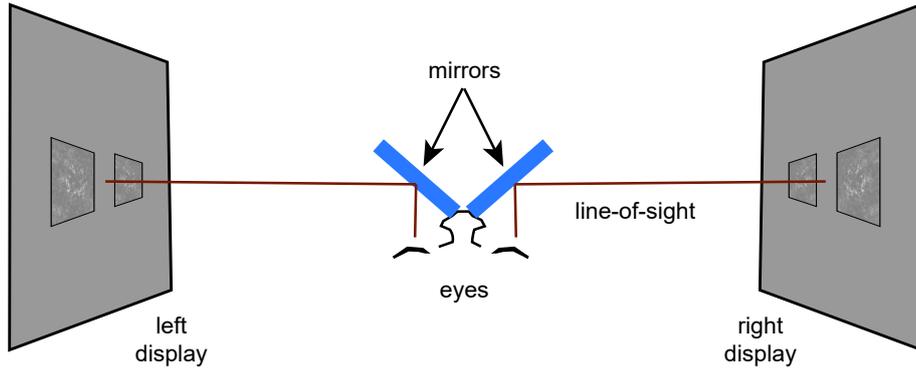


Figure 1.1: Top-down view of the haploscope used to present stimuli independently to the two eyes.

### 1.2.3 Procedure

On each trial, two images were presented on each monitor (four total). Each image consisted of a  $2^\circ$  diameter circular region embedded in a surrounding  $4^\circ \times 4^\circ$  square region (Figure 1.2A). When participants viewed these stimuli binocularly, the four images were fused into two: one image positioned in the upper half of the screen and the other positioned in the lower half of the screen (Figure 1.2B). One pair of images comprised the reference stimulus and the other pair comprised the adjustable stimulus (i.e., the test stimulus). The positions of the reference and adjustable stimuli were swapped for half of the participants (i.e., half of the participants adjusted the lower stimulus and half adjusted the upper stimulus).

The participant’s task was to change the physical contrast of the adjustable stimulus’s central region until it perceptually matched, as closely as possible, the reference stimulus’s central region. There was no time limit for the matching, and participants were able to look back and forth between the two stimuli to compare. On each trial, the contrast of the reference shown to the left and right eyes were selected from a predetermined set. During most trials, these contrasts differed between the two eyes – that is, the reference stimulus was dichoptic. There were some trials in which the reference stimulus was non-dichoptic. These primarily served as catch trials. On the other hand, the adjustable stimulus always had the same contrast in both eyes (non-dichoptic). The reference stimulus and the adjustable stimulus always shared the same physical contrast in the surround region.

Stimulus contrast in the central region was adjusted by applying a multiplicative scale factor to an original image with 100% contrast. Image contrasts were adjusted according to the following formula:

$$N = c(M - u) + u, \quad (1.1)$$

where  $N$  is the contrast adjusted image,  $c$  is the contrast adjustment scalar,  $M$  is the original full contrast image, and  $u$  is the mean pixel intensity of the original image. For grating stimuli, this adjustment was equivalent to adjusting the Michelson contrast. The initial contrast of the adjustable stimulus was randomly selected between levels 0 and 1, and the participants made adjustments in steps of 0.05.

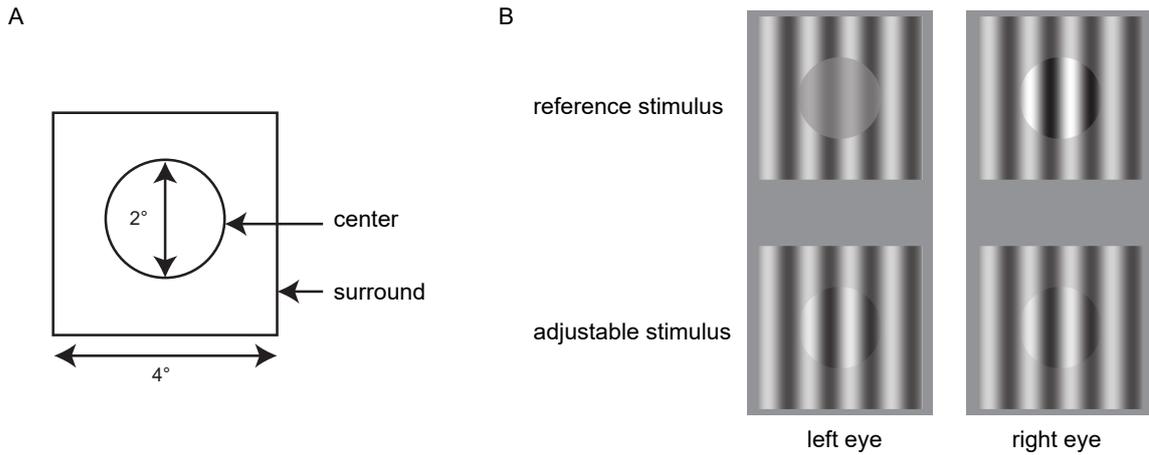


Figure 1.2: A) Each image had a center-surround layout, such that the contrast of the center and surround could differ. B) In each trial, two stimuli (four images) were shown. The dichoptic reference stimulus was fixed and had different contrasts between the two eyes except during catch trials. Participants increased or decreased the contrast of the central region of the non-dichoptic adjustable stimulus to match the appearance of the reference stimulus. The surround contrast was always the same in both eyes and both stimuli.

### 1.2.4 Contextual Modulation

The surround contrasts were fixed on each trial and not adjustable by the participant. To examine the effects of surrounding context, we explored three different surround conditions (Figure 1.3). In Experiment 1, three contextual modulations were considered: 1) mean surround, 2) low surround, and 3) high surround. In Experiment 2, mean surround and high surround conditions were tested on a new set of stimulus patterns (see Stimulus Types section).

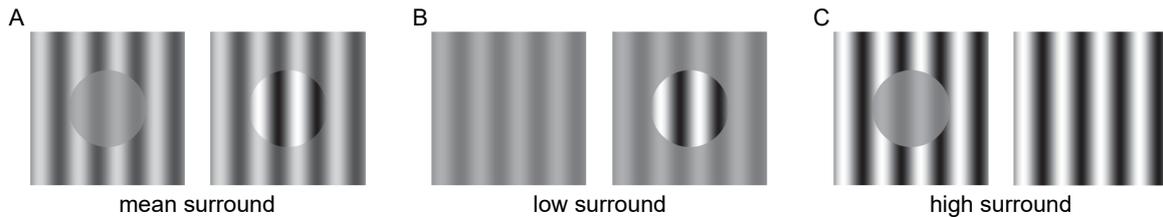


Figure 1.3: A pair of dichoptic reference images shown to the left and right eye in the: A) mean surround condition (binocular edge), B) low surround condition (monocular edge in the eye viewing higher contrast target), and C) high surround condition (monocular edge around the lower contrast target).

In the mean surround condition, the surround contrast of all four images on a given trial was equal to the mean contrast of the reference's central regions (Figure 1.3A). In the low surround condition, the contrast of all images' surround was equal to the lower contrast of the dichoptic reference images (Figure 1.3B). Finally, in the high surround condition, the contrast of the surround was equal to the higher contrast of the dichoptic reference images (Figure 1.3C). It is worth noting that in both the low and high surround conditions there is a visible edge around the circular target that is present in one eye but not the other. We hypothesized that this contrast edge serves a similar purpose to the monocular contours in prior dichoptic luminance combination studies, where the presence of the contour in one eye boosts the contribution of that eye's image in binocular combination [42, 62, 76, 104]. In this sense, the mean surround condition is more similar to the existing literature in

which there is a contour in both eyes’ images and the effect of contours in the two eyes may cancel each other out and become less relevant to shaping binocular interaction. On the other hand, a monocular stimulus commonly used in the literature (i.e., one eye sees uniform gray and the other eye sees a target with some contrast) would be most similar to a trial in the low surround condition with zero contrast in the surround [39, 43, 74].

## 1.2.5 Stimulus Types

Another goal of this project was to assess whether rules of binocular combination of simple stimuli can be generalized to more complex stimuli across different contexts. In Experiment 1, we examined four types of stimuli: a low spatial frequency sine wave grating (1 cycle per degree (cpd)), a higher spatial frequency sine wave grating (5cpd), natural image patches, and 1/f 2D noise (Figure 1.4). The grating stimuli were vertical gratings (Figure 1.4A). We selected three natural image patches from the McGill Calibrated Image Dataset as the natural image stimuli [124]. These images contained natural textures, such as foliage, without obvious recognizable objects or scenes to reduce the influence of higher perceptual organization on perceived contrast (Figure 1.4B). The noise stimuli were generated by phase scrambling the natural image patches so that the amplitude spectra matched closely with the natural images. We also included a synthetic 1/f noise stimulus with an amplitude spectrum slope of -1 in the log-log space and a Gaussian intensity distribution (Figure 1.4C). All stimuli were greyscale. In summary, we tested all three surround conditions (mean, low, high) on the nine stimulus patterns (2 gratings, 4 noise, 3 natural images). For each stimulus type and surround condition, the contrast level of the center region of the reference was set to five pre-selected contrast levels for each eye (0, 0.25, 0.5, 0.75 and 1), resulting in 25 contrast combinations between the two eyes. In total, there were 675 trials with no repeats.

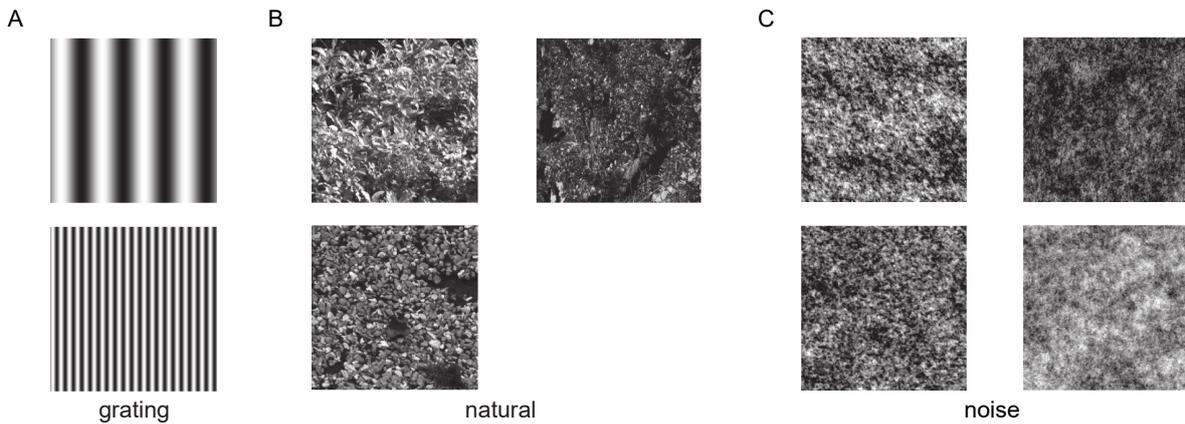


Figure 1.4: Stimulus types used in Experiment 1, each shown with center and surround both having contrast of 1. Stimuli included: A) sine wave gratings (1 cycle per degree and 5 cycles per degree), B) natural textures (with amplitude spectra slopes of -0.9, -0.7, and -0.9 on a log-log scale [124]), and C) 1/f noise. Three of the 1/f noise images were generated by phase scrambling the natural textures, the fourth (bottom right) was synthesized to have a slope of -1 and a Gaussian intensity histogram.

In Experiment 2, we aimed to replicate and extend the findings from Experiment 1. We only included the mean surround and high surround conditions. We included the 5cpd sine wave grating and the synthetic 1/f noise from Experiment 1 as baselines in Experiment 2. Based on the results of Experiment 1, we also created a set of intermediate stimuli to match specific properties of the grating and noise so that we could examine

what factors contributed to the differences between these stimuli observed in Experiment 1.

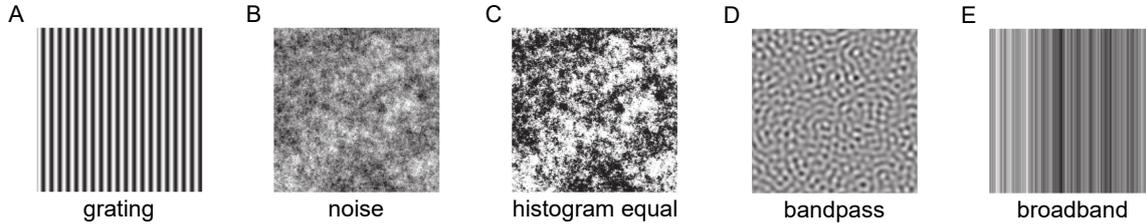


Figure 1.5: Example images of the stimuli used in Experiment 2, with center and surround both having contrast of 1. Stimulus types included: A) the vertical 5cpd grating from Experiment 1, B) the 1/f noise pattern from Experiment 1, C) the noise pattern with histogram adjusted to match the grating, D) the noise pattern with bandpass filtering centered at 5cpd, and E) a broadband grating.

Specifically, we adjusted the noise stimulus in three ways to make it more similar to the grating stimulus (Figure 1.5C-E). First, we created a noise pattern with its pixel histogram matched to the grating (Figure 1.5C). Second, we bandpass-filtered the noise image around 5cpd to create a narrowband noise image (Figure 1.5D). Third, we took the first row of pixels in the noise image (1D noise) and repeated it across all rows to create a broadband grating that contained a single orientation as with the grating stimulus (Figure 1.5E). As a control, we also included a horizontal 5cpd sine wave grating in addition to the vertical 5cpd for comparison (not shown).

Lastly, to evaluate the role of edge contrast, we also included an edge-blurred version of the vertical grating and noise stimuli (Figure 1.6). The blurring was done by applying two different masks to the images of center and surround separately. A Gaussian filter was applied to the masks such that the mask for the center peaks at the middle of image and has a decreasing ramp towards the surround. The mask for the surround peaks at the surround and ramps in the opposite direction. Effectively, the masks crop the center and surround with blurred border. The masked center and surround images were then added together. The standard deviation ( $\sigma$ ) of the Gaussian filter was  $0.5^\circ$ . By eye, it is notable that visibility of the edge is more pronounced for the grating than the noise without edge blur (Figure 1.6A, C). The edge blur seems to reduce the visibility of the edge for both stimuli (Figure 1.6B, D).

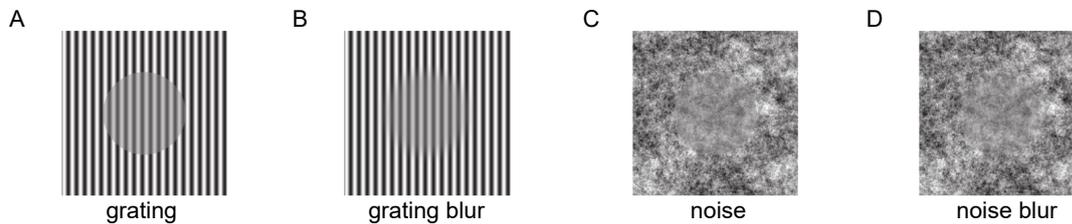


Figure 1.6: Example images of high surround stimuli (surround contrast = 1) for the low contrast center eye (center contrast = 0.5) used in Experiment 2. A) and B) show the original 5cpd grating and edge-blurred 5cpd grating, C) and D) show the original noise and edge-blurred noise.

In summary, for Experiment 2 we tested the mean surround and high surround conditions, with eight different stimulus patterns (2 baselines, 3 adjustments, 2 edge-blurred, and 1 horizontal grating) and four contrast levels for each eye’s reference stimulus: 0, 0.25, 0.5, and 1, resulting in 15 combinations. We did not include the 0 and 0 combination

since this combination is simply uniform gray for both eyes’ reference images, and results from Experiment 1 showed that participants matched this condition close to 0 contrast as expected. In total, there were 240 trials in Experiment 2.

## 1.3 Data Analysis

### 1.3.1 Outlier Criteria

Outliers were determined by assessing the performance on catch trials in which both the reference and the adjusted stimuli were non-dichoptic. On these trials, we expect the adjustment to be a very close match to the reference if the participants are motivated to perform the task correctly. We calculated the square root of the mean squared error (RMSE) for each participant on the catch trials to capture the average error between an individual’s match and the theoretical perfect match. If a participant’s RMSE exceeded 1.5 times the interquartile range (IQR) above and below the 75th and 25th percentiles across all participants, they were considered to be an outlier. Based on this criterion, none of the participants were excluded from Experiment 1. For Experiment 2, based on the outlier analysis, two of the 34 participants were excluded from further analysis.

### 1.3.2 Data Transformation

There are three naive predictions we can make about the appearance of the dichoptic reference stimulus. If the perceived contrast of the reference matches the eye seeing the higher contrast, then the results would follow a winner-take-all pattern illustrated in Figure 1.7A. In this prediction plot, the rows and columns of the heatmap correspond to the contrast levels shown to the two eyes for the reference stimulus and the matches for the adjustable stimuli are indicated by the gray levels. The predicted contrast match in each row/column pair is always equal to the higher contrast between the left and right eye. On the other hand, if the dichoptic reference percept is matched to the average contrast of the two eyes’ images, the resulting data would be similar to that shown in Figure 1.7B. Finally, in some conditions the perceived contrast of the reference stimulus could be matched to the lower contrast reference image (Figure 1.7C). Note that these are illustrative predictions and there are known cases, such as the Fechner’s paradox [56, 104], that do not fall into these three simple categories across all contrast combinations.

To create a summary measure of binocular combination and facilitate comparisons across different surround conditions and stimulus types, we wanted a method to fit the data represented by each heatmap with a single summary parameter. In a similar experiment to ours, Legge and Rubin used a nonlinear equation to fit their data, where an exponent parameter was varied to capture the three types of predictions [103]. However, one disadvantage of this method is that the possible parameter space is not constrained. The exponent is equal to 1 for averaging, but needs to be at infinity to fit winner-take-all and negative infinity to achieve loser-take-all.

We decided instead to express our data as a multiplicative weight ( $w$ ) on the higher contrast stimulus, which has a more restricted parameter space, between 0 and 1. We model the matched non-dichoptic contrast ( $b$ ) as a weighted combination of the two eyes’ contrasts ( $h$  and  $l$ ). For example, a weight ( $w$ ) of 1 for the higher contrast stimulus ( $h$ )

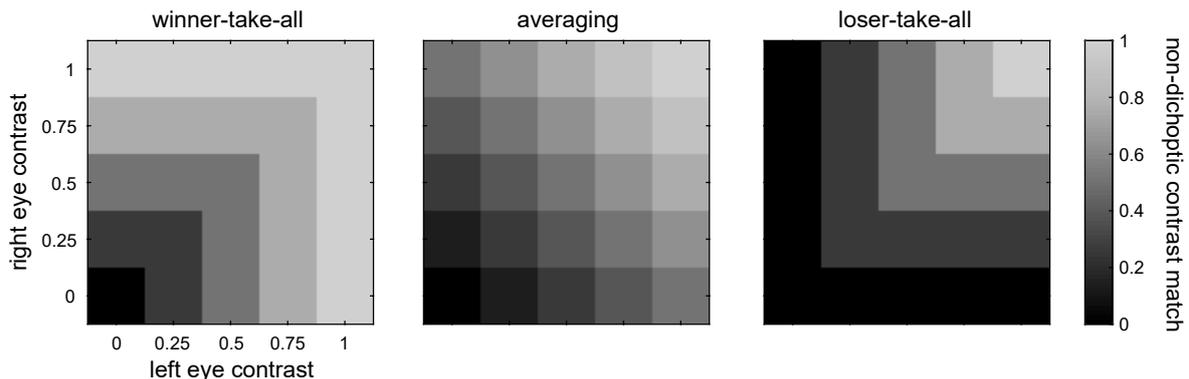


Figure 1.7: Hypothetical data showing three different naive predictions about the perceived contrast of the reference stimulus. The grey color of each square corresponds to the matched adjustable stimulus’ contrast for a given left ( $x$  axis) and right eye contrast ( $y$  axis) of the reference stimulus. A) adjustable stimulus matches the higher contrast reference image, B) adjustable stimulus matches the average contrast of the two reference images, C) adjustable stimulus matches the lower contrast reference image.

would mean that the perceived contrast is solely determined by the high contrast image (winner-take-all), a weight of 0.5 means averaging, and a weight of 0 means loser-take-all:

$$b = wh + (1 - w)l$$

To determine the best fitting weight, we performed a grid search with weights ranging from 0 to 1 in steps of 0.01, and minimized the RMSE for each participant and each unique stimulus combination. This model is meant to capture the data rather than suggest the underlying neural computation involved. Since the weights were very similar across different exemplars of noise and natural images, we averaged the fitted weights across these exemplars for each participant. We kept the two grating stimuli separate to examine potential differences between higher and lower spatial frequencies. We compared our linear combination method with the non-linear method used by Legge and Rubin and confirmed that our simple weighted combination method achieves similar RMSE with the best fit parameters. A comparison of the fitting performance between the two methods are shown in Table 1. In general, the fits of both models to the data from Experiment 2 were worse. In the Results section, we describe an exploratory analysis of eye dominance effects in Experiment 2 which may explain the larger errors associated with simple model.

Table 1.1: Fitting performance of our method compared to the Legge and Rubin method. The mean RMSE for each method across all stimulus types and surround conditions for all participants is indicated, and the standard deviation of the RMSEs for the different best fit parameters is shown in parentheses.

	Legge and Rubin model	ours
Experiment 1	0.08 (0.02)	0.09 (0.02)
Experiment 2	0.24 (0.05)	0.13 (0.04)

### 1.3.3 Statistical Tests

Planned analyses included two-way ANOVAs (surround condition by stimulus type). A preliminary analysis of the results suggested that the responses did not satisfy the assumptions of a conventional ANOVA, so two-way permutation ANOVAs were used to determine statistical significance (aovp function from lmPerm R package). We denote the

p values from the permutation ANOVA as  $pp$ . We also run the regular ANOVA and report the F statistics and effect size (generalized eta squared). Follow up pairwise comparisons were conducted using Wilcoxon sign rank tests (`wilcox.exact` from `exactRankTests` R package) and p values were corrected for multiple comparisons using Bonferroni correction. A threshold of  $p \leq 0.05$  is used to determine statistical significance throughout.

## 1.4 Results

### 1.4.1 Experiment 1

To examine whether different stimulus types produced different binocular combination rules, we first compare all stimulus types in the mean surround condition, which has similar contour information in the two eyes. In Figure 1.8, the average response across subjects is shown in the same format as the predictions in Figure 1.7. The fitted weights that best describe each subject’s data are shown in Figure 1.9A. Qualitatively, all stimulus types produced percepts that closely followed the winner-take-all pattern with a high weight given to the eye viewing the higher contrast image. The 5cpd grating had a slight reduction in weight compared to the other stimuli (but this reduction was not statistically significant, see below).

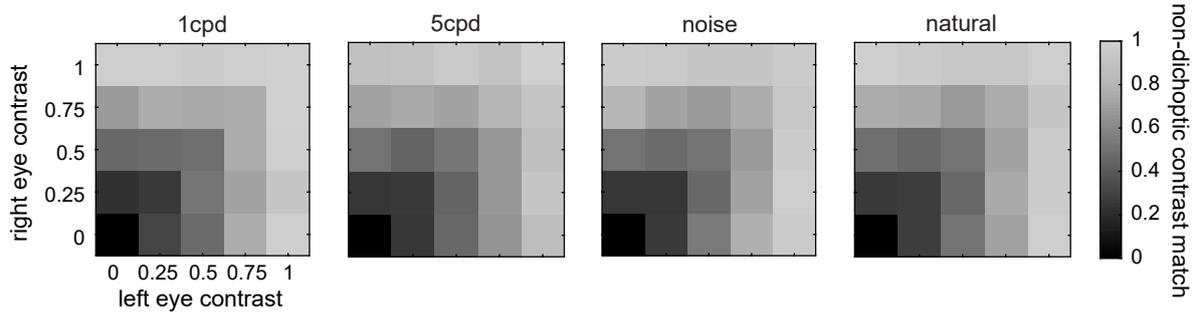


Figure 1.8: Experiment 1 ( $N = 10$ ), mean surround condition data averaged across all participants for the different stimulus types: A) and B) gratings, C) noise stimuli, D) natural texture

How do manipulations of the surrounding context influence this pattern of results? In the low surround condition (Figure 1.9B), the eye seeing a higher contrast center contains an edge contour and the other eye does not. We expected that the contour would bias the binocular percept even more towards the high contrast eye, giving it more weight. Indeed, qualitatively we see a weight close to 1 for all stimuli with no clear difference among the stimulus types. Finally, we expected the high surround condition (Figure 1.9C) to be associated with a lower weight on the higher contrast eye (and more weight on the lower contrast eye) compared to the previous two surround conditions, due to the monocular contour in the eye seeing the lower contrast center. Consistent with this prediction, all stimulus types were associated with a reduced weight on the higher contrast eye in this condition. Interestingly, however, the noise and natural stimuli appeared less affected by this manipulation.

The main effects of surround condition and stimulus type were both statistically significant (Surround:  $F(2, 18) = 23.28$ ,  $pp < 0.001$ ,  $\eta^2 = 0.54$ ; Stimulus:  $F(3, 27) = 23.40$ ,  $pp < 0.001$ ,  $\eta^2 = 0.24$ ) and the interaction between surround condition and stimulus type was also statistically significant ( $F(6, 54) = 27.77$ ,  $pp < 0.001$ ,  $\eta^2 = 0.31$ ). We asked

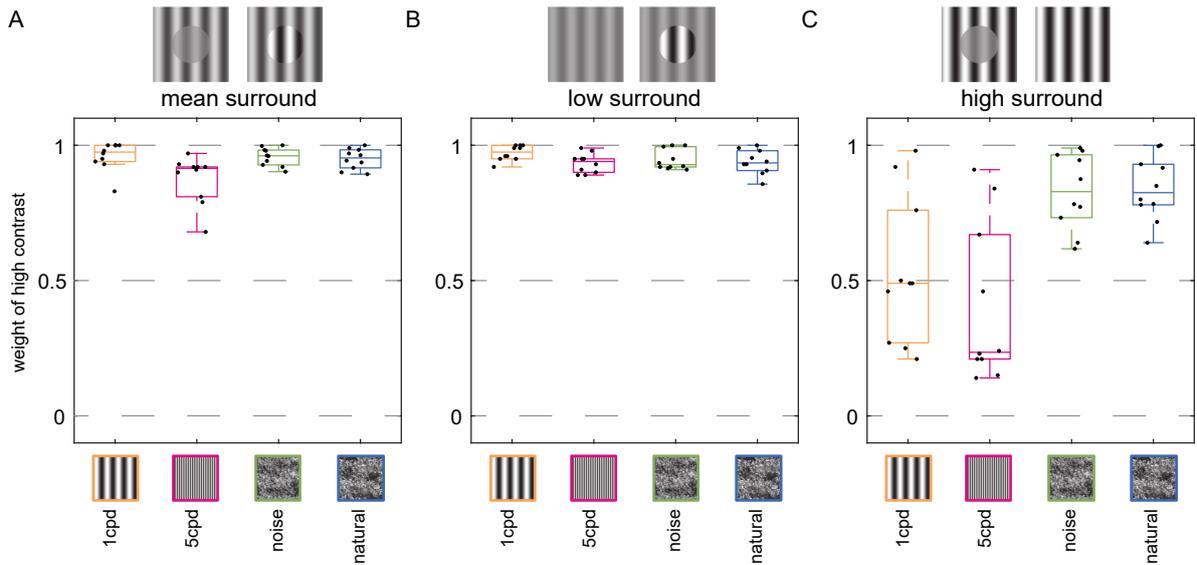


Figure 1.9: Experiment 1 results ( $N = 10$ ) for the four stimulus types in the A) mean surround, B) low surround, and C) high surround condition. The box-and-whisker plots show the median weight of the higher contrast image across individuals, the 25th and 75th percentiles, and the non-outlier range. The black dots indicate each participant's weight. The gray dash lines represent the weights for the three types of combination rules (winner-take-all, averaging, loser-take-all).

two questions in the follow up tests. First, we asked how the weights associated with the different stimuli compared to each other within each surround condition (Bonferroni correction for 18 tests). In the mean surround condition, the weight associated with the 5cpd grating was statistically lower than the weight associated with the 1cpd grating ( $p = 0.04$ ), but not the other stimuli. There were no significant differences amongst the stimulus types in the low surround condition. In the high surround condition, the 5cpd grating was associated with a significantly lower weight than all other stimuli ( $ps = 0.04$ ). Second, we asked how the surrounding context affected each different stimulus type (Bonferroni correction for 12 tests). For both grating stimuli, we found that the high surround condition was associated with significantly lower weight as compared to both the mean surround and low surround conditions ( $ps = 0.02$ ). No other stimuli were associated with statistically significant differences across the different surround conditions. These results suggest that the surround significantly affected perceived binocular contrast for the grating stimuli, but not for the more complex stimuli.

## 1.4.2 Experiment 2

In Experiment 2, we aimed to replicate the results of Experiment 1 and to understand why the effects of the surround differed between stimulus types. As such, we tested several stimuli with visual properties that were intermediate between the gratings and naturalistic stimuli, focusing on the mean surround and high surround comparison. There are many differences between these stimuli, so we considered several hypotheses about what might contribute to the different results: 1) the gratings are narrowband in spatial frequency, whereas the noise/textures are broadband, 2) the gratings have a different pixel value histogram than noise/textures (more values close to blacks and whites), and 3) the gratings have only one orientation, whereas the noise/textures are broadband in orientation.

Qualitatively, the results for the grating and noise stimuli in the mean surround condition were consistent with Experiment 1, with a higher weight given to the eye viewing the high contrast stimulus (Figure 1.10A, magenta and green boxes). The intermediate stimuli were also all associated with a high weight on the high contrast eye in this condition (Figure 1.10A, blue boxes). However, in this experiment the 5cpd grating stimulus was associated with even more individual variation and a more neutral weighting compared to the other stimulus types in the mean surround condition. In the high surround condition, the results were again qualitatively similar to Experiment 1 for the grating and noise: we observed a shift in weight towards the eye viewing the lower contrast stimulus, but the shift was stronger for the grating stimulus (Figure 1.10B, magenta and green boxes). Matches for all the intermediate stimuli were also shifted towards the lower contrast eye, but only the broadband grating was associated with a substantial shift similar to the grating (Figure 10B, blue boxes).

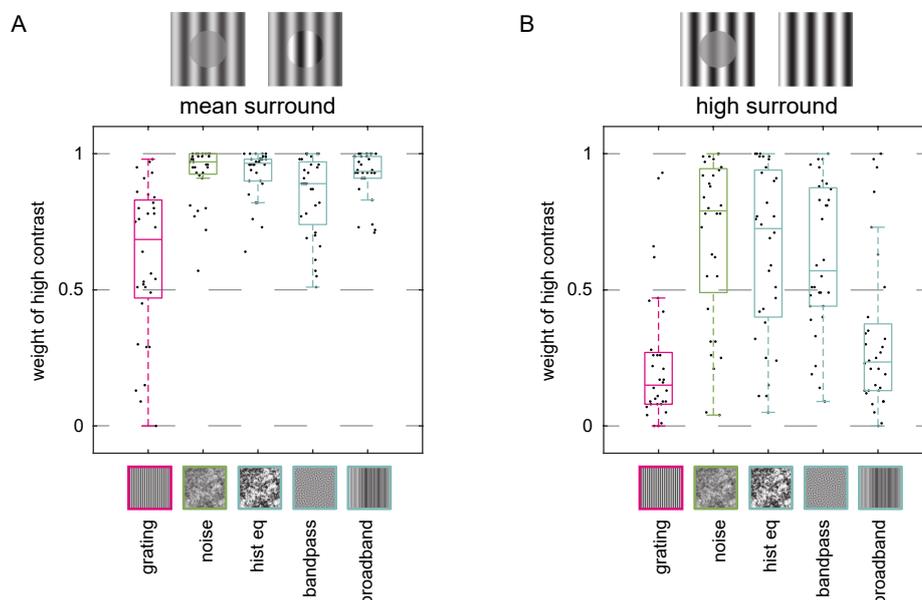


Figure 1.10: Experiment 2 results ( $N = 32$ ) across the two surround conditions for the five stimulus types: the narrowband 5cpd grating and noise baseline from Experiment 1 (grating and noise), histogram matched noise (hist eq), bandpass noise (bandpass), and broadband grating (broadband).

An ANOVA conducted on the results of Experiment 2 indicated a significant main effect of stimulus type ( $F(4, 124) = 64.15$ ,  $pp < 0.001$ ,  $\eta^2 = 0.28$ ) and surround ( $F(1, 31) = 115.06$ ,  $pp < 0.001$ ,  $\eta^2 = 0.37$ ). The interaction between stimulus type and surround was also significant ( $F(4, 124) = 17.29$ ,  $pp < 0.001$ ,  $\eta^2 = 0.08$ ). Follow up tests were conducted again to examine how the results for different stimuli differed from each other in each surround condition (Bonferroni correction for 20 comparisons), and how the results for each stimulus type varied across different surround conditions (Bonferroni correction for 5 comparisons). When comparing the mean and high surround conditions, there was a significant reduction in the weights associated with the 5cpd grating from the mean surround to the high surround condition ( $p < 0.001$ ) as in Experiment 1. However, there was also a significant weight reduction between the surround conditions for the noise stimulus and the three intermediate stimuli ( $ps < 0.001$ ). That is, the weights for all five stimulus types were reduced in the high surround condition.

Unlike in Experiment 1, the weight associated with the 5cpd stimulus was different from

the weight associated with the noise stimulus in the mean surround condition ( $p < 0.001$ ; in Experiment 1, this difference was visible but not statistically significant,  $p = 0.18$ ). Indeed, the weight associated with the 5cpd grating was significantly lower than all other stimuli in this condition ( $ps < 0.001$ ). The weight associated with the bandpass noise also differed from noise ( $p = 0.004$ ), histogram equalized noise ( $p < 0.001$ ), and broadband grating ( $p < 0.001$ ). Turning to the high surround condition, we found that both the 5cpd grating and broadband grating stimuli were associated with statistically lower weights than the other stimulus types ( $ps < 0.001$ ), but were not different from each other ( $p = 0.16$ ).

Taken together, these results suggest that the broadband grating was the most similar intermediate stimulus to the 5cpd grating in the high surround condition. These results suggest that the luminance histogram and spatial frequency of the grating are not key factors for generating strong surround modulation, but the orientation structure may be. Specifically, the broadband grating appeared more like the noise and other intermediate stimuli in the mean surround condition but appeared more like the 5cpd grating stimulus in the high surround condition. We conclude that the singular orientation structure of the narrow and broadband gratings likely interacts with the edge contour to modulate contextual effects on binocular contrast perception.

### 1.4.3 Stimulus Orientation and Edge Blur

In Experiment 2, we also tested a horizontal grating as a control, and edge blurred versions of the grating and noise stimuli to examine whether reducing the edge contrast between the center and the surround would weaken the bias towards the lower contrast eye in the high surround condition. Starting with the orientation manipulation (Figure 1.11A), the main effects of orientation ( $F(1, 31) = 7.08$ ,  $pp = 0.01$ ,  $\eta^2 = 0.01$ ) and surround ( $F(1, 31) = 75.32$ ,  $pp < 0.001$ ,  $\eta^2 = 0.32$ ) were significant, but there was no significant interaction between these two factors. Thus, we conclude that the surround modulation effect was not substantially disrupted by making the grating horizontal.

For the edge blurred stimuli (Figure 1.11B), the main effects of blur ( $F(1, 31) = 14.57$ ,  $pp = 0.001$ ,  $\eta^2 = 0.01$ ), stimulus type ( $F(1, 31) = 116.60$ ,  $pp < 0.001$ ,  $\eta^2 = 0.37$ ), and surround type ( $F(1, 31) = 119.77$ ,  $pp < 0.001$ ,  $\eta^2 = 0.30$ ) were each significant. However, the effect size for blur was quite small. There were small but significant interactions between surround and stimulus type ( $F(1, 31) = 7.26$ ,  $pp = 0.01$ ,  $\eta^2 = 0.03$ ), and between stimulus types and blur ( $F(1, 31) = 5.40$ ,  $pp = 0.04$ ,  $\eta^2 < 0.01$ ). In follow up tests, we looked at the effects of blur and surround for gratings and noise separately. For the grating stimuli, the main effect of edge blur ( $F(1, 31) = 16.01$ ,  $pp < 0.001$ ,  $\eta^2 = 0.02$ ) and surround ( $F(1, 31) = 74.05$ ,  $pp < 0.001$ ,  $\eta^2 = 0.38$ ) were significant, and there was no significant interaction between surround and edge blur. For the noise, the effect of blur was not significant ( $F(1, 31) = 1.47$ ,  $pp = 0.28$ ,  $\eta^2 < 0.01$ ), while the surround effect was significant ( $F(1, 31) = 30.33$ ,  $pp < 0.001$ ,  $\eta^2 = 0.21$ ). In summary, the effect of blur was significant for the 5cpd grating, but minimal compared to the effect of surround modulation.

### 1.4.4 Eye Dominance

In Experiment 2, we sought to replicate and extend the findings from the grating and noise stimuli in Experiment 1. While different sample size and the different number of

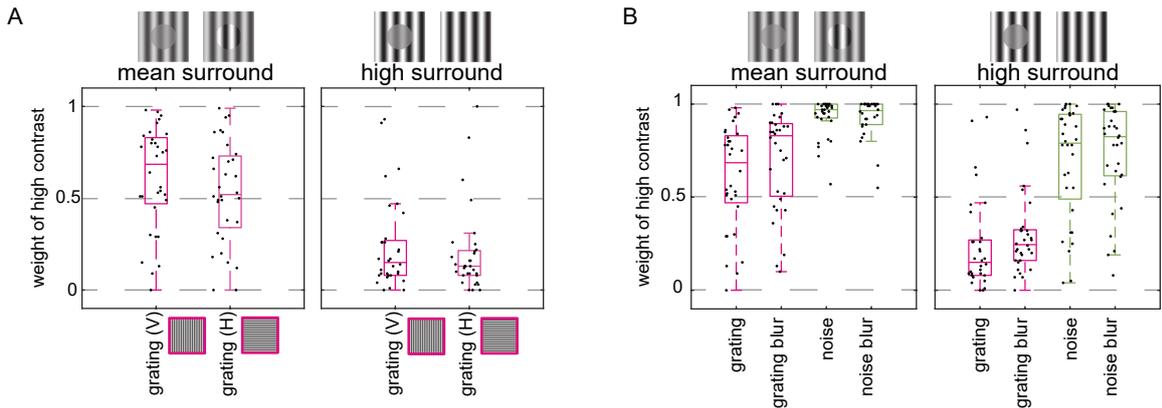


Figure 1.11: Additional stimuli tested in Experiment 2. A) Comparison between vertical (V) and horizontal (H) 5cpd gratings. B) The effect of edge blur on the baseline vertical 5cpd grating and noise.

multiple comparison correction applied could lead to Experiment 2 having more statistical power, and thus detecting smaller differences that we did not detect in Experiment 1, the effects of interest were generally replicated. However, we also found an unexpected variation in the best-fit weights for the 5cpd grating in the two experiments. It was not our original intent to examine eye balance, so we used a post hoc analysis to explore the possibility that the large individual variations observed in our data could be a result of interocular imbalance (i.e., different binocular percepts are observed depending on whether the dominant eye sees the higher contrast image or the lower contrast image). We can visualize eye dominance effects by looking at individual participant’s response heatmap. In Figure 1.12, the responses from two different participants in Experiment 2 are shown. Subject A shows strong asymmetry depending on which eye was presented with the higher contrast stimulus in many conditions (in this case, the participant appears to be left eye dominant), whereas Subject B does not show such pattern. Thus, in this post hoc analysis, we separately examined trials in which the high contrast reference images were presented to the dominant eye versus the non-dominant eye.

To determine whether the left or right eye was the dominant eye for each participant, we used a criterion defined by Legge and Rubin [103]. We separated out the trials in the mean surround condition based on whether the left eye saw the higher contrast reference target or the right eye saw the higher contrast reference target. After separating the two sets of data, we assessed which eye’s data was closer to a winner-take-all model by comparing the RMSE of both eyes’ data against a perfect winner-take-all weighting scheme. The eye that resulted in a lower RMSE was labeled as the dominant eye. We tried three ways to separate the data. Since the 5cpd stimulus is the one that we are interested in understanding, we first considered using only trials with the 5cpd grating stimulus. We also tried using the noise stimulus and all stimuli. The results were similar for all three approaches, so here we report the results when eye dominance was determined from the 5cpd stimulus (Figure 1.13). By this measure, 60% of participants in Experiment 1 and 63% of participants in Experiment 2 are left eye dominant.

Quantitatively, the trend for different surround and stimulus patterns that we described previously in the main analysis holds both when the dominant eye and non-dominant eye were presented with the higher contrast stimulus. However, there are some differences between the dominant eye data and the non-dominant eye data. The dominant eye data has much less variation and higher weight for the 5cpd grating blur compared to the non-

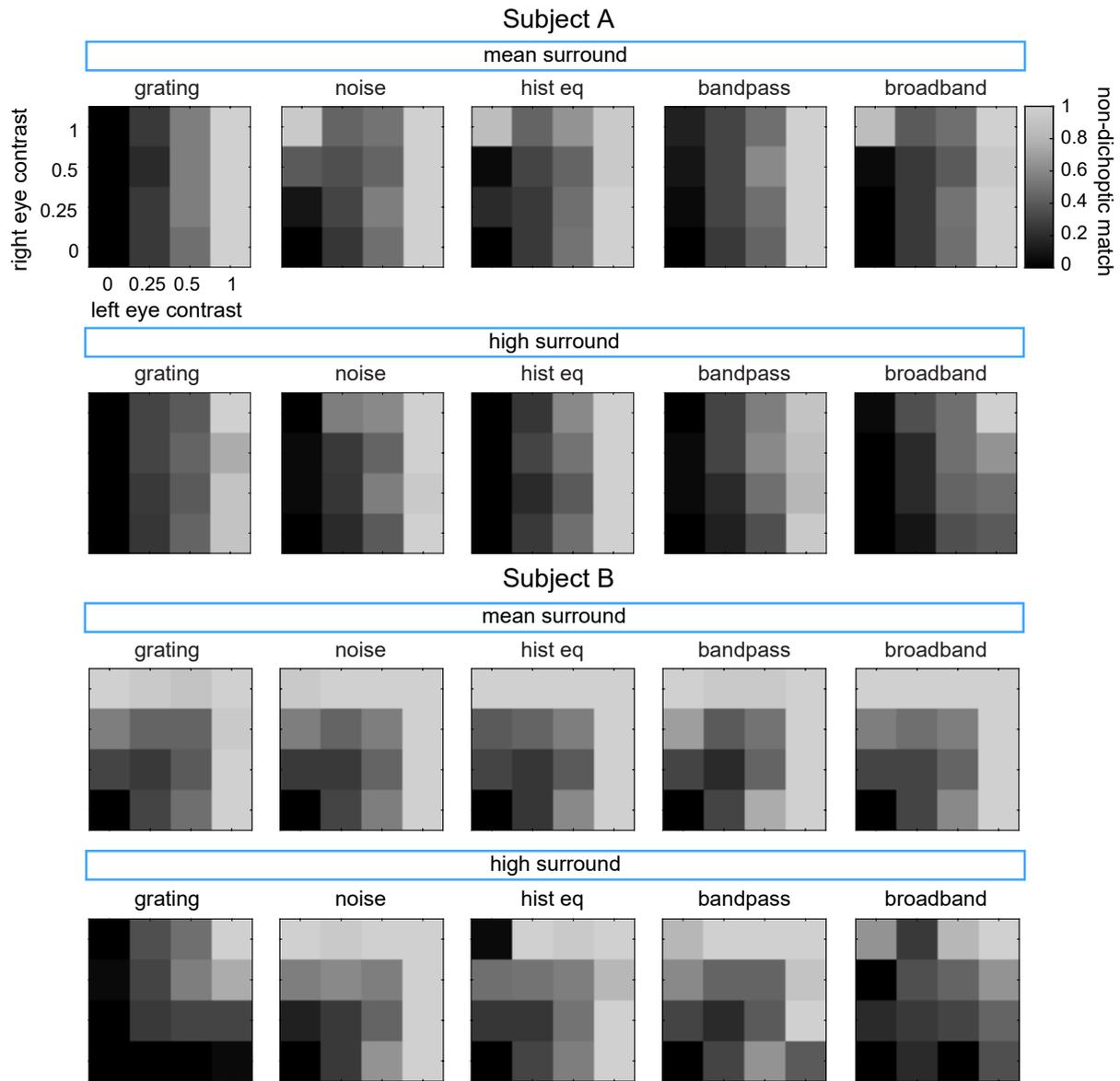


Figure 1.12: Heatmaps showing the non-dichoptic contrast match for all stimuli in the mean surround and high surround conditions for two observers in Experiment 2. Subject A shows winner-take-all when the left eye sees the higher contrast, but not when the right eye sees the higher contrast in most situations. Subject B generally has the same response regardless of which eye saw higher contrast.

dominant eye, and when both eye's data are combined in the main analysis (Figure 1.10). This suggests that the differences in the two experiments could be related to a difference in the two groups' eye dominance. Interestingly, for Experiment 2, eye dominance changes the responses for the bandpass noise centered at 5cpd in the mean surround condition as well, suggesting a spatial frequency dependence.

## 1.5 Discussion

In the current study, we examined how the spatial complexity of binocular stimuli affects binocular contrast perception. We found that dichoptic contrast percepts for both grating and naturalistic stimuli follow a roughly winner-take-all model when the surrounding

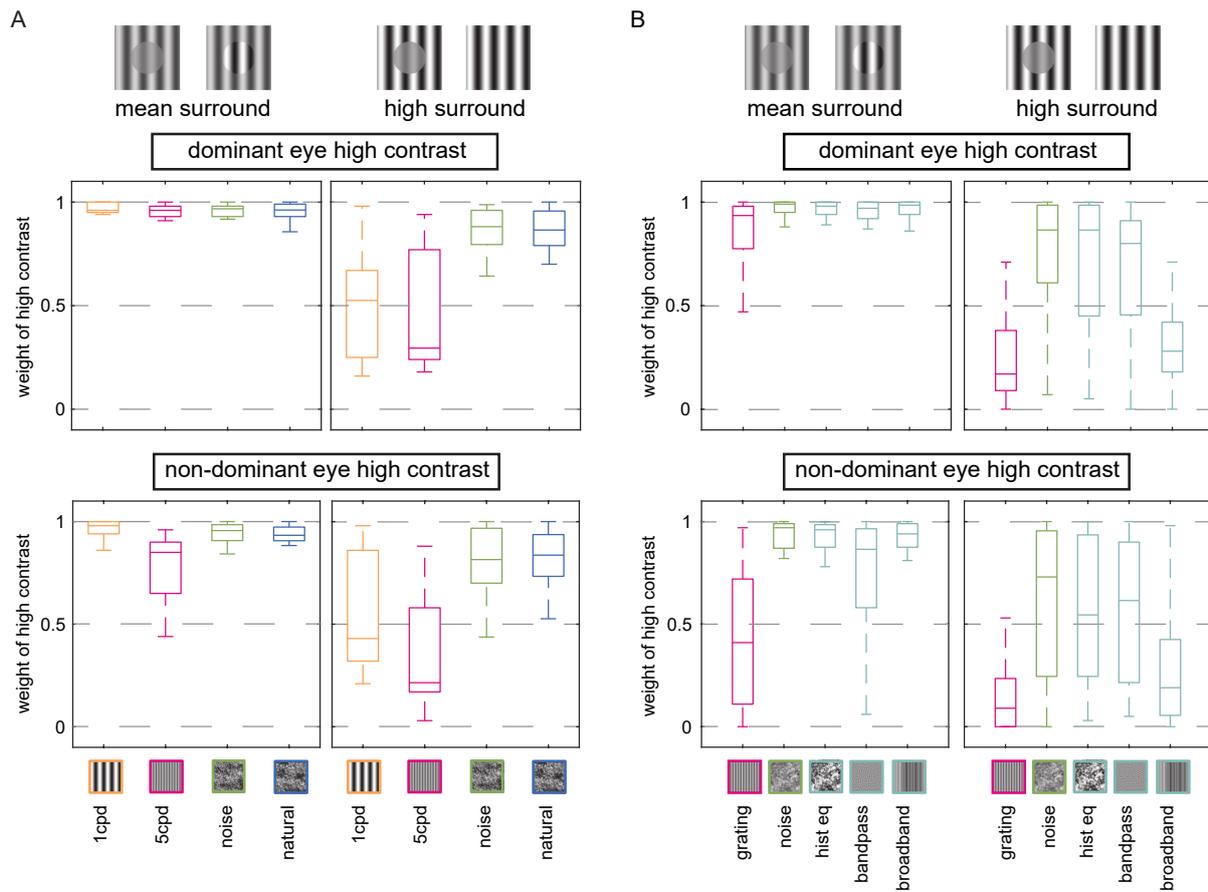


Figure 1.13: Comparison between trials where the dominant eye (top panels) or non-dominant eye (bottom panels) was presented with the higher contrast image. Eye dominance is determined by 5cpd mean surround responses. A) Experiment 1 B) Experiment 2.

contrast matches in the two eyes (but the results from Experiment 2 suggest that the interocular weighting is more variable for gratings). Across a range of stimulus types, we also found evidence that monocular edge information due to a contrast difference with the surround biases dichoptic contrast perception toward the eye seeing the edge.

Our results with respect to the effects of a monocular edge are directly in line with previous work that assessed the role of a monocular contour in dichoptic luminance perception [42, 104] and dichoptic contrast perception [62]. This set of results can all be well-described in terms of interocular suppression, in which the strength of the suppression depends on the “strength” of the winning stimulus. We can think of the winner-take-all behavior as a complete suppression of the weaker stimulus by the stronger stimulus. Introducing a monocular feature, such as contour, to the eye with the weaker stimulus could lead to averaging between the two eyes if the added feature causes the weaker stimulus to be less suppressed by the stronger stimulus (i.e., the contour feature “strengthens” the stimulus).

Evidence of similar reductions in interocular suppression has also been found in studies of dichoptic masking. Dichoptic masking refers to the observation that the detection threshold for a stimulus presented in one eye can be increased due to the presence of a masking stimulus presented in the other eye [118]. We can reason that “winner-take-all” dichoptic contrast perception is congruent with a complete masking of the weaker stimulus

by the stronger stimulus. In dichoptic masking, it has been shown that the suppression of a masked target can be reduced to the level of no masking when a monocular ring surrounds the target [76], similar to the monocular contour effects observed in dichoptic luminance/contrast perception.

Together, these results support the theory that monocularly-visible contours, however induced, contribute to the perception of dichoptic stimuli. It is worth pointing out that previous studies have also found that adding binocularly matched features to both eyes can modulate suppression effects in dichoptic color masking [79], dichoptic color perception [81], binocular luster [168], and binocular rivalry [20]. Our results thus contribute to converging lines of evidence that both monocularly-visible and binocularly-matched features modulate suppression, further motivating the notion that contextual relationships must be considered as a whole for determining interocular suppression.

We also found a surprising result that gratings showed a larger contextual change than more complex noise and natural stimuli. In Experiment 2, we tested several properties that could contribute to the difference and found evidence that the spatial orientation of the grating pattern (single orientation) contributes to the difference. How and why does the spatial organization of the stimulus affect binocular contrast perception? Here, we discuss two related possibilities: low level suppression and higher-level segmentation.

Surround suppression is a well-documented characteristic of the visual system. It has been demonstrated that both grating and non-grating stimuli can have different apparent contrast depending on the surrounding contrast, and that the similarity between the center and surround plays a key role in the amount of modulation [32, 149, 172]. It is possible that the visual similarity between center and surround in our grating and noise stimuli differs, which contributes to a difference in the amount of surround modulation on perceived contrast. For example, the similar orientation between the center and surround regions of the grating stimuli could cause stronger surround suppression in monocular pathways, which would thus lead to a stronger monocular edge between the center and surround and cause the eye with the edge to dominate more. On the other hand, for the broad orientation stimuli, there may be less suppression because the center and surround differ more in their orientation properties. It is important to note that in our study, the noise and natural textures do not have a dominant orientation. However, in real-world scenarios, one can imagine viewing a dichoptic scene with trees or buildings that have a strong orientation bias, similar to the grating stimuli. Quantifying surround suppression for stimuli with a combination of spatial frequencies, phases, orientations, and contrast/luminance is difficult but may be important for ultimately explaining the full pattern of binocular contrast perception with natural stimuli. In addition, there are likely other types of suppression happening both within and between the monocular pathways. For example, previous work has modeled suppression effects within each eye (cross orientation suppression/masking) and between the eyes (dichoptic masking) at different stages of visual processing [11]. It would be interesting to see if this framework could be adopted to reflect surround suppression and dichoptic contrast perception as well. In general, monocular suppression mechanisms are likely key for determining how monocular channels interact upon binocular combination. Existing binocular combination models can generate good predictions for binocular perception of simple stimuli [10, 39, 42, 76]; however, a versatile model that can predict binocular perception for any dichoptic image pair would be very desirable and would likely need to consider all the aforementioned characteristics of each eye's input. A

precise computational instantiation of stimulus strength and its respective suppression remains a topic of ongoing empirical and modeling research.

Another potential explanation for the contextual effects observed in this study is a difference in higher order segmentation for the single orientation versus broad orientation stimuli. It may be that the single orientation and regularity of the narrow and broadband grating patterns made it possible for observers to visually segment the center and surround better. In other words, the local contrast variation in the center and surround regions of the orientation-broadband stimuli (e.g., noise) could make it harder to segment them, which in turn may make it more difficult for the visual system to assign contours. This hypothesis is similar to earlier work showing that orientation-induced boundary contours influence eye dominance dynamics during binocular rivalry [125] and that the subjective strength of a phase-shift induced contour determines which eye dominates the binocular percept [62]. There is some evidence that the visual system first determines the contour information then processes the interior of the enclosed contour via a filling-in mechanism [152]. In our experiments, perhaps a harder-to-define contour in the broad orientation stimuli leads the visual system to fill-in the center by evaluating the pattern/contrast in the surround. Whether this segmentation is performed in the brain by the same mechanism underlying surround suppression of simple stimuli is unclear [134].

From the perspective of real-world applications, these results provide a foundation for assessing how visual information beyond the target of interest can influence binocular contrast perception. In our study, the high surround condition approximates a partial monocular scotoma where one eye has a region of low contrast compared to the rest of the view while the other eye may have no vision defects. We found that this condition biases the binocular percept towards lower contrast stimulus in persons with normal vision. We hypothesized that blurring the edge would reduce this bias, but the blurring manipulation in our study ultimately had minimal effect. It is possible that larger amounts of edge blur may produce stronger results, though. However, it is unclear how closely this manipulation applies to real-life scotomas. Indeed, it is still an ongoing area of research to characterize the vision of people with scotomas and accurately simulate scotomas [46, 139, 156, 157, 159]. It is possible that the high surround stimulus does not correlate well with what people with scotomas experience, particularly if patients engage adaptation mechanisms to correct for local contrast sensitivity loss. Nonetheless, the importance of surrounding context suggests that the way the two eyes combine information relies on not just the corresponding areas in the two eyes, but also what is outside of those areas.

Finally, we explored the effect of eye dominance on dichoptic contrast perception and found evidence that eye dominance exerts a larger influence over the dichoptic percept for spatially narrowband stimuli than broadband stimuli (i.e., differences due to eye imbalance were more noticeable for the 5cpd gratings and bandpass noise). Assessing eye dominance can have important implications for training and rehabilitation of eye imbalance problems, in the extreme case of amblyopia. In the research literature, several methods have been proposed to measure sensory eye dominance [126]. One method involves assessing eye balance by adjusting the interocular contrast difference between the two eyes to null the effects of an interocular phase difference [43], orientation difference [176], or masking [14]. In amblyopic vision, the binocular imbalance depends on both spatial frequency and stimulus contrast, with more imbalance at higher frequencies and higher contrast [38, 40]. In addition, it has been shown that in adults with normal vision, there is a spatial

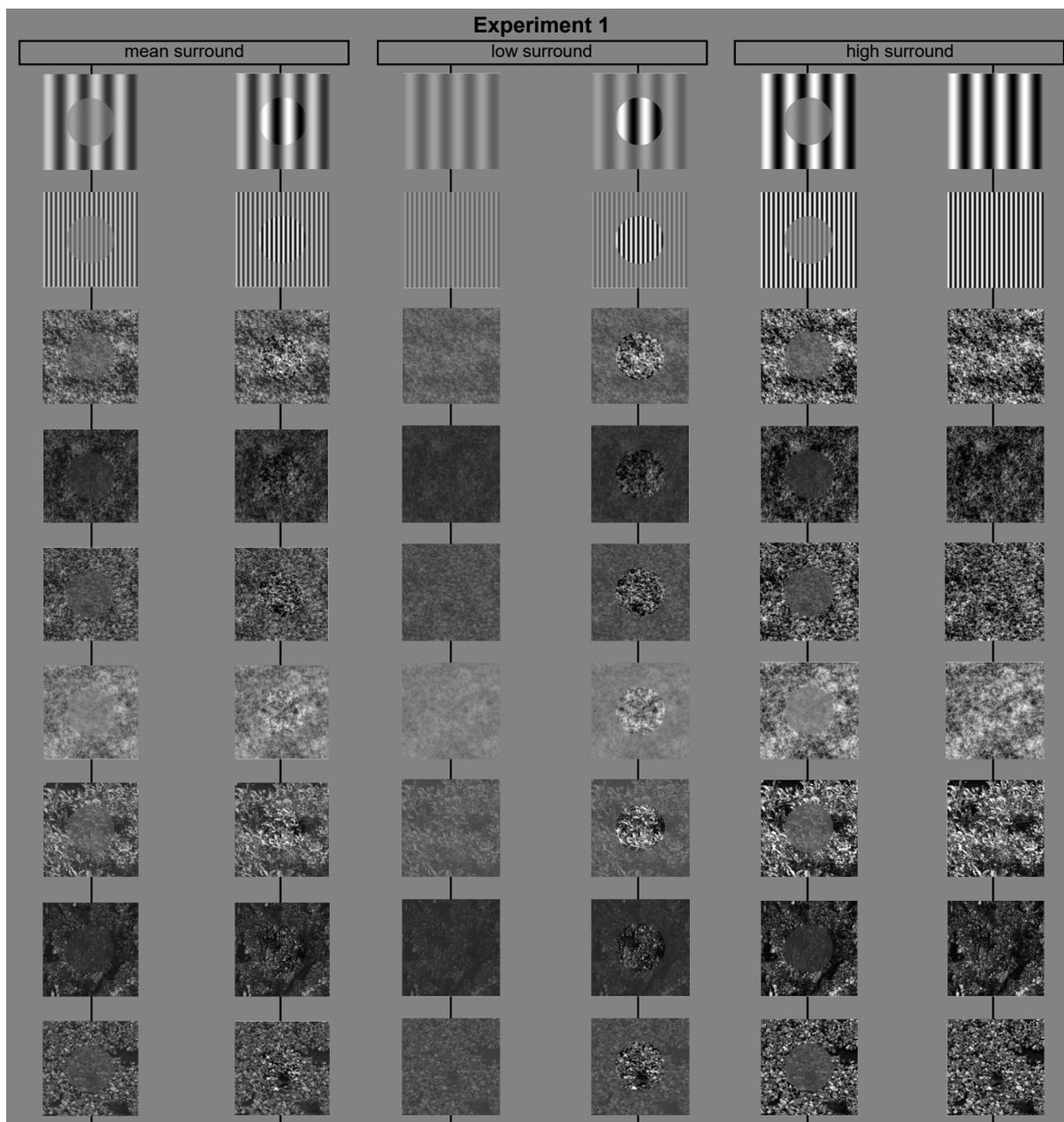
frequency dependency of eye dominance as well, such that there is more imbalance at higher spatial frequencies measured using phase and orientation nulling tasks [164]. Our study used the dichoptic contrast appearance as the measure instead, but found a similar result: eye imbalance was spatial frequency dependent, having a stronger effect on the 5cpd grating than the broadband or lower frequency stimuli. It remains to be explored why there is a spatial frequency difference in eye imbalance in the normal population.

## 1.6 Conclusion

Binocular contrast combination rules learned from simple stimuli such as sine wave gratings do not fully generalize to more naturalistic stimuli. Here, we uncovered one property that contributes to the difference between gratings and naturalistic stimuli: the restricted spatial orientation of gratings results in context modulations that differ from orientation-broadband stimuli. The underlying mechanism for why the orientation context matters for binocular contrast perception warrants future research. We hypothesize that surround suppression and/or image segmentation may be important factors. Going beyond simple stimuli to study binocular contrast combination of complex spatial patterns will inform the development of better models that can predict binocular perception during natural vision.

## 1.7 Supplemental Material

Cross fusible stimuli from Experiment 1 and 2 are shown in Figure 1.14 and 1.15. The black lines are added to aid cross fusion. Experiment 1 stimuli includes two grating patterns, four noise patterns, and three natural textures (Figure 14). Experiment 2 stimuli includes the baseline grating and noise pattern, along with noise histogram matched to grating, bandpass noise, and broadband grating (1D noise) (Figure 15). Here, we are also showing what Experiment 2 stimuli would look like for the low surround condition, which was not tested in Experiment 2. We encourage the reader to look at these stimuli and see how the contrast of the centers for the 1cpd, 5cpd, and broadband grating change between the high surround condition and the other two conditions.



*Figure 1.14: Example stimuli from Experiment 1.*

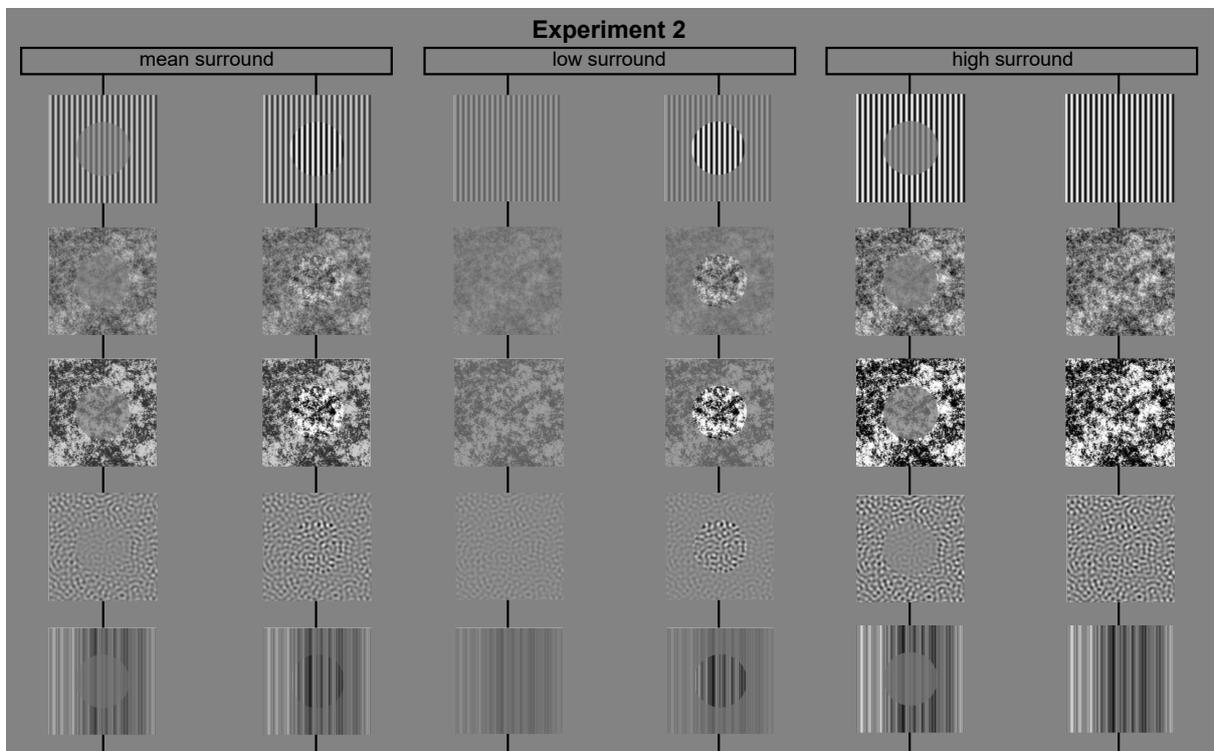
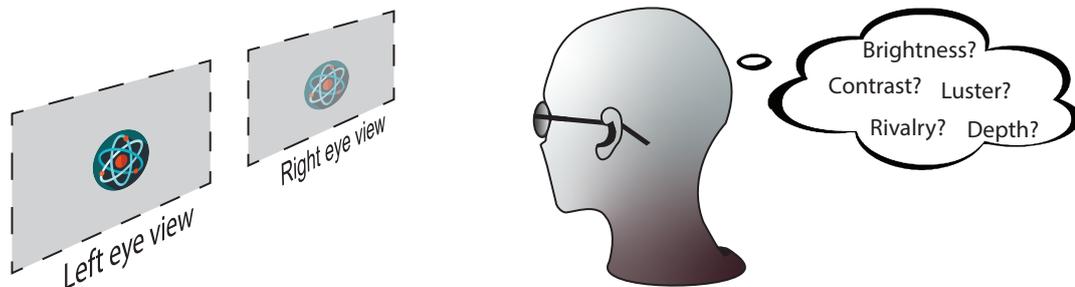


Figure 1.15: Example stimuli from Experiment 2.

# CHAPTER 2

## The effect of interocular contrast differences on the appearance of augmented reality imagery

Augmented reality (AR) devices seek to create compelling visual experiences that merge virtual imagery with the natural world. These devices often rely on wearable near-eye display systems that can optically overlay digital images to the left and right eyes of the user separately. Ideally, the two eyes should be shown images with minimal radiometric differences (e.g., the same overall luminance, contrast, and color in both eyes), but achieving this equality can be challenging in wearable systems with stringent demands on weight and size. Basic vision research has shown that a spectrum of potentially detrimental perceptual effects can be elicited by imagery with radiometric differences between the eyes, but it is not clear whether and how these findings apply to the experience of modern AR devices. In this work, we first develop a testing paradigm for assessing multiple aspects of visual appearance at once, and characterize five key perceptual factors when participants viewed stimuli with interocular contrast differences (brightness, contrast, luster, rivalry, and depth). In a second experiment, we simulate optical see-through AR imagery and use the same paradigm to evaluate the multifaceted perceptual implications when the display luminance differs between the two eyes. We also include a simulation of monocular AR systems. We find that interocular contrast difference is the main predictor for potentially detrimental perceptual effects such as binocular luster, rivalry,



*Figure 2.1: Binocular display systems present separate images to each eye. These systems are commonly used for augmented reality and are subject to unintended spatial, temporal, or radiometric differences between the two eyes due to hardware limitations or imperfections. Here, the user's view of an icon has higher contrast in the left eye than in the right eye.*

and spurious depth differences. In addition, monocular AR displays tend to have more artifacts than binocular displays that have a large contrast difference in the two eyes. A better understanding of the range and likelihood of these perceptual phenomena can help inform design choices that support a high-quality user experience.

## 2.1 Introduction

Designing new display systems often requires understanding whether and how the display’s visual limitations adversely affect the user experience. Display systems for augmented reality (AR) pose a unique set of challenges because they aim to merge virtual information into the user’s natural vision using a system with demanding design specifications (e.g., a wearable optical see-through near-eye display system) [94]. When these wearable systems are binocular, they employ independent displays and optics for the two eyes, introducing the potential for spatial, temporal, and radiometric differences in the virtual content that each eye sees (Figure 2.1). Here, we aim to explore the range of perceptual effects that can result when the user of an AR system receives a higher intensity image in one eye than the other.

From a display engineering perspective, differences between the left and right eye views can be desirable or detrimental. Importantly, binocular display systems enable the presentation of images with binocular disparities – the natural spatial offsets between the two eyes’ views that can elicit a compelling sense of depth via a perceptual process called stereopsis. However, patterns of imperfections in display panels (sometimes called mura) and spatial distortions introduced by optical architectures are not necessarily the same between the two eyes [92, 101, 122, 154]. These factors can introduce additional spatial and radiometric differences that are not intended by the designer, and understanding their potential perceptual consequences is key for optimizing the user experience.

Basic vision science studies, using simple shapes and gratings as stimuli, suggest that large interocular differences in brightness, contrast, and pattern between the two eyes are likely to elicit troublesome percepts in which the stimulus appears to shimmer or alternate in appearance over time (see [19, 169] for review articles). However, small differences can go unnoticed [53]. Recent applied research has begun exploring whether and how these phenomena might affect the appearance of AR content. For example, in AR systems with a small eyebox, the two eyes can be subject to different patterns of luminance vignetting (nonuniformity), which may result in degraded image quality [26]. However, a recent perceptual study suggests that a binocular AR display system with different vignetting patterns between the two eyes can result in reduced salience of vignetting artifacts as compared to a monocular system [31]. This prior work shows that certain types of radiometric differences may not be detrimental; in fact, it is possible to take advantage of binocular combination to achieve certain desired design goals (i.e. display uniformity). However, prior work has focused largely on assessing just one aspect of perceptual experience at a time [12, 39, 42, 69, 103, 167], while it is likely that binocular image differences cause multifaceted perceptual effects that are not well captured by a single perceptual measurement.

Here, we adopt the term *dichoptic* to refer to stimuli that differ radiometrically between the two eyes (e.g., differ in luminance, contrast, or color). A better understanding of the visual phenomena that occur when viewing dichoptic imagery in AR is needed to make

well-informed design decisions. We conducted two perceptual experiments to evaluate the implications of dichoptic imagery for user experience in AR display systems. Drawing on the basic vision science literature, we identified several perceptual factors pertinent to the appearance of dichoptic imagery in AR: perceived brightness, contrast, luster, rivalry, and depth (Figure 2.1). Using a battery of subjective response prompts, we studied all the aforementioned effects together across different levels of interocular difference and for different stimulus patterns. Studying these factors all together, rather than focusing on a single effect, enables us to characterize a broad gamut of potential perceptual consequences to AR display design.

Our primary findings are:

1) Across a broad range of visual stimuli, participants judged the appearance of dichoptic images to differ from non-dichoptic images with respect to all five perceptual factors tested. We found that luster was the perceptual effect reported most often with dichoptic stimuli.

2) As the contrast difference between the two eyes increased, the prevalence of dichoptic differences for all perceptual effects increased.

3) Monocular viewing (i.e., viewing display content in only one eye) resulted in a similar set of perceptual results, but with a higher prevalence of rivalry with AR-like images.

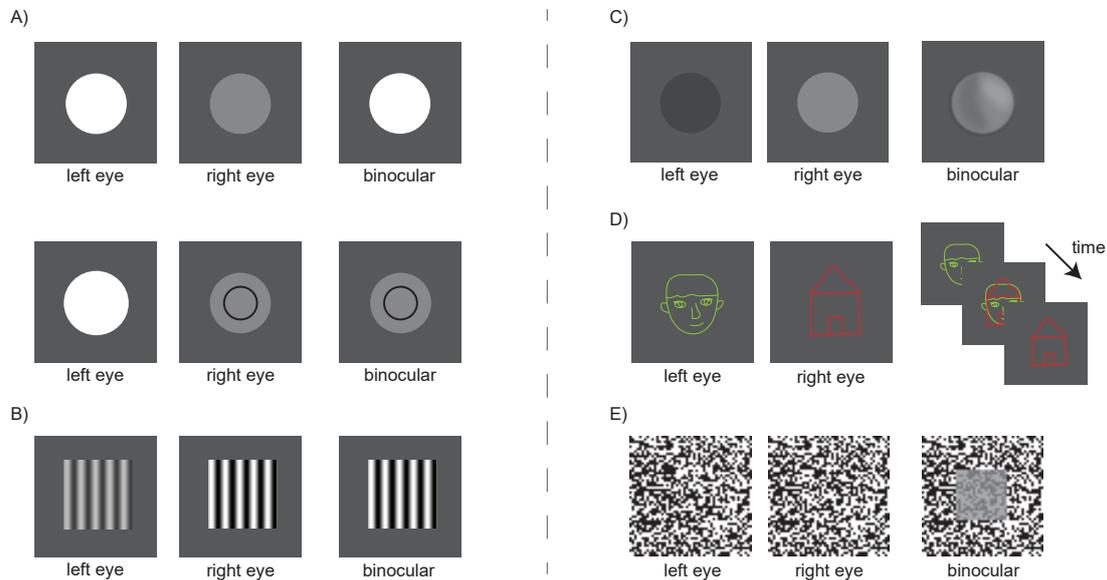
## 2.2 Related Work

In this section, we briefly summarize the range of perceptual phenomena associated with viewing dichoptic stimuli that differ in luminance or contrast between the two eyes. Here, we focus on achromatic effects, and will briefly discuss chromatic-related effects in the Discussion.

### 2.2.1 Brightness

For most people, closing one eye does not make the world appear any dimmer under normal viewing conditions. This observation suggests that perceived brightness is not a simple average of the luminance levels reaching the two eyes and has motivated a range of psychophysical research characterizing dichoptic brightness perception. Brightness perception is well-modelled as a weighted combination of the inputs to the two eyes, with the weights varying depending on the context. For example, when simple stimuli (e.g., uniform gray discs) with different luminance levels are shown to the two eyes, we can ask what the resulting perceived brightness is. Generally, the stimulus with a greater contrast is found to dominate the binocular brightness percept (sometimes termed 'winner-take-all') [12, 42, 104] (Figure 2.2A). That is, if both stimuli are bright compared to the background (increments), the binocularly perceived brightness tends to match the brighter stimulus and if both stimuli are dark compared to the background (decrements), the perceived brightness matches the darker one. However, percepts can shift towards binocular averaging under certain viewing situations [42]. For example, in *Fechner's paradox*, viewing a dichoptic image pair with different luminance levels in the two eyes results in a darker percept than if the observer closes one eye and just views the brighter of the pair monocularly [104]. Lastly, under certain viewing conditions, the brightness

percept can be more like 'loser-take-all' and biased towards the lower contrast image. In particular, if additional contours or edges are added to the stimulus with lower contrast, the perceptual biases can switch towards that stimulus [42, 104].



*Figure 2.2: Artistic illustrations of the different binocular perceptual phenomena that can result from interocular luminance and contrast differences. Readers are encouraged to cross-fuse the left and right eye images to observe the effects since the artistic depiction is not exact. A) Two pairs of stimuli with dichoptic luminance increments. The top row shows the winner-take-all brightness perception phenomenon. The bottom row, with a monocular contour in the eye seeing the lower luminance disc, illustrates the resulting bias towards that eye (in this case, loser-take-all). B) Dichoptic contrast perception of more complex patterns, for example if the contrast of a grating pattern differs between the two eyes, is often dominated by the eye seeing higher contrast. C) Binocular luster percepts can be elicited by dichoptic luminance stimuli. D) Binocular rivalry can be elicited by pairs of images with different visual patterns in the two eyes. E) Lastly, imagery that is anti-correlated and without binocular disparity between the two eyes can result in anomalous depth percepts. In this example, the binocular percept of the middle region (highlighted by the grey square) appears to be at a different depth from the surrounding.*

## 2.2.2 Contrast

A related line of research has asked how people perceive the contrast of more complex dichoptic stimuli when the average luminance is matched between the two eyes. Contrast refers to the range between the brightest and darkest regions of an image. Research participants can be asked to match or rate the perceived contrast of a binocularly-perceived sine wave grating when the two eyes view gratings with different contrast levels (Figure 2.2B). Research using this type of stimulus has shown that dichoptic contrast perception also tends to follow a 'winner-take-all' pattern similar to dichoptic luminance perception [39, 103]. This finding is not surprising given that a sine wave grating can be thought of a set of alternating luminance increments and decrements. But, like luminance perception, contextual effects can alter the balance between the two eyes for perceived contrast. For example, recent work showed that the dichoptic contrast percept can be strongly influenced by a lower contrast stimulus if it is embedded within a contour, similar to brightness percepts [163]. These modulations were also found to depend on the spatial properties of the stimulus: the influence of the contour was stronger for simple grating-like stimuli with a single orientation and weaker for other more complex stimuli.

### 2.2.3 Luster

If our perception of dichoptic stimuli could be completely modelled as a weighted mixture of the luminance and contrast of the two eyes inputs, then the challenge of predicting these percepts would just be a matter of determining the appropriate weights for a given stimulus. However, this is not the case. There are unique forms of binocular appearance that can emerge with dichoptic stimuli such as binocular luster. The *lustrous* appearance of dichoptic stimuli is often subjectively described as shimmery, shiny, or metallic (Figure 2.2C). A classic stimulus targeted to elicit binocular luster has opposite contrast polarity in the two eyes (i.e., a luminance increment in one eye and decrement in the other eye), but binocular luster can also be elicited when the two eyes have unequal increments or decrements [115, 167], or color [110]. For AR (and also for virtual reality, VR) applications, luster may be troublesome if it interferes with the perceived realism or material properties of the stimulus. However, it may also be a tool that designers desire to leverage, for example, to make a virtual object stand out visually or to break color metamerism [59](see [169] for review).

### 2.2.4 Binocular Rivalry

Another binocular phenomenon that occurs with dichoptic imagery is rivalry. During binocular rivalry, the appearance of a stimulus changes over time. When a stimulus elicits binocular rivalry, it may appear to match one or the other eye's input at any moment in time, or it may be perceived as a mixture (Figure 2.2D). For example, the binocular percept may be a patchy mix of the two eyes inputs, in which some parts of the percept look like one eye's input while other parts look like the other eye's input, or in which both eyes' inputs may appear to be superimposed [147]. To study binocular rivalry, a pair of highly dissimilar images (e.g., gratings with different orientations or two disparate images) are often used, but rivalry can also be elicited by more subtle interocular differences [136]. For many binocular AR devices, there is unlikely to be extremely dissimilar content seen by the two eyes, but for monocular devices that show virtual content to only one eye, rivalry may be more of a concern [130]. It is thought that the relative strength of each eye's input determines rivalry dynamics, and the eye with the stronger stimulus (e.g., brighter, higher contrast) is the predominant percept [105]. This observation holds true for simple stimuli, but not necessarily for more complex stimuli [148]. Compared to the other binocular effects covered in this section, salient rivalry is likely to be universally considered as an undesirable visual artifact that compromises the visibility of the displayed content.

### 2.2.5 Depth

It is well established that the visual system can use positional differences in the two eyes' images (binocular disparities) to infer depth information. It has been recently shown, however, that dichoptically tonemapped natural imagery with interocular contrast and luminance differences can generate a sense of depth as well [161, 183]. However, this depth effect has been elusive to vision science research, as it is harder to elicit consistently compared to binocular luster, rivalry, and stereoscopic depth. Pertinent psychophysical studies have demonstrated an anomalous depth effect (also referred to as the 'sieve effect' and 'rivaldepth') with anticorrelated images in which a white pixel in the left eye matches to a black pixel in the right eye, but there is no binocular disparity [73, 114, 127] (Figure

2.2E). There is individual variation for this anomalous depth effect, however, such that some participants can perceive a reversal in depth but not others [69, 138]. It also is highly dependent on the specific stimulus configuration used [68, 69]. These depth effects may also be associated with luster and rivalry. One small study found that these three effects could all be induced with the same amount of dichoptic luminance difference by simply changing the stimulus size [133]. Depending on the use case, this depth effect may be an additional tool for display designers to enhance depth impressions, since people often underestimate the distance of objects simulated via near-eye displays [47]. On the other hand, any anomalous depth effects may also be problematic for tasks that require fine depth accuracy.

## 2.2.6 Modelling Dichoptic Percepts

Considering the importance of the perceptual appearance of dichoptic imagery for display design, it would be useful to be able to predict binocular appearance given any pair of input images for the two eyes. Efforts have been made to develop models to predict various aspects of binocular percepts, but no model exists yet that has been shown to reliably predict a range of perceptual factors at once. For example, some models of binocular combination focus on implementing the mechanisms of early stages of interocular interaction (e.g, interocular suppression) based on basic stimulus properties (e.g., contrast) [43, 55, 80, 106], while other models, particularly those focused on rivalry, employ higher-level frameworks such as perceptual inference and decision making [28, 71]. However, oftentimes these perceptual models intend to predict only a single aspect of appearance, and most prior work has focused on using controlled stimuli targeted to elicit one type of effect only. To support models that can predict the multifaceted appearance of dichoptic stimuli, a better understanding of how multiple perceptual effect might co-occur is needed.

## 2.3 Perceptual Experiments

In this report, we present the results of two perceptual experiments designed to examine all five of the aforementioned perceptual factors in dichoptic appearance together. We aim to provide a more holistic picture of what dichoptic stimuli may look like to users and in this way inform display design decisions. For example, it would be beneficial to know if there is any systematic relationship between the different perceptual effects. Are different effects associated more or less with different amounts of interocular image differences? Is there a 'sweet spot' for optimal user experience where perceptual artifacts like binocular rivalry are minimized but the sense of depth or contrast is enhanced? How does the perceptual outcome change when viewing different spatial patterns?

In Experiment 1, we examine how spatial complexity and interocular contrast differences influence the occurrence of the different perceptual effects. This experiment uses conventional psychophysical stimuli. It aims to validate our multi-question experimental procedure and understand the potential relationships between the perceptual factors of interest. Experiment 1 was conducted as part of a larger psychophysical study, and some non-overlapping results from this study were already reported in [163] (Chapter 1). In Experiment 2, we leverage the paradigm from Experiment 1 to more directly examine how dichoptic imagery varies in appearance in optical see-through AR scenarios. We simulated

stimuli in which the AR display is brighter in one eye than the other, which results in both interocular differences in luminance and contrast together. Contrast for AR content is defined as the ratio of the maximum AR luminance over the maximum luminance of the background. While many of the presented results are relevant to VR as well, we focus on AR in this report because the optical and electronic demands on such systems necessitate challenging trade-offs that can be informed by a deeper understanding of dichoptic perception. In addition, current wearable systems for AR have an idiosyncratic appearance because the optically-overlaid virtual content is often semi-transparent. Stimuli with this appearance warrant dedicated investigation as our perceptual interpretations of them can be complex [65, 180]. In addition to exploring dichoptic contrast in binocular AR systems, we include conditions simulating monocular AR viewing, because monocular designs may be sufficient for some AR applications.

### 2.3.1 Participants

Two groups of thirty-four adults participated in Experiment 1 (23 females, ages 19-32 years) and Experiment 2 (25 females, ages 18-34 years). All participants had normal or corrected-to-normal visual acuity and normal stereo vision (measured with the Randot Stereotest). The experimental procedure was approved by the Institutional Review Board and all participants gave informed consent prior to beginning the study.

### 2.3.2 Experimental Setup

Stimuli were displayed on a mirror haploscope (Figure 2.3A) to allow for independent presentation of images to the left and right eyes (presented on two LG 32UD99-W LCD displays). The spatial resolution of each display was 3840 x 2160 pixels per eye (approx. 60-pixels-per degree), with a maximum luminance of 163  $cd/m^2$ . We measured the gamma nonlinearity of each display and used a calibrated look-up table such that the displays' output was linear. Participants were head-fixed with a chin rest and sat in the dark during the experiment.

### 2.3.3 Task

In a series of trials, participants were presented with pairs of stimuli to compare. One stimulus was presented on the top half of the screens and the other on the bottom half. One stimulus was identical in the two eyes (non-dichoptic) and the other stimulus (usually) comprised a dichoptic pair as described below. We call this second stimulus the *reference*. Participants used keyboard presses to adjust the contrast of a target pattern in the non-dichoptic stimulus to match the appearance of the target in the reference stimulus as best as they could (Figure 2.3A). They could look back and forth between the stimuli and could spend as much time as they needed to obtain the best match. The positions of the reference stimulus and the adjustable non-dichoptic stimulus were swapped for half of the participants.

After participants indicated that they had found the best match, the stimuli disappeared and they were shown several prompts to assess which, if any, perceptual differences there were between the reference and their best match (Figure 2.3B). They were first asked whether they were able to find an exact match or not. If the answer was no, they were asked to judge the contrast, brightness, luster, rivalry and depth of their best match against

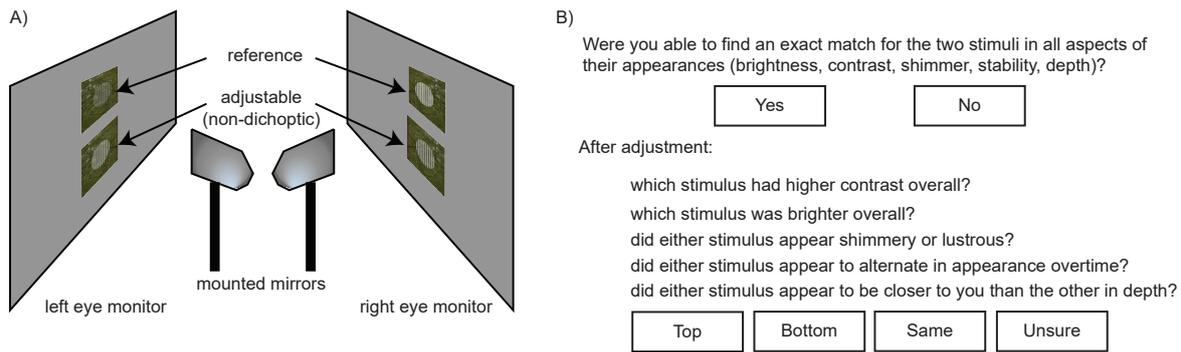


Figure 2.3: A) Experimental setup, in which two stimuli were shown to participants in a mirror haploscope. At the start of each trial, participants were asked to match the appearance of the two stimulus targets (e.g., the circular pattern) as best they could by varying the adjustable stimulus. B) Following the matching task, a set of follow up questions (exact wording in figure) and response options (boxed) were shown on the screen for the participants to select based on what they saw during the matching phase.

the reference stimulus. The prompts shown in Figure 2.3B were presented sequentially on screen. Response options to each prompt were top, bottom, same, and unsure. Responses of 'top' or 'bottom' indicated which stimulus was associated with the stronger perceptual effect. Based on pilot testing, we selected wording to describe luster and rivalry that best matched how participants described these effects (third and fourth questions, respectively). For the luster, rivalry, and depth questions, people were instructed to use the response option 'same' when neither stimulus had the effect. Prior to starting the experiment, participants were shown images to help them understand what was meant by luster and rivalry. We showed them a square stimulus with different shades of grey in each eye to explain what luster looks like and orthogonally oriented gratings in each eye to demonstrate binocular rivalry. Participants also completed ten practice trials to get familiar with the task.

### 2.3.4 Stimuli

In Experiment 1, we used gray-scale pattern stimuli to probe the nature of people's responses to the visual appearance questions. In Experiment 2, we used stimuli designed to mimic the appearance of optical see-through AR systems.

#### Experiment 1

We used two common types of vision research stimuli in this experiment: vertical sine wave gratings with a spatial frequency of 5 cycles-per-degree (cpd) and a 1/f ('pink') noise pattern with a broad frequency amplitude spectrum similar to that of natural images (Figure 2.4A). In a previous experiment, we found differences between the dichoptic contrast percepts of these grating and noise stimuli [163]. Therefore, in addition to these two stimuli, we also included three intermediate noise patterns that shared some similarities with the grating patterns (also shown in Figure 2.4A): we matched the pixel intensity distribution of the 1/f noise pattern to the grating through histogram matching (histogram-matched), we bandpass-filtered the 1/f noise image and only kept spatial frequencies near 5cpd (bandpass), and we repeated the first row of the 1/f noise image for all rows to create a broadband vertical grating (broadband). Each stimulus image was 8-bit and spanned the full range of 0-255 bit levels.

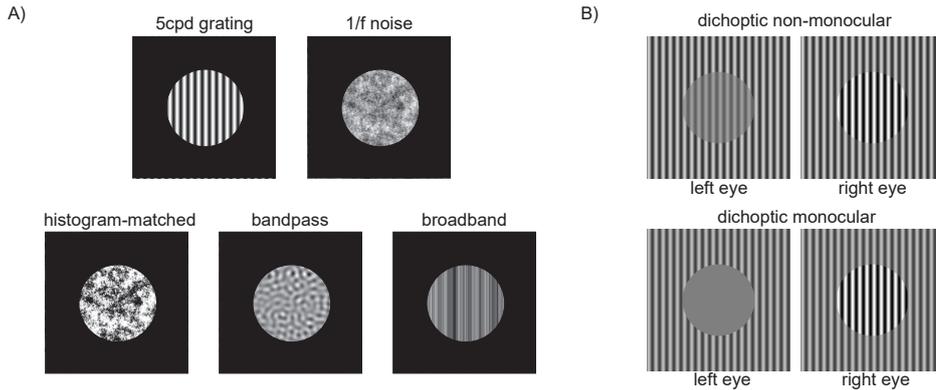


Figure 2.4: A) Five stimulus target patterns used in Experiment 1. B) An example of two types of dichoptic references that were used: non-monocular reference stimuli had different contrast for each eye’s target and both eyes’ target contrasts were greater than 0, whereas monocular reference stimuli only had a target visible in one eye. Recall that all targets were embedded in a square surround region that matched the average contrast in the reference targets.

Under realistic viewing conditions, targets of visual inspection are rarely viewed in isolation (i.e., against a uniform background). Instead, the surrounding context provides additional visual information that may play a role in determining the appearance. Thus, each image of the stimulus consisted of a  $2^\circ$  circular target of interest embedded in a binocular  $4^\circ$  by  $4^\circ$  surround region (Figure 2.4B). To vary the contrast of each eye’s target region, we normalized the image range from 0-1 and rescaled the values around the mean value as follows:

$$I = c(I_0 - \mu) + \mu, \quad (2.1)$$

where  $I_0$  denotes the original image (a 2D matrix),  $\mu$  denotes the mean pixel intensity of that image,  $I$  denotes the new image, and  $c$  is a scalar value that determines the amount of contrast reduction. To generate the reference stimuli, the contrast ( $c$ ) of the target for the left and right eyes ( $c_L$  and  $c_R$ ) was set to 0, 0.25, 0.5 or 1, resulting in 16 possible combinations between the two eyes (e.g.,  $c_L = 0.25$  and  $c_R = 1$ ,  $c_L = 1$  and  $c_R = 0.5$ , etc.). We did not present a stimulus with zero contrast in both eyes, so only 15 combinations were used. Of these, 6 combinations had  $c = 0$  in one eye and  $c > 0$  in the other eye, which we refer to as the special case of dichoptic stimuli with a monocular target. The other 6 dichoptic combinations were non-monocular (visible target in both eyes) (Figure 2.4B). The remaining 3 combinations resulted in non-dichoptic stimuli ( $c_l = c_R$ ) that were used as control/catch trials. The contrast of the square outside of the target region was always equal to the average contrast of the two eyes’ target regions and non-dichoptic. All stimuli were shown on a uniform mid-grey background. In total, there were 75 trials (5 stimulus patterns x 15 contrast combinations).

## Experiment 2

We created stimuli that simulated AR visual experiences by compositing a virtual icon with a naturalistic background image. The virtual icon was then used as the target for the perceptual task. We tested four different patterns for the virtual icons. To have a baseline comparison with Experiment 1, we included the 5cpd grating and 1/f noise pattern again. Based on the results of Experiment 1, we were interested in understanding if more realistic AR content would appear similar to the two baseline stimuli or not. We thus selected two different icon patterns from an existing library [1], which we refer to as simple and

complex icons (Figure 2.5A). The grating, noise, simple, and complex icon stimuli were all overlaid on an image of a natural background from the SYNS dataset [2] (Figure 2.5B, the same background image was used for all icons). Similar to Experiment 1, target regions (the icons) subtended  $2^\circ$  circles and the background region subtended a  $4^\circ$  square.

The contrast adjustment of the icons was performed similarly to Experiment 1, with some key differences to more closely simulate the joint contrast/luminance modulations that can occur when one display in an optical see-through AR system is brighter than the other. In particular, since these systems use additive light, we simulated the addition of the icon image onto the natural background. Pixel values of the background image ( $B$ ) were scaled down by a factor of 2 so that only half of our display’s dynamic range was used to simulate the background and the other half could be used for the icons. This effectively provides a maximum AR contrast of 2:1 against the background. The normalized 8-bit, 3 color channel icon images ( $A$ ) were also downscaled by a factor of 2 before being multiplied by the different scale factors, such that the maximum normalized pixel value in the combined image was equal to 1:

$$I = c \left( \frac{A}{2} \right) + \left( \frac{B}{2} \right). \quad (2.2)$$

All contrast adjustments were made in linear units based on the assumption that all color channels were encoded with a gamma non-linearity of 0.45 (e.g., normalized bit values from the background and icon images were exponentiated to  $1/0.45$  prior to being combined). That is,  $A$  and  $B$  in Equation 2.3 have the gamma non-linearity corrected. We used the same contrast combinations as described for Experiment 1 for the AR target in this experiment (Figure 2.5C). The surround region was identical in the two eyes. In total, there were 120 trials in Experiment 2 (4 stimulus conditions x 15 luminance combinations x 2 repeats).

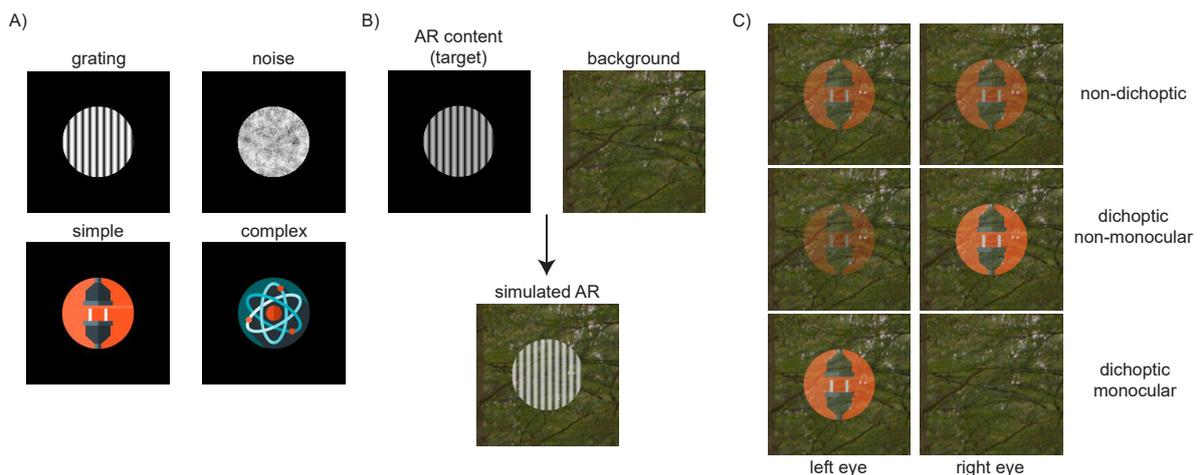


Figure 2.5: A) Four icon stimulus patterns used in Experiment 2. B) We simulated the appearance of an AR target on a natural background by compositing each icon with a forest scene. C) The AR target in the reference stimulus could be non-dichoptic, dichoptic but non-monocular, or fully monocular.

## Catch trials

On some trials in both experiments, we presented a non-dichoptic reference to check that participants were following the instructions. These also served as catch-trials because

participants should be able to find an exact match between the reference and the adjustable stimulus. We used the matching performance during these trials to exclude participants who were not performing the task reliably. Two participants from Experiment 1 and three participants from Experiment 2 were excluded because their matching error exceeded 1.5 times the interquartile range of all participants' errors. For the results presented below,  $N = 32$  for Experiment 1 and  $N = 31$  for Experiment 2.

## Statistical analyses

We used mixed-effect logistic regression models to fit the data and evaluate which stimulus properties were associated with different perceptual reports, with participants modelled as random effects. For each analysis, we include tables that report the coefficients, 95% confidence intervals,  $t$  statistics, and  $p$  values associated with a set of stimulus properties modelled as fixed effects. A qualitative examination of the data did not suggest that any notable interactions were present. We therefore do not report interactions. For some analyses, we use a separate model to examine the difference between responses to monocular targets and other dichoptic stimuli so that we can treat monocular versus non-monocular targets as a categorical predictor.

## 2.4 Results

### 2.4.1 Experiment 1

#### Probability of finding an exact match

The probability that participants could find an exact perceptual match to the reference stimulus varied systematically as a function of the interocular difference and the stimulus pattern, although there was substantial individual variation (Figure 2.6). Figure 2.6A shows how the magnitude of the contrast difference between the two eyes was associated with dramatic changes in the probability of a perceptual match. To characterize the contrast differences, we use the ratio of the higher contrast target to the lower contrast target, or interocular contrast ratio (ICR). A ratio of one means that the reference stimulus had the same target contrast in each eye and was non-dichoptic. As expected, participants were able to find an exact perceptual match close to 100% of the time when this was the case.

A larger ratio indicates a larger contrast difference in the two eyes (i.e., an ICR of four means one eye's contrast is four times the contrast of the other eye). As ICR increased from one to four, participants were on average less likely to find an exact match, with only about a quarter of the stimuli resulting in an exact match when the ICR was equal to four. We can take the coefficients from the regression model (Table 2.1) and exponentiate them to obtain the odds ratios for the predictors. The coefficient of -1.62 for the ICR translates to an odds ratio of 0.20, meaning that for each one unit increase in ICR, there's an 80% decrease in the odds of getting an exact match (probability of getting a match divided by the probability of no match).

Monocular reference stimuli, in which one eye had a target contrast of 0 (i.e., uniform gray embedded in a binocular surround region), have an ICR of infinity (Figure 2.6A, labelled as monoc). For these stimuli, we ran a separate regression model containing a

categorical predictor on a subset of the data (monocular trials versus trials with an ICR of 4). The results suggest that the probability of finding an exact perceptual match was not notably lower for monocular stimuli as compared to dichoptic stimuli with a large ICR (Table 2.2).

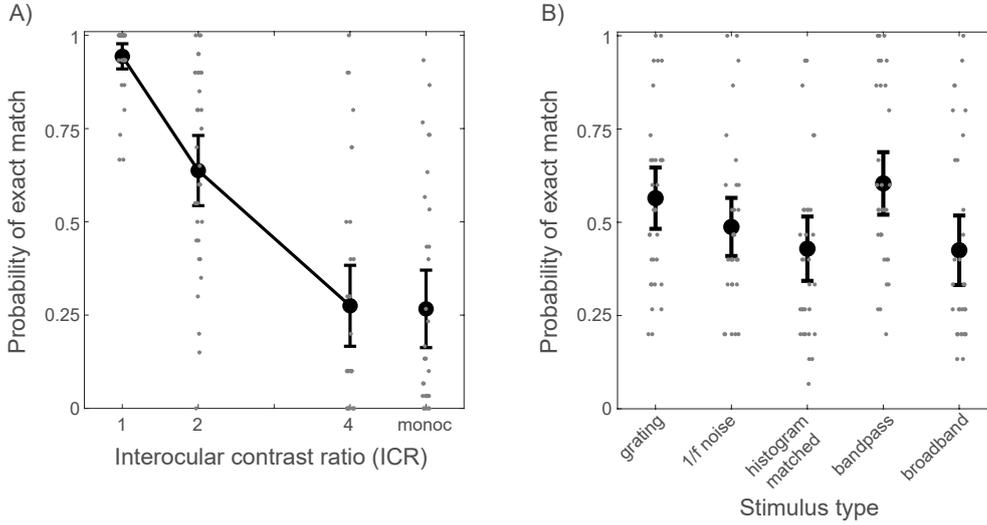


Figure 2.6: Results for the exact match question in Experiment 1. Large black dots represent the average probability of finding an exact perceptual match across subjects. Error bars are 95% confidence intervals. The smaller grey dots represent each participant’s data. A) The probability of exact matches across different interocular contrast ratios (ICR) in the two eyes, and for monocular (monoc) targets. B) The probability of exact matches for each stimulus pattern.

The probability of finding exact matches was much less affected by stimulus pattern (Figure 2.6B, 2.1). We found that the probability associated with the grating stimulus was not significantly different from the 1/f noise stimulus, but was significantly different from all three intermediate patterns. Compared to the match probability associated with the grating stimulus, there was a 75% increase in odds for the bandpassed noise, and 65% and 73% decreases in odds for the histogram-matched noise and broadband grating, respectively. This result suggests that the spatial pattern of the stimulus may influence the chances that people see phenomena like luster, rivalry, and depth in dichoptic stimuli, but the effect is much smaller compared to ICR.

Table 2.1: Logistic regression model for Experiment 1 exact match question. ICR is a continuous variable and stimulus types are categorical predictors (with grating used as the baseline). The coefficient reflects an increase or decrease in the probability that an exact match was obtained. Positive values indicate more exact matches, and negative values indicate fewer exact matches. Coefficients that are significantly different from zero based on the  $t$ -statistics are marked with asterisks (\*).

Experiment 1	Coefficient (95% CI)	$t$	$p$
Intercept	5.13 (4.27, 5.99)	11.74	< 0.001*
ICR	-1.62 (-1.81, -1.44)	-17.37	< 0.001*
1/f noise	-0.46 (-0.95, 0.02)	-1.84	0.06
histogram matched	-1.05 (-1.53, -0.56)	-4.28	< 0.001*
bandpass	0.56 (0.04, 1.08)	2.10	0.04*
broadband	-1.31 (-1.79, -0.83)	-5.36	< 0.001*

Table 2.2: Logistic regression model for comparing dichoptic trials with non-monocular targets (ICR equals 4) to trials with monocular targets. The non-monocular trials were used as the baseline. Data are reported in the same format as Table 2.1.

Experiment 1	Coefficient (95% CI)	$t$	$p$
Intercept	-1.70 (-2.51, -0.88)	-4.08	< 0.001*
monocular	-0.07 (-0.45, 0.30)	-0.39	0.70

## Perceptual appearance of dichoptic stimuli

What perceptual effects do people experience when they are unable to find an exact perceptual match by varying stimulus contrast, and how do these perceptual effects vary across different interocular contrast ratios and stimulus types? To answer these questions, we next look at participants' responses to the follow up questions about perceived contrast, brightness, luster, rivalry, and depth.

First, as a sanity check, we asked which stimulus participants reported seeing luster and rivalry in. We expected participants to select the dichoptic reference stimulus as the one that elicits these perceptual phenomenon, because the adjustable stimulus was always non-dichoptic and should not elicit luster or rivalry. When looking at the average response for each effect, we indeed find that the dichoptic stimulus was selected more often than the non-dichoptic stimulus when participants perceived luster or rivalry. When luster was detected in one of the stimuli, the reference stimulus was selected 98% of the time. For rivalry, it was 94%. When a depth difference was detected, participants also tended to indicate that the dichoptic stimulus was closer to 87% of the time. For the brightness and contrast questions, we did not expect participants to systematically select either stimulus because we do not have a strong hypothesis that dichoptic stimuli should appear systematically higher or lower in contrast or brightness than non-dichoptic ones. Indeed, the choices were closer to chance (56% and 61% of the time, respectively).

For the main analysis, we re-coded the data to simply indicate whether people perceived a difference or not for each perceptual factor. When participants made any response other than 'same' for a given prompt, a perceptual difference was considered to be present. The average number of 'unsure' responses across all the prompts was low (mean=5.95%, median=2.67% of all trials) and the results do not notably change if we omit these responses.

When the reference was non-dichoptic, there were minimal perceptual differences, as expected from the analysis of exact matches (Figure 2.7A). That is, on these trials participants were unlikely to indicate any perceptual differences between the two stimuli. As the ICR increased, all five effects started to become more noticeable. The most common perceptual differences across all contrast ratio levels were binocular luster, depth, and rivalry, in the order from most likely to less likely. The results also suggest that different effects tended to co-occur to some extent, because the proportions for high ICR trials sum to a value greater than one. Indeed, experiences of these perceptual phenomena were not mutually exclusive. Across all participants, the mean number of perceptual differences per dichoptic trial was greater than 1, with marginal statistical significance (mean=1.27, median=1.17,  $t(31) = 1.94$ ,  $p = 0.06$ ), and this amount increased with increasing ICR (for example, the mean and median were 1.60 and 1.31 for an ICR of 4).

We performed five logistic regressions to look at the occurrence of each perceptual

effect separately. Table 2.3 (left) shows the association between ICR (ICRs of one to four) and the presence of each perceptual difference. The ICR coefficients for all effects were positive and statistically significant, suggesting that the occurrence of all perceptual differences increased systematically as ICR increased. Based on the magnitude of the ICR coefficients, luster had the largest increase. In terms of odds ratios, we observed about a 314% increase in the odds of luster effects for each one unit increase in ICR, as compared to a 252% increase in rivalry, 263% in depth effects, and a 116% and 88% increase in odds of contrast and brightness effects respectively.

Next, we directly compared the trials with a monocular target to trials with a non-monocular but high ICR target (ICR = 4) (Table 2.3 right). The fits to this subset of trials indicates that the monocular targets were associated with a relative increase in binocular rivalry. The odds of perceiving binocular rivalry increased by 67% on trials with monocular targets, while no other effects were notably different.

Lastly, we looked at how the perceptual effects differed among different stimulus patterns. Figure 2.7B shows the occurrence of perceptual differences for each stimulus type out of the all dichoptic trials (ICR of 2, 4 and monocular). The brighter the color in the matrix, the higher likelihood that there was a difference associated with each effect (x-axis label) for the given stimulus pattern (y-axis label). The results suggest that different stimulus patterns can have a different set of perceptual effects. For example, the grating stimulus had fewer perceptual differences overall, and a slightly higher rate of rivalry than luster. The more complex patterns have relatively higher rates of luster, all of which exceeded the occurrence of rivalry. Taken together, this set of results suggests that rivalry may be a concern particularly for monocular stimuli and for simple grating stimuli. These results serve to highlight the importance of investigating these perceptual effects using visual stimuli that mimic the visual appearance of genuine AR experiences.

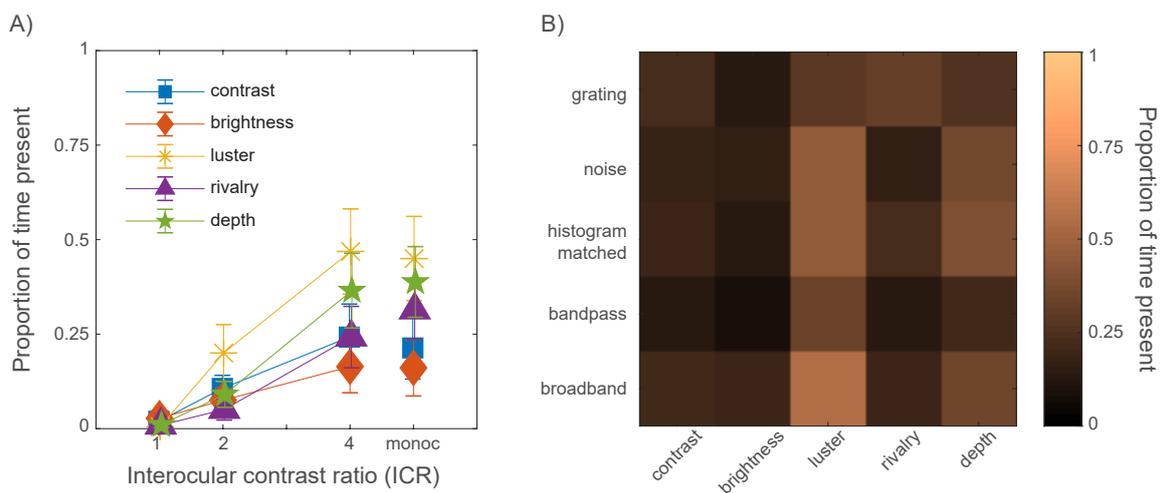


Figure 2.7: Results for the five perceptual effects measured in Experiment 1. A) The average proportion of trials across participants (with 95% confidence interval) in which each of the effects was present as a function of interocular contrast ratio (ICR), and for monocular targets. B) Heatmap showing the average proportion of time that each effect (x axis) was present for each stimulus type (y axis) across all dichoptic trials.

Table 2.3: Logistic regression models for the occurrence of perceptual effects in Experiment 1. Left: Each effect’s occurrence as a function of ICR (ICRs of 1,2, and 4). Right: Each effect’s occurrence difference between non-monocular (ICR of 4) and monocular dichoptic trials. Data are reported in the same format as Tables 2.1 and 2.2.

Expt 1	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Contrast			
Intercept	-4.41 (-5.07, -3.76)	-13.20	< 0.001*
ICR	0.77 (0.61, 0.93)	9.62	< 0.001*
Brightness			
Intercept	-4.46 (-5.14, -3.78)	-12.80	< 0.001*
ICR	0.63 (0.46, 0.80)	7.13	< 0.001*
Luster			
Intercept	-5.27 (-6.13, -4.41)	-12.02	< 0.001*
ICR	1.42 (1.25, 1.60)	15.59	< 0.001*
Rivalry			
Intercept	-6.65 (-7.65, -5.65)	-13.04	< 0.001*
ICR	1.26 (1.02, 1.49)	10.46	< 0.001*
Depth			
Intercept	-5.59 (-6.35, -4.82)	-14.40	< 0.001*
ICR	1.29 (1.12, 1.46)	14.65	< 0.001*

Expt 1	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Contrast			
Intercept	-1.48 (-2.07, -0.88)	-4.86	< 0.001*
monocular	-0.27 (-0.63, 0.08)	-1.51	0.13
Brightness			
Intercept	-2.25 (-2.95, -1.56)	-6.37	< 0.001*
monocular	0.01 (-0.40, 0.42)	0.05	0.96
Luster			
Intercept	0.13 (-0.68, 0.95)	0.32	0.75
monocular	-0.20 (-0.54, 0.13)	-1.18	0.24
Rivalry			
Intercept	-1.63 (-2.23, -1.02)	-5.25	< 0.001*
monocular	0.51 (0.17, 0.86)	2.91	0.004*
Depth			
Intercept	-0.61 (-1.36, 0.13)	-1.61	0.11
monocular	-0.02 (-0.33, 0.29)	0.12	0.90

## 2.4.2 Experiment 2

The stimuli used in Experiment 2 were designed to more closely mimic the visual experience of AR with natural backgrounds, partially transparent imagery, and a coupling of contrast changes with stimulus brightness.

### Probability of finding an exact match

The effect of ICR on the probability of finding an exact perceptual match was similar to Experiment 1 for the AR-like stimuli used in Experiment 2 (Figure 2.8A, Table 2.4). In this experiment, the probability of finding exact matches when the contrast ratio was high (ICR of 4) or when the stimulus was monocular was quite low. For each unit increase in ICR, the odds of getting an exact match decreased by 88%. For the four AR icon patterns used in this experiment, only the complex icon condition was associated with a match probability that was significantly different from the grating baseline, and this modulation was again substantially less than the differences associated with manipulating ICR (Figure 2.8B, Table 2.4). When comparing the trials with a monocular AR target against the trials with target ICR of 4, we found the monocular AR was associated with a significantly lower probability of finding an exact match (Table 2.5), with a 62% decrease in odds.

Taken together, these results replicate and extend the findings from Experiment 1 and show that there is substantial variation in the appearance of dichoptic stimuli that differ in interocular contrast, and that these appearances are subtly but lawfully modulated by the stimulus pattern.

### Perceptual appearance of dichoptic stimuli

We performed a set of analyses on perceptual appearance responses mirroring those described for Experiment 1. Similar to Experiment 1, the response pattern for luster and rivalry fits our expectations: when these effects were present, participants indicated that they saw them in the dichoptic reference 97% of the time. Again, we coded an effect as not present if participants responded with ‘same’, and present if they responded with one

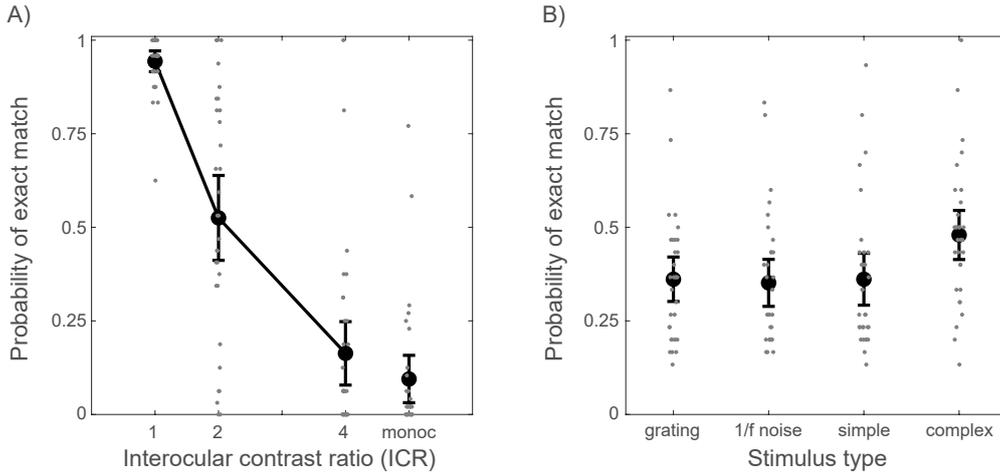


Figure 2.8: Results for the exact match question in Experiment 2. Large black dots represent the average probability of finding an exact perceptual match across subjects. Error bars are 95% confidence intervals. The smaller grey dots represent each participant’s data. A) The probability of exact matches across different interocular contrast ratios (ICR) in the two eyes, and for monocular (monoc) targets. B) The probability of exact matches for each stimulus pattern.

Table 2.4: Logistic regression model for Experiment 2 exact match question. ICR is a continuous variable and stimulus types are categorical predictors (with grating used as the baseline). The coefficient reflects an increase or decrease in the probability that an exact match was obtained. Positive values indicate more exact matches, and negative values indicate fewer exact matches. Coefficients that are significantly different from zero based on the  $t$ -statistics are marked with asterisks (\*).

Experiment 2	Coefficient (95% CI)	$t$	$p$
Intercept	4.82 (3.99, 5.66)	11.29	< 0.001*
ICR	-2.12 (-2.31, -1.93)	-21.74	< 0.001*
noise	-0.20 (-0.54, 0.14)	-1.14	0.26
simple	-0.18 (-0.52, 0.16)	-1.05	0.29
complex	1.52 (1.15, 1.91)	7.86	< 0.001*

Table 2.5: Logistic regression model for comparing dichoptic trials with non-monocular targets (ICR equals 4) to trials with monocular targets. The non-monocular trials were used as the baseline. Data are reported in the same format as Table 2.4.

Experiment 2	Coefficient (95% CI)	$t$	$p$
Intercept	-2.67 (-3.47, -1.87)	-6.54	< 0.001*
monocular	-0.98 (-1.36, -0.61)	-5.09	< 0.001*

of the other three options. The number of ‘unsure’ responses per participant was again low (mean=3.92%, median=1.67% of trials) and recoding these responses does not change the pattern in the results.

Across all participants, the mean number of perceptual differences per dichoptic trial was 2 (median 1.82,  $t(30) = 6.72$ ,  $p < 0.001$ ), and this amount increased with increasing ICR (for example, the mean and median were 2.54 and 2.42 for an ICR of 4). Similar to Experiment 1, the probability that participants reported any perceptual effect increased as the stimulus ICR increased (Figure 2.9A, Table 2.6 left). Luster was again the most commonly reported perceptual phenomenon associated with dichoptic imagery, and it increased the most with ICR. Across ICRs of 1, 2, and 4, there was a 479% increase in the odds of luster per unit increase in ICR, as compared to a 191% increase in rivalry, 262% increase in depth effects, 174% increase in contrast differences, and 191% increase

in brightness differences.

Importantly, there were notable differences in the responses associated with AR targets that had high ICR and AR targets that were fully monocular (Table 2.6 right). The probability of all effects except for luster substantially increased for the monocular target compared with the non-monocular high ICR target, resulting in a 87% increase in the odds for contrast differences, 31% for brightness differences, 421% for rivalry, and 320% for depth. Qualitatively, the probability of reporting luster was lower for the monocular target, but this difference did not reach statistical significance.

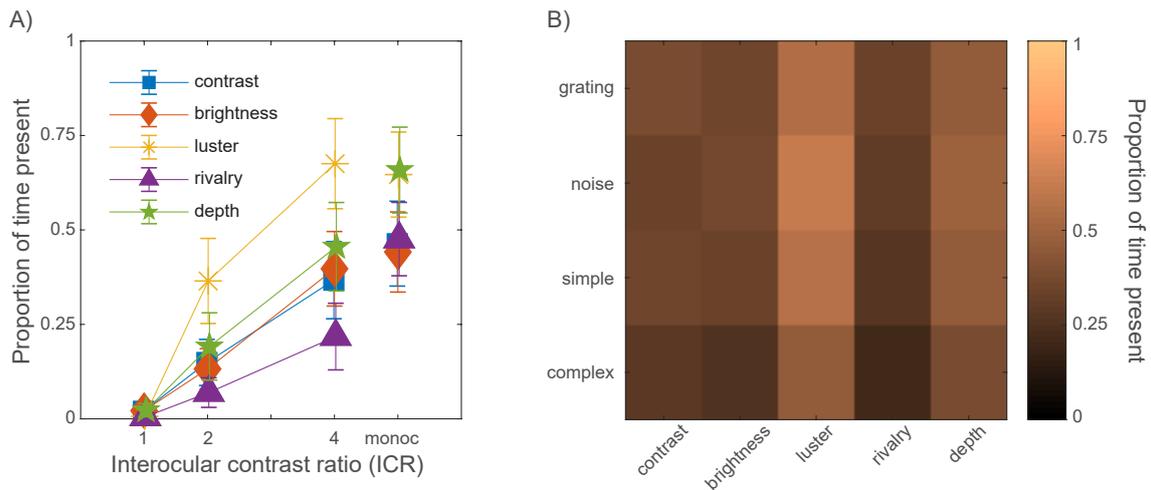


Figure 2.9: Results for the five perceptual effects measured in Experiment 2. A) The average proportion of trials across participants (with 95% confidence interval) in which each of the effects was present as a function of interocular contrast ratio (ICR), and for monocular targets. B) Heatmap showing the average proportion of time that each effect (x axis) was present for each stimulus type (y axis) across all dichoptic trials.

The association of each stimulus type with each perceptual difference is shown in Figure 2.9B. Overall, there was no strong stimulus dependent pattern. Unlike in Experiment 1, the grating stimulus was not associated with a unique pattern that deviated from the other stimuli when presented as a semi-transparent stimulus over a natural background. All stimulus types had binocular luster and depth differences as the predominant reported effects.

Table 2.6: Logistic regression models for the occurrence of perceptual effects in Experiment 2. Left: Each effect’s occurrence as a function of ICR (ICRs of 1,2, and 4). Right: Each effect’s occurrence difference between non-monocular (ICR of 4) and monocular dichoptic trials. Data are reported in the same format as Tables 2.4 and 2.5.

Expt 2	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Contrast			
Intercept	-4.67 (-5.30, -4.04)	-14.54	< 0.001*
ICR	1.01 (0.89, 1.13)	16.05	< 0.001*
Brightness			
Intercept	-4.49 (-5.23, -4.14)	-16.76	< 0.001*
ICR	1.07 (0.94, 1.19)	17.03	< 0.001*
Luster			
Intercept	-5.33 (-6.21, -4.44)	-11.83	< 0.001*
ICR	1.76 (1.59, 1.92)	20.87	< 0.001*
Rivalry			
Intercept	-6.01 (-6.80, -5.22)	-14.90	< 0.001*
ICR	1.07 (0.90, 1.23)	12.66	< 0.001*
Depth			
Intercept	-5.31 (-6.13, -4.49)	-12.73	< 0.001*
ICR	1.29 (1.15, 1.43)	18.10	< 0.001*

Expt 2	Coefficient (95% CI)	<i>t</i>	<i>p</i>
Contrast			
Intercept	-0.79 (-1.53, -0.06)	-2.13	< 0.001*
monocular	0.63 (0.36, 0.89)	4.69	< 0.001*
Brightness			
Intercept	-0.48 (-1.12, 0.16)	-1.48	0.14
monocular	0.27 (0.02, 0.52)	2.10	0.04*
Luster			
Intercept	1.20 (0.41, 1.99)	2.99	0.003*
monocular	-0.22 (-0.50, 0.06)	-1.54	0.12
Rivalry			
Intercept	-1.74 (-2.36, -1.12)	-5.54	< 0.001*
monocular	1.65 (1.36, 1.94)	11.12	< 0.001*
Depth			
Intercept	-0.32 (-1.12, 0.48)	-0.78	0.44
monocular	1.43 (1.15, 1.72)	10.00	< 0.001*

## 2.5 Discussion

### 2.5.1 Predicting Perceptual Artifacts in Dichoptic AR Stimuli

Our results highlight the importance of understanding dichoptic percepts with visual stimuli that closely match genuine AR experiences. For example, with the simple stimuli used in Experiment 1, we did not consistently observe any of the perceptual differences more than 50% of the time, but when we switched to AR-like stimuli in Experiment 2, we observed higher rates of luster and depth effects (about 75% in extreme dichoptic cases). We also observed a notable increase in reports of rivalry for AR-like monocular stimuli. We propose that perceptually-motivated guidelines for acceptable levels of ICR between two AR displays should thus be more conservative than might be assumed based on simpler stimuli. As another example, with the AR-like stimuli in Experiment 2, a complex natural background visible through the icon seemed to dampen the stimulus dependent variations. This equalizing effect may have to do with the natural background visibility balancing the two eyes’ differences – when both eyes can see the background behind an AR-icon, this reduces the average difference between the two eyes.

We can use the results from Experiment 2 to provide preliminary design guidelines for AR applications. As an example, we can consider a case in which we want to adopt a strict threshold on the probability that a dichoptic stimulus contains any perceptual effects that deviate from a comparison non-dichoptic stimulus. Collapsing across all of the stimulus types (i.e., removing them as model parameters) and refitting the data for the exact match question with a logistic regression model, we come to the following equation:

$$ICR_{max} = \frac{\log\left(\frac{P}{1-P}\right) - 4.70}{-1.95}, \quad (2.3)$$

where  $P$  is the designer-select minimum probability that the dichoptic stimulus matches a non-dichoptic one (i.e., no perceptual effect), and  $ICR_{max}$  is the maximum acceptable ICR. For example, if the designer aims for a threshold of  $P = 80\%$ , they should aim for an ICR of no more than 1.7. However, this result reflects the data on average, and given the large individual variation in our data a more stringent threshold may be appropriate to

accommodate users who are more sensitive to dichoptic perceptual effects. For example, a 90% threshold would be associated with an ICR of 1.3 or less.

## 2.5.2 Modeling Interocular Differences and Their Effect on Perception

The best metric for quantifying interocular differences remains an open area of research. Here, we used the ratio between the overall contrast in each eye as the summary measure of interocular difference. Our rationale is that human vision tends to follow Weber’s Law – for example, the amount of luminance difference required to detect a luminance change is proportional to the background luminance – and as such, this ratio is more likely to reflect the salience of the contrast differences in our stimuli than simply subtracting the two values [53]. However, there may be other metrics that could be more informative and practical. In particular, in Experiment 2 the stimuli differed in more than just contrast, so this metric is incomplete.

Ideally, perceptual metrics of interocular differences should account for both luminance and contrast, and even color. For example, the luminance adjustment applied to the colored AR icons in Experiment 2 could also result in interocular color differences, especially when viewing monocular AR on a binocular background (e.g., a red monocular target against a green binocular forest background), which is known to elicit perceptual effects such as luster as well [110]. Building better image-computable models of binocular combination will be crucial for developing unified metrics that can account for any arbitrary differences between the two eyes. However, modelling monocular human perceived contrast, let alone binocular contrast perception, for complex imagery remains an ongoing area of research [66, 113, 117].

Such models also have great appeal for developing tonemapping methods intended to improve image quality through binocular combination. For example, a recent line of research has looked at developing tonemapping methods that intentionally display different luminance and contrast information to the two eyes in order to improve the perceived visual quality of a stereoscopic display system that cannot reproduce the full dynamic range of luminance found in the natural environment. Yet, at present the results of these approaches are mixed [161, 174, 181, 182, 183]. In our experiments, we found that people did find there to be contrast and brightness difference between the best match and the dichoptic reference stimulus. Furthermore, looking at the four response types, participants tended to selected the dichoptic reference stimulus to be higher contrast and brighter in Experiment 2, suggesting the potential for dichoptic imagery to boost subjective image quality (The reference and the adjustable stimuli were about equally selected in Experiment 1).

## 2.5.3 Potential Benefits of Dichoptic Contrast in AR

Here, we focused primarily on the potential negative consequences of interocular differences in display brightness/contrast for users of AR systems. However, some of the perceptual phenomena we characterized may be desirable. For example, the appearance that a dichoptic stimulus is closer in depth might be helpful for heads-up AR systems that display icons floating in front of the environment. However, we found that this depth effect generally co-exists with other phenomena and may be challenging to isolate. For example,

as the likelihood of the depth effect increased, binocular rivalry also increased. Therefore, we did not find a 'sweet spot' of interocular contrast differences where desired effects may dominate undesired ones. However, some effects were more readily detected at a lower interocular differences than others. In both experiments, binocular luster was more detectable than the other effects, and the good news is that rivalry remained relatively uncommon in comparison. This may be beneficial if designers want to leverage binocular luster to create a shiny metallic appearance for the object without rivalry effects. The exception to this observation was during monocular viewing, particularly in the AR-like situation simulated in Experiment 2. In this experiment, observers detected rivalry about half of the time during monocular trials, suggesting that even if a binocular display has large interocular differences, it may be preferable to a monocular system if rivalry is a concern.

### **2.5.4 Potential Effects of Stimulus Motion**

Our study was limited to static imagery, but in the future it would be useful to investigate the effects of interocular differences for perception of dynamic content. For example, the Pulfrich effect is a change in perceived motion when one eye's image is darkened (for example, by applying a light filter such as sunglasses) compared to the other eye. It is thought Pulfrich effect is due to a neural delay in the registration of the two eye's inputs that lead to incorrect stereopsis [121]. For dynamic AR stimuli, there may therefore be additional constraints on interocular differences depending on the people's sensitivity to this illusory motion.

## **2.6 Conclusion**

Binocular displays can introduce unwanted visual differences between the left and right eye's views. Here, we focused on the perceptual consequences of contrast differences for optical see-through AR systems in particular, but such interocular differences can occur in any binocular display system. Across two experiments, our results suggest that the binocular appearance of dichoptic imagery is quite complex and multifaceted, and the magnitude of the interocular difference between the two eyes is a main predictor for the intrusion of potentially detrimental perceptual effects such as luster and rivalry. We also found evidence that the use of monocular AR displays notably increases the likelihood of eliciting binocular rivalry compared to other dichoptic scenarios. Our study results provide an overview of supra-threshold perceptual effects, but understanding detection thresholds for these effects will provide valuable and complementary information for display design. As we continue to improve our understanding of the perceptual phenomena associated with binocular differences in AR devices, a careful consideration of both the scope and strength of these phenomena can help guide design choices that support a high quality user experience.

# CHAPTER 3

## A re-examination of dichoptic tone mapping

Dichoptic tone mapping methods aim to leverage stereoscopic displays to increase visual detail and contrast in images and videos. These methods, which have been called both *binocular tone mapping* and *dichoptic contrast enhancement*, selectively emphasize contrast differently in the two eyes' views. The visual system integrates these contrast differences into a unified percept, which is theorized to contain more contrast overall than each eye's view on its own. As stereoscopic displays become increasingly common for augmented and virtual reality (AR/VR), dichoptic tone mapping is an appealing technique for imaging pipelines. We sought to examine whether a standard photographic technique, exposure bracketing, could be modified to enhance contrast similarly to dichoptic tone mapping. While assessing the efficacy of this technique with user studies, we also re-evaluated existing dichoptic tone mapping methods. Across several user studies, however, we did not find evidence that either dichoptic tone mapping or dichoptic exposures consistently increased subjective image preferences. We also did not observe improvements in subjective or objective measures of detail visibility. We did find evidence that dichoptic methods enhanced subjective 3D impressions. Here, we present these results and evaluate the potential contributions and current limitations of dichoptic methods for applications in stereoscopic displays.

### 3.1 Introduction

A good digital reproduction recreates the visual experience of a real scene using a different medium (e.g., a television, computer monitor, or head-mounted display). The reproduction process is primarily mediated by two devices: a camera and a display. Current pipelines face a number of challenges, particularly related to presenting the full range of visible light in natural scenes [37]. Two common challenges are the reproduction of the absolute range (lowest and highest values) and the reproduction of visually distinguishable differences in light intensity within that range. While both of these factors are perceptually relevant, the latter is more important for creating a reproduction with good fidelity of visual details.

Typical digital reproductions often lack visible details in the highlights or low-lights (shadows) of a scene. This loss of detail occurs because the dynamic range of typical

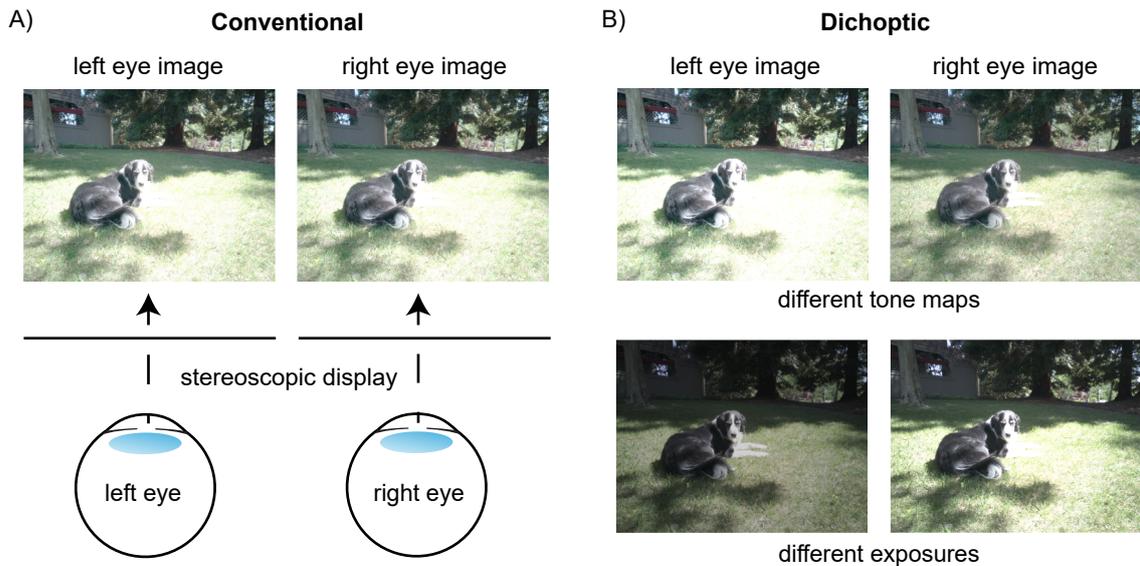


Figure 3.1: Example images shown to each eye on a stereoscopic display using either conventional digital reproduction methods (A) or dichoptic tone mapping/exposure methods (B,C). (A) In conventional methods, both eyes view images that were produced with the same tone mapping algorithm and camera settings (usually captured from different viewpoints to create binocular disparities). With dichoptic methods, the eyes see images that have different tone maps, emphasizing different areas of detail/contrast in the two eyes (B), or images that have the same tone map, but that were captured with different camera exposures (C). These examples were derived from the HDR+ Burst Photography Dataset [64], and panels A and B use tone maps from [182], reproduced with the authors permission.

cameras—their ability to represent visually distinguishable light intensities—cannot fully capture the range of detail that the human eye can see. High-end cameras can now represent a wider dynamic range directly, however, these cameras are not broadly available [70]. Instead, various methods exist to synthesize high dynamic range (HDR) images from low dynamic range (LDR) camera images, commonly by computationally combining multiple exposures [8, 29, 35, 107].

Importantly, even when HDR images can be generated, scene reproduction is still limited by the light levels that a display can produce. A tone mapping algorithm is typically used to define a function that maps values from HDR space to an LDR space appropriate for the display. However, this process is lossy and can introduce artifacts [37]. For example, traditional compressive tone mapping curves preserve visible contrast across a certain range of light levels at the expense of reducing contrast in other ranges. Local tone mapping algorithms preserve contrast in fine details, at the expense of reducing global luminance differences across regions of a scene (e.g. [45, 50, 140]). At present, there is no practical way to faithfully reproduce the full range of visible light intensities in natural scenes with typical devices.

With the rise of stereoscopic displays to support virtual reality (VR) experiences, a range of new approaches have been proposed that aim to leverage binocular vision to increase visible contrast and detail [51, 174, 181, 182, 183]. These approaches use different tone mapping algorithms for the images shown to the left and right eye. The term *dichoptic* refers to situations in which the two eyes are presented with different images, so we will use *dichoptic tone mapping* to refer to these techniques. We will refer to the two different images that make up a given dichoptically tone mapped pair as the *component*

*images*. Examples of a conventionally versus dichoptically tone mapped image pair are shown in Fig. 3.1 (A,B). If the increased overall contrast from dichoptic tone mapping can be integrated effectively by the visual system, these approaches are appealing for use in VR systems. Motivated by this idea, we sought to extend existing dichoptic tone mapping algorithms with a simple dichoptic exposure method inspired by photography (Fig. 3.1C). However, in the process of conducting a set of user studies, we were unable to find evidence that dichoptic tone mapping methods [182, 183] (selected as prior state-of-the-art) or the proposed exposure-based method consistently improved overall subjective image quality or detail visibility. We did observe a substantial enhancement in subjective 3D impressions [183]. These results are important to consider when evaluating how to capture and tone map imagery for stereoscopic displays, and highlight the challenges and opportunities associated with designing dichoptic methods that consistently improve upon conventional imaging approaches.

Our work makes three primary contributions:

1. We re-evaluate previous dichoptic tone mapping methods, along with dichoptic exposures, in a series of six user studies.
2. We provide a new guideline for assessing dichoptic tone mapping, whereby the dichoptic pairs should be compared to each of the component images.
3. We introduce a performance-based perceptual measure for quantifying detail visibility in images to complement subjective metrics used in typical user studies.

## 3.2 Related Work

### 3.2.1 Binocular Combination

During normal experience, the visual system is tasked with processing slightly different views in the two eyes, which must be combined to generate a unified percept of the world. The binocular differences encountered during normal experience primarily manifest as *binocular disparities*: slight offsets between features in the two eyes' images that occur due to the horizontal displacement between the eyes. These binocular disparities are used by the visual system to infer the relative depths of objects and surfaces in the environment [78]. Stereoscopic displays show different images to the two eyes, creating binocular disparities and a compelling three-dimensional (3D) percept.

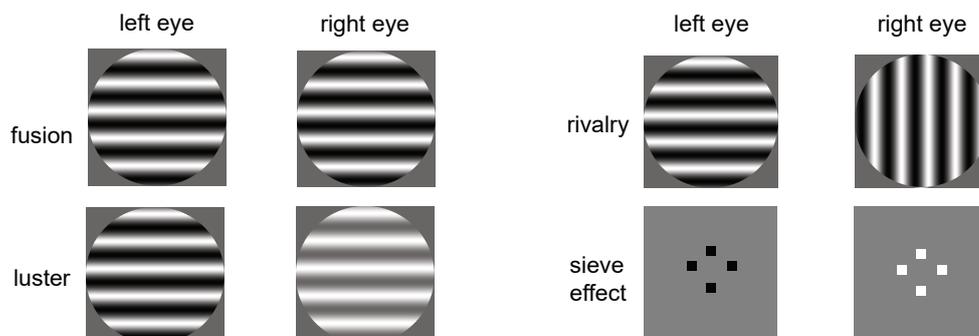


Figure 3.2: Examples of simple images used to study binocular visual combination. These stimuli provide illustrative examples of those that elicit fusion, luster, rivalry, and the sieve effect.

The binocular differences created by dichoptic tone mapping are different from binocular disparities: the two tone maps create images that differ in luminance, color and/or contrast, rather than viewpoint. This manipulation can be presented alone (i.e., the eyes see the same scene view but with different tone maps) or in addition to binocular disparity (i.e., the two eyes see different views of the scene and different tone maps).

Basic vision research suggests that introducing a luminance/contrast difference between the two eyes' images can elicit a range of perceptual phenomena. When the two eyes view images that are similar in appearance, the visual system is said to 'fuse' the images into a single, unified percept (Fig. 3.2, top). Extensive psychophysical studies using simple, synthetic stimuli have shown that fusion can be achieved even when the two eyes' images differ in luminance or contrast [34, 103, 104]. However, the fused percept is not typically well-described as a simple average of the two eye's images. For example, studies in which sine wave gratings of different contrast were presented to each eye showed that the binocularly fused percept is generally dominated by the eye seeing higher contrast [82, 103]. If this "winner-take-all" phenomenon is applicable to viewing more complex natural images, fusion may be exploited by dichoptic tone mapping to produce a binocular percept of greater overall contrast than either of the component images. For example, if each component image has some regions with more visible contrast than others, the binocularly fused percept may combine the higher contrast regions of both components to better convey the visual details of the physical scene.

However, viewing mismatched images in the two eyes can also result in a range of secondary perceptual effects. Two such effects are luster and binocular rivalry. *Luster* tends to occur when the overall pattern seen by the two eyes is relatively similar but the luminance, contrast, or color differs [53, 160] (Fig. 3.2, top middle). The term luster refers to the fact that stimuli with this property often appear shimmery, glossy or metallic. *Binocular rivalry* tends to occur when the eyes view two geometrically dissimilar patterns [105]. For example, if one eye sees a horizontally oriented sine wave grating and the other eye sees a vertically oriented sine wave grating, people typically see an unstable pattern alternating over time between the horizontal and vertical gratings, interleaved with periods in which a mixed pattern is seen (Fig. 3.2, bottom middle). Finally, in some cases viewing different luminance/contrast in the two eyes can elicit an enhanced perception of depth, even when binocular disparities are absent [73]. This is referred to as the *sieve effect* (Fig. 3.2, bottom). The perceptual mechanisms that drive the sieve effect are not understood but appear to differ from those that rely on binocular disparities [158]. While luster and binocular rivalry may often be considered undesirable perceptual artifacts, the sieve effect may actually produce a desirable enhancement of 3D percepts.

## 3.2.2 Dichoptic Tone Mapping Methods

### Algorithms

Several dichoptic tone mapping methods have been proposed in recent years that rely on similar principles [51, 174, 181, 182, 183]. Essentially, one eye's image is tone mapped to maximize the visible contrast across a certain luminance range and the other eye's image is tone mapped to maximize the visible contrast across a different range.

Yang et al. [174] were the first to propose this method to improve visual richness. They created an optimization algorithm that runs on an image-by-image basis, using an existing

standard tone mapping algorithm for one eye’s image and creating a paired image with a different tone map for viewing on a stereoscopic display. The algorithm for generating the second image was designed to maximize the visual differences between the pair of images while maintaining conservative visual comfort and fusion criteria, based on a perceptual model.

This method was followed up on in several additional studies [51, 181, 182]. In particular, Zhang et. al. [182] proposed a method for creating dichoptic pairs in which both tone maps can be optimized simultaneously. For this approach, they used an existing bilateral tone mapping algorithm in order to maximize global contrast in one eye’s image and local contrast in the other eye’s image [45]. The global contrast tone map is obtained by using a filter that maintains sharp edges but smooths out local contrast and therefore substantially lacks detail. This algorithm was used in concert with a perceptual model similar to the one used by Yang et al. [174] to optimize visible contrast and fusibility.

Earlier dichoptic tone mapping methods were not applicable to real-time applications (as is needed for interactive VR experiences) because running the optimization could take several seconds per image [174, 182]. More recent work has thus aimed to speed up the required processing time. Feng et al. [51] proposed a method to ensure temporal continuity for dichoptically tone mapping video frames, but their algorithm still required off-line processing. In a promising development, Zhang et al. [181] recently used GPU acceleration and a neural network approach to substantially reduce processing time, allowing the potential for real-time frame-by-frame deployment. Most recently, Zhong et. al. [183] created a dichoptic tone mapping method (referred to as DiCE, for Dichoptic Contrast Enhancement) that uses fixed tone mapping curves for each eye’s image. DiCE tone mapping curves can be applied to either HDR or LDR inputs and are pre-optimized with user studies to reduce binocular rivalry, eliminating the image-by-image optimization step present in previous methods.

## User studies

The effectiveness of new tone mapping algorithms is often assessed subjectively with user studies in which a group of observers are asked to view a set of example images and either rate or rank them according to some subjective criterion (e.g., [25, 44, 98, 100, 177, 178]). Two types of procedures are commonly found in subjective image assessments: image ratings (i.e., Likert scales) and forced choice (i.e., comparing images and choosing the best one).

Previous reports evaluating dichoptic tone mapping methods have used two-alternative forced choice paradigms (2AFC) to ask observers to make a choice between a standard tone mapping method (e.g., a nondichoptic method) and the newly proposed dichoptic method on some image quality criteria. The strength of 2AFC is that it gives reliable and sensitive results [112], but the downside is that it is not always obvious what the best approach is for selecting the *standard* method for comparison.

What is an appropriate standard to compare dichoptic methods against? At the core of the dichoptic tone mapping idea is the notion that the fused image is better than a conventional nondichoptic image along some desired perceptual dimension, such as image quality or visual contrast. Thus, it is important to establish that a given dichoptic tone mapping method is preferred over a reasonable state-of-the-art nondichoptic method.

However, it is also important to rule out the possibility that either of the component images is consistently preferred over the dichoptic pair. For example, if nondichoptic viewing of a component image is preferred over viewing the dichoptic pair, it would mean that just one of the tone mapped images (rather than the dichoptic presentation per se) may be driving subjective preferences. On the other hand, if dichoptic viewing is consistently preferred over both component images on some desired perceptual dimension, it provides evidence for the theory that binocular fusion can be exploited to increase subjective assessments. That is, that the dichoptic percept benefits from incorporating desirable features from both eyes' images.

Previous studies have used a range of nondichoptic standards for comparison during 2AFC tasks, but often omitted direct comparisons with both component images. Yang et al. [174] included a comparison with just one of the component images and found their dichoptic method to be better in terms of visual richness. Zhang et al. [182] used a mixture of standard conditions that included a nondichoptic tone map (the average of the two component images) and several dichoptic methods. The results suggested that their new dichoptic method was better than the nondichoptic average in terms of overall contrast, detail, and preference. But no comparisons with the component images were reported. Zhong et al. [183] evaluated both their DiCE method and Zhang et al.'s [182] method against one of Zhang et al.'s component images. With respect to overall image preference, the authors did not find consistent improvement in image quality for either dichoptic method relative to the component. When participants were instructed to make their response on the basis of visual contrast, they found consistent evidence for improved perception of contrast with both methods.

### 3.2.3 Current Study Motivation

When our own pilot studies began producing negative results for improved image quality and perceived contrast/detail with selected dichoptic methods, we developed a hypothesis that this important choice of the standard for comparison may account for some of these differing results. We thus designed a set of user studies to address this possibility, particularly exploring a range of nondichoptic alternatives for comparison. We used these studies to evaluate two previously proposed methods: the dichoptic tone mapping methods described in Zhang et al. and Zhong et al. [182, 183], because they provide two recent examples of complementary approaches to tone map generation. In addition, we evaluated a simple exposure-based dichoptic method as described in the next section.

## 3.3 Exposure-Based Dichoptic Method

To explore additional possibilities for real-time dichoptic contrast enhancement, we tested a method that creates dichoptic images during the capture stage instead of during post-capture tone mapping. This method can serve as a comparison to the more sophisticated tone mapping techniques. Specifically, rather than applying two different tone maps, we simply showed differently exposed images to the two eyes (Fig. 3.1C): one eye sees a high exposure image that contains better details in low-light areas and the other eye sees a low exposure image that contains better details in highlight areas. One benefit of this *dichoptic exposure* method is that it does not rely on having an HDR image as the input

and could bypass the HDR imaging and tone mapping pipeline. In this way, it is conceptually similar to computational photography algorithms that generate an HDR image from multiple LDR images with bracketed exposures [35, 109, 111], but we are asking the human visual system to combine two different exposures binocularly, rather than digitally. This process is advantageous for saving computing time during real-time applications (similar to the DiCE method [183]). However, a downside of this exposure-based method is that there is no explicit optimization for reducing perceptual artifacts. Specifically, inter-ocular differences in luminance and contrast are not constrained systematically as in dichoptic tone mapping. Therefore, we would predict that binocular rivalry artifacts might be more likely to occur and reduce perceived image quality. In this sense, this exposure method can serve as a baseline dichoptic method to evaluate the efficacy of the perceptual optimizations performed for other dichoptic methods. In the stimuli for our user studies, we simulated dichoptic exposures, as described in more detail below (Section 4.1.2).

### 3.4 User Studies

To determine whether dichoptic reproduction methods consistently produce superior images as compared to nondichoptic methods, we asked participants to subjectively judge the overall image quality of a set of images with different tone maps and exposures applied. We also included judgements based on two more specific criteria: detail visibility and 3D impression. To thoroughly test each dichoptic method, we first conducted a rating study that allowed us to compare dichoptic methods against several nondichoptic methods in a time-saving manner. The best rated nondichoptic methods were then directly compared to each dichoptic method in a 2AFC task. Lastly, to gain a better understanding of the practical implications of dichoptic methods beyond subjective judgements, we conducted an exploratory experiment to assess the visibility of details using an objective performance measure. Overall, we conducted six experiments with a total of 88 participants. The tasks and response criteria used in each of the experiments is outlined in Table 3.1. All participants had normal or corrected-to-normal visual acuity in both eyes and normal stereo vision as determined by a Titmus test. All participants gave informed consent prior to starting the experiment, were compensated for their time, and were naïve to the study hypotheses and goals. The procedures were approved by the institutional review board at the University of California, Berkeley.

Table 3.1: Overview of experiments. Asterisk indicates that the same 16 people participated in Experiments 1 and 2.

Exp.	Task	Criterion	N	Scenes
1	rating	image quality	16*	18
2	rating	detail visibility	16*	18
3	rating	3D impression	16	18
4	2AFC	image quality	16	18
5	2AFC	detail visibility	16	18
6	performance	detail visibility	24	8

### 3.4.1 Stimulus Generation

For each experiment, we included several different dichoptic and nondichoptic conditions (listed in Table 3.2). By condition, we refer to a method for generating a left/right image pair. Image pairs were generated from natural photographs inline with the stimuli used in previous studies. The image pairs were all generated from a publicly available dataset of HDR images (the HDR+ Burst Photography dataset) [64]. Note that while these image pairs were viewed on a stereoscopic display, because they were generated from a single HDR image they did not have any binocular disparities. This is consistent with the majority of stimuli used in previous dichoptic tone mapping studies [174, 181, 182, 183]. In Experiments 1-5, we generated the stimuli for each condition from 18 unique HDR scenes. Because Experiment 6 involved more trials, we used a subset of eight scenes (see Section 4.4.2 for details).

Table 3.2: Summary of dichoptic and nondichoptic conditions in experiments.

Dichoptic Conditions	nondichoptic Conditions
Zhang et al. 2018 [182]	average tone map global tone map local tone map
Dichoptic exposure	proper exposure high exposure low exposure
Zhong et al. 2019 (DiCE) [183]	C1 (low-light detail) C2 (highlight detail)

#### Zhang et al. (2018) dichoptic tone mapping

Four conditions were selected to examine the perceptual effects of the dichoptic tone mapping algorithm proposed by Zhang et al. in 2018 [182]. The tone mapped images for these conditions were obtained directly from the authors. In the *dichoptic* tone map condition, one eye’s image emphasized local contrast, and the other eye’s image emphasized global contrast (Fig. 3.3 C, D). The details of how these two images were generated is described in the original paper [182]. In brief, the authors optimized a tone mapping parameter that defines the global contrast of a base-layer, which is then combined with the image details [45]. We included three *nondichoptic* conditions for comparison. In two conditions, both eyes viewed one or the other of the component images that made up the dichoptic pair: we will refer to these two conditions individually as the *local* and *global* conditions. In the *average* condition, both eyes viewed an image generated with the average parameter from the local and global images. This was the standard used for comparison in the original study [182].

#### Dichoptic exposure

We also simulated the proposed dichoptic exposure method and generated a set of nondichoptic alternatives. One can think of increasing exposure as increasing the number of photons arriving at the camera sensor by some gain factor. To simulate increasing or decreasing camera exposure, we thus applied multiplicative gain factors to the HDR pixel values, with larger gain factors simulating longer exposures. First, we normalized the HDR pixel values to range from 0-1. Next, we applied a gain factor (see below) and gamma corrected by 0.45. We capped any resulting values that were greater than

1 (simulating camera saturation) and converted to 8-bit precision. Following standard practice in photography, we applied a range of gains (from 0.03 to 4, in steps of  $\sim 0.2$ ) and selected a proper exposure with minimal clipping (the percentage of pixel values in the lower and upper 4% of the 8-bit range). If multiple exposures of the same scene had similar clipping, we manually selected one. We next simulated a high and low exposure by generating images that were three gain steps above and two gain steps below the proper exposure. These steps were the same for all images, and were chosen to maximize differences in detail visibility without a substantial difference in mean luminance. Matlab code for generating this simulation on a typical HDR image is available upon request.

It is important to note that this approach only simulates the proposed dichoptic exposure pipeline. In a real-time application of this pipeline, the two cameras in a stereo camera pair would be simply set to a constant difference in exposure or gain. This could be implemented with a modification to existing automatic exposure methods, whereby after the ideal exposure level is detected, a constant offset is introduced to set one camera slightly lower and one camera slightly higher. The subsequent image capturing would be done without further adjustment or exposure bracketing needed.

Four conditions for the exposure manipulation were generated from these simulations. In the *dichoptic* exposure condition, one eye viewed the low exposure and one eye viewed the high exposure. Example dichoptic pairs are shown in Fig. 3.3 (A, B). There were three *nondichoptic* conditions, that were either *proper* exposures, *high* exposures, or *low* exposures.

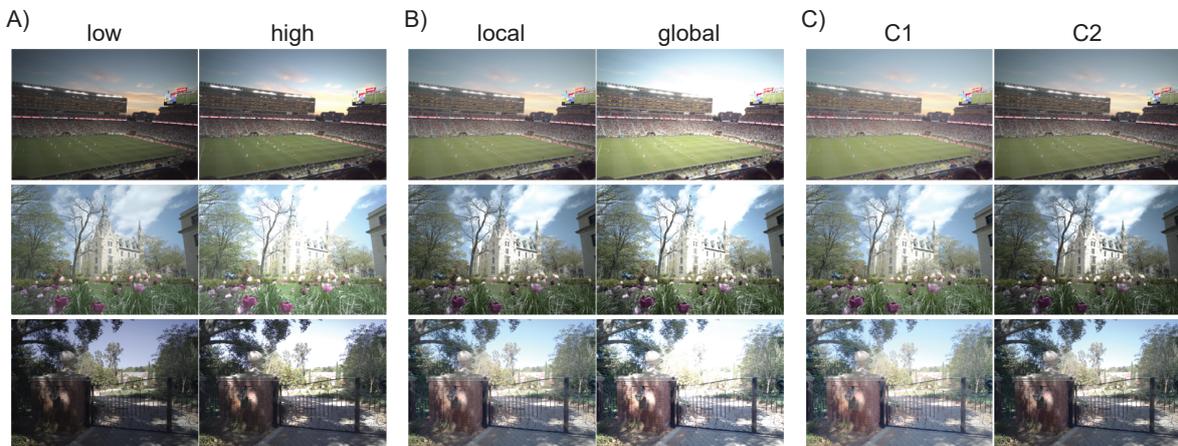


Figure 3.3: Three example scenes from the HDR+ Burst Photography dataset [64], with each dichoptic image pair illustrated. These include dichoptic exposures (A,B), the dichoptic tone mapping method from Zhang et al. [182] (C,D), and the dichoptic tone mapping method from Zhong et al. (DiCE) [183] (E,F). In A/B, low/high refer to the exposure level. In C/D, local/global refer to the tone mapping algorithm. In E/F, C1/C2 refer to the DiCE tone mapping curves.

### Zhong et al. 2019 (DiCE)

For Experiments 4 and 5, we also generated conditions to include the recently proposed DiCE dichoptic tone mapping method [183]. We implemented the DiCE algorithm in Matlab and confirmed that the resulting images matched the DiCE test images available online. The original LDR input image for DiCE always had the local tone map from Zhang et al. applied [182]. We used the tone curve parameters from the authors' main evaluation experiment (their Experiment 2), with an 1/h ratio of 0.63 and two segments

for each tone curve. We included the full *DiCE* dichoptic pair, as well as each component tone map, which we refer to as *C1* (better contrast for low-lights) and *C2* (better contrast for highlights). Example dichoptic pairs are shown in Fig. 3.3 (E, F).

### Stimulus presentation

For all experiments, a mirror haploscope with two LCD displays was used to present separate images to the left and right eye. The displays (DELL U2415) had a pixel resolution of 1920x1200, screen size of 52cm x 32.5cm, and refresh rate of 60Hz. The maximum luminance was approximately 174 cd/m<sup>2</sup>. During the experiments, the participant sat in a dark room and viewed the displays from approximately 60 cm away with their head on a chin rest. The experiments were controlled by Psychtoolbox [83] in Matlab and responses were made on a keyboard.

### 3.4.2 Experiments 1, 2, and 3: Preference, Subjective Detail Visibility, and 3D Impression Ratings

In the first three experiments, we re-evaluated the dichoptic tone mapping method proposed by Zhang et al. [182] along with the new dichoptic exposure method using a set of image rating tasks. Instead of doing a 2AFC task with a single nondichoptic standard, we asked participants to rate the images one at a time, allowing us to efficiently compare dichoptic methods with multiple nondichoptic conditions [112].

#### Participants.

Sixteen adults participated in both Experiments 1 and 2 (13 F, 3 M, age range 19-30). The order of the experiments was randomized such that half of the participants completed Experiment 1 first and the rest completed Experiment 2 first, with at least 1 week in between. Sixteen additional adults participated in Experiment 3 (13 F, 3 M, age range 18-40).

#### Procedures.

In Experiment 1, participants were instructed to rate a series of images based on their overall preference on a scale 1 to 5, with 5 being best. The specific instructions were to rate based on their “preference for image quality captured by the camera and not the scene content”. Prior to starting the experiment, participants were briefly shown each stimulus once to give them an understanding of the range of quality. There were 144 unique stimuli (8 conditions x 18 scenes) presented in randomized order, and each stimulus was shown twice for a total of 288 trials. During the repeated trials, we switched which eye saw which image. For example, for the *dichoptic exposure* condition, on half of the trials the left eye saw the high exposure and on the other half of the trials the right eye saw the high exposure. The presentation time for each stimulus was 6 seconds. In the Supplementary Material, we provide an analysis of potential contributions of left or right eye dominance to user preferences. Consistent with prior work, we do not find evidence for a substantial contribution of eye dominance, so we averaged across the two repeats for the main results [174, 183].

The procedures for Experiments 2 and 3 were the same as Experiment 1, except that the instructions differed. In Experiment 2, participants were instructed to rate based on

their “impression of detail visibility across the scene”. During the instructions, they were shown an example scene (not from the actual experiment) to explain the concept of detail visibility. We showed them a low exposure, high exposure, and proper exposure of this scene, pointing out the details in the highlights and low-lights. We refined this method over pilot testing, to ensure that the naïve participants understood the concept of image contrast and visible detail. In Experiment 3, participants were instructed to rate how 3D each scene looked. They were told that “we are interested in whether certain types of image adjustments look more 3D to you. Some images might have worse quality in terms of visible details, but please only focus on how 3D each scene appears.” We refer to participants’ responses to this prompt as their *3D impression*. While 3D enhancement is not the stated goal of dichoptic tone mapping and exposure, we included this task because previous work, and our own pilot testing, suggested that these methods can create an enhanced 3D impression (perhaps related to the sieve effect) [183].

### Results for dichoptic versus nondichoptic.

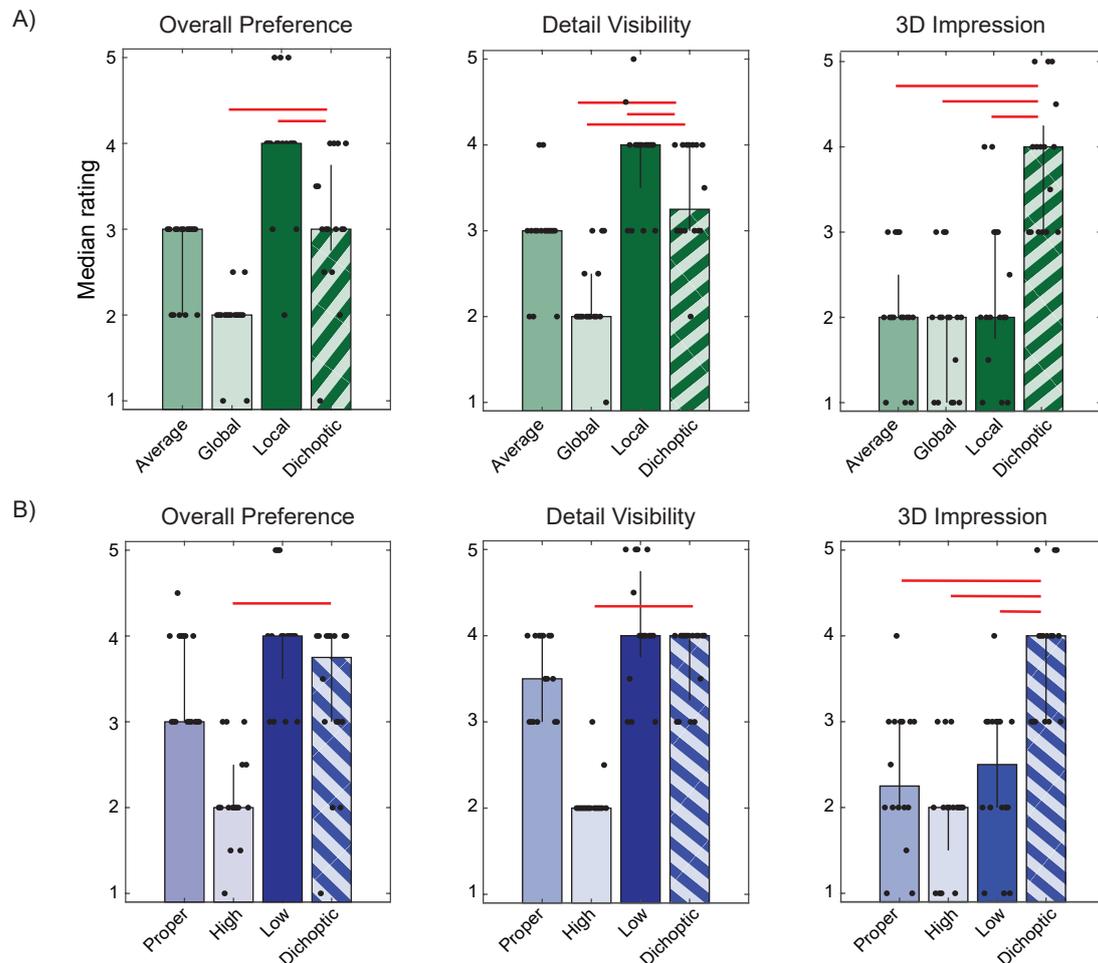


Figure 3.4: Results of rating tasks in Experiments 1, 2, and 3. The top panels (A,B) correspond to the overall preference ratings for the tone mapping (A) and the exposure (B) manipulations in Experiment 1. The middle panels (C,D) correspond to the perceived detail ratings in Experiment 2, and the bottom panels (E,F) correspond to the 3D impression ratings in Experiment 3. Each black dot represents an individual participant’s median rating for that task and condition (labeled on the x-axis). Dots are jittered for visibility. Each bar represents the median rating across all participants’ medians. Error bars indicate the 75<sup>th</sup> and 25<sup>th</sup> percentiles. Red lines mark statistically significant pairwise comparisons (see Table 3.4)

We analyzed the participant ratings first with a Friedman test to separately compare the tone mapping conditions and the exposure conditions. The results showed that there were significant differences among the groups of conditions across all three experiments (Table 3.3). We also found that there was relatively high agreement across participants (as determined by Kendall’s coefficient of concordance, Table 3.3), except that there was only moderate agreement for the ratings of overall preference in the exposure conditions.

The results for overall preference (Experiment 1) and detail visibility (Experiment 2) are plotted in Fig. 3.4 (A-D) and pairwise statistical tests are reported in Table 3.4. Starting with the four tone mapping conditions (Fig. 3.4 A,C), we did not find evidence that images in the dichoptic condition were consistently rated higher than the highest rated nondichoptic condition. When compared to the two component images, the dichoptic condition was rated significantly better than the global tone map in both Experiment 1 and 2, but was rated lower than the local tone map (this difference was only statistically significant in Experiment 1). The low ratings for the global tone maps are not too surprising. Viewing these images with smoothed local contrast likely does not result in good subjective image quality unless they are combined with more local detail. It is perhaps also not surprising that the local tone map exceeded all other conditions for the detail visibility rating, since this tone map prioritizes local detail. Ideally, however, a dichoptic tone map would completely, or nearly completely, preserve the visibility of local details. Some reduction in detail might be desirable so long as overall image quality was improved with the dichoptic tone map, however, we also found that the local tone map exceeded all other conditions in terms of overall preference (Fig. 3.4 A). This result suggests that the inclusion of the global tone map in the dichoptic condition did not increase users’ overall preferences. Ratings of dichoptically tone mapped images were not significantly different than the ratings for the average tone map.

Next, we turn to the exposure conditions (Fig. 3.4 B,D), for which we again did not find a consistent dichoptic preference. The median rating for the dichoptic exposure was descriptively higher than the median rating for the proper exposure, however this difference was small and not statistically significant. As compared to the two component images, the dichoptic exposure was rated significantly higher than the high exposure, but did not differ significantly from the low exposure. These results suggest an overall preference of the participants for underexposed images as compared to overexposed images, perhaps because the underexposed images offered a more appealing balance of light and dark points [57]. Again, we found that the ratings of the dichoptic condition did not systematically exceed the ratings for the most preferred nondichoptic component image (in this case, the low exposure).

The results from the 3D impression ratings in Experiment 3 differed notably from the other two experiments (Fig. 3.4 E,F). These results supported the observation that dichoptically tone mapped/exposed images create an enhanced 3D impression. For both the tone mapping and exposure conditions, participants consistently rated the dichoptic conditions to be more 3D than each of the nondichoptic conditions (these comparisons were all statistically significant). The 3D impression results stand in contrast to the results from the other two experiments: they suggest that dichoptic methods can generate a substantially different perceptual experience than nondichoptic methods. However, participants seem to attribute this experience to a difference in 3D information rather than an increase in visible details.

Table 3.3: Results of omnibus statistical tests of the ratings in Experiments 1, 2, and 3. For each dichoptic method and task, we report the  $\chi^2$  statistic and  $p$ -values from Friedman tests, along with Kendall’s coefficient of concordance for the participants ( $W$ ). Statistically significant comparisons with a  $p$ -value threshold of 0.05 are marked with an asterisk.

Tone Map			Exposure		
<b>Overall Preference</b>					
$\chi^2$	$p$	$W$	$\chi^2$	$p$	$W$
34.10	<0.001*	0.71	27.00	<0.001*	0.56
<b>Detail Visibility</b>					
$\chi^2$	$p$	$W$	$\chi^2$	$p$	$W$
34.91	<0.001*	0.73	37.81	<0.001*	0.79
<b>3D Impression</b>					
$\chi^2$	$p$	$W$	$\chi^2$	$p$	$W$
34.74	<0.001*	0.72	29.83	<0.001*	0.62

### Results across all conditions.

Recall that our baseline dichoptic exposure method did not incorporate strategies to reduce perceptual artifacts, such as binocular rivalry. Thus, it is also interesting to examine how well this exposure method compares to tone mapping methods optimized to reduce artifacts. In these experiments, the dichoptic exposure method was not rated significantly different for overall preference ( $z=-1.46$ ,  $p=0.14$ ) or 3D impression ( $z=0$ ,  $p=1$ ) as compared to the optimized dichoptic tone mapping method from Zhang et al. [182]. The exposure method was rated significantly higher for detail visibility ( $z=-2.27$ ,  $p=0.02$ ). These results suggest that a relatively naïve dichoptic method may be sufficient to obtain the gains in 3D impression, although it is key to consider how important these gains are for stereoscopic content (see Discussion).

Lastly, to ask whether there is a best tone mapping method based on these results combined, we also descriptively compared the rankings among all 8 conditions. For overall preference, the highest rated method was tied between the local tone map and the low exposure. For detail visibility, the highest rated method was tied between the local tone map, the low exposure, and the dichoptic exposure. For 3D impression, the highest rated method was tied between the dichoptic tone map and dichoptic exposure. These rankings highlight the fact that different methods may be more suitable for supporting different aspects of perceptual experience. While overall preference may be the best global criterion to use for many applications, it is interesting to consider whether it is useful to combine multiple criteria when picking a tone mapping method. As a first step towards a multi-dimensional approach, we also calculated an overall rating score for each method by simply summing the median ratings across the three experiments. Considering all three criteria in this way, the best method would be dichoptic exposure, with an overall median score of 11.75 out of 15 possible. Importantly, this approach gives equal weight to all three criteria, which may not be appropriate. However, in some cases it might be possible to select a set of criteria and weights for a specific application. For example, if a compromise on detail visibility were desirable in order to obtain an enhanced 3D impression, both dichoptic methods would likely be more desirable than any nondichoptic ones.

Table 3.4: Results of follow up statistical tests for the participant ratings in Experiments 1, 2, and 3. Wilcoxon signed-rank tests were run comparing the ratings of each dichoptic condition against its three nondichoptic comparisons, using z-statistics to assess significantly different ratings with a p-value threshold of 0.05. Each row indicates the results for comparing one dichoptic condition (tone map or exposure) with the listed nondichoptic conditions. We corrected for multiple comparisons using a false discovery rate (fdr) of 0.05 [18]. Comparisons for which the ratings were significantly different following fdr correction are marked with an asterisk.

Tone Map			Exposure		
<b>Overall Preference</b>					
	<i>z-Stat</i>	<i>p</i>		<i>z-Stat</i>	<i>p</i>
Average	-1.52	0.13	Proper	0.90	0.37
Global	-3.01	0.003*	High	-2.60	0.01*
Local	2.37	0.02*	Low	2.01	0.04
<b>Detail Visibility</b>					
	<i>z-Stat</i>	<i>p</i>		<i>z-Stat</i>	<i>p</i>
Average	-1.89	0.06	Proper	-1.32	0.19
Global	-3.21	0.001*	High	-3.53	< 0.001*
Local	2.12	0.03	Low	1.68	0.09
<b>3D Impression</b>					
	<i>z-Stat</i>	<i>p</i>		<i>z-Stat</i>	<i>p</i>
Average	-3.45	< 0.001*	Proper	-3.09	0.002*
Global	-3.44	< 0.001*	High	-3.33	< 0.001*
Local	-3.07	0.002*	Low	-2.97	0.003*

## Results summary

With respect to overall preference and detail visibility, our results suggest that both dichoptic methods were better on average than one of their nondichoptic component images (global tone map and high exposure), but not the other (local tone map and low exposure). These results suggest that dichoptic methods may not yet consistently yield improvements along these perceptual dimensions, which differs from the conclusions drawn from previous user studies [174, 181, 182, 183]. However, when 3D impression is considered as well, we found more evidence to support substantial subjective improvements with dichoptic tone mapping. The difference in our conclusions with respect to prior work might be explained by the decreased sensitivity of the rating task as compared to the 2AFC tasks used in previous studies. However, the current results also suggest another potential explanation: that the preferences for the dichoptic methods over some standards may be explained by people preferring one component image of the dichoptic pair, rather than the dichoptic pair *per se*.

### 3.4.3 Experiments 4 and 5: Preference and Subjective Detail Visibility Forced Choice Comparisons

For the next set of experiments, we focused in on comparing the higher rated nondichoptic conditions against the dichoptic conditions with a 2AFC task. This allowed us to test the hypothesis generated by the previous experiments with a targeted comparison. We only focus on preference and detail visibility, as the results for 3D impression were quite clear. For these experiments, we created four dichoptic versus nondichoptic 2AFC comparisons. These included the dichoptic conditions from the previous experiments and one new one. While earlier experiments were in progress, a new dichoptic tone mapping method was published by Zhong et al. [183], called DiCE, which we incorporated into

these next studies.

## Participants

Two different groups of sixteen adults participated in these studies. One group completed the preference task (10 F, 6 M, ages 18 to 23) and the other group completed the detail visibility task (11 F, 5 M, ages 18 to 28).

## Procedure

The paired conditions for comparison are shown in Table 3.5. For the Zhang et al. [182] dichoptic tone map and for the dichoptic exposure conditions, the nondichoptic standards were selected as the most preferred component images from Experiments 1 and 2. For the DiCE conditions, we included comparisons with both component images, because we did not have ratings data to justify choosing one over the other.

Table 3.5: Summary of conditions that are compared in the 2AFC tasks in Experiments 4 and 5.

<b>Dichoptic</b>	<b>vs.</b>	<b>nondichoptic</b>
dichoptic tone map [182]	vs.	local tone map
dichoptic exposure	vs.	low exposure
DiCE [183]	vs.	C1 (low-light detail)
DiCE [183]	vs.	C2 (highlight detail)

On each trial, participants viewed a dichoptic and a nondichoptic pair of the same scene in sequence, for 3 seconds each. They indicated whether they preferred the first or the second image in terms of overall image quality (Experiment 4) or detail visibility (Experiment 5), using the same basic instructions from Experiments 1 and 2. There were four repeats of each comparison, for a total of 288 trials. Before starting the experiments, participants viewed all stimuli once.

## Results.

For both experiments, we calculated the mean proportion of trials in which each participant chose dichoptic over nondichoptic images. The results are plotted in Fig. 3.5 and statistical tests are summarized in Table 3.6. In the figure, a preference for the dichoptic method is indicated by data points falling above the dashed line. Qualitatively, in both experiments there was no consistent preference for dichoptic viewing compared to nondichoptic viewing of the component images across all four conditions. In fact, the only consistent preferences were in favor of the nondichoptic component images. The results from the Zhang et al. [182] tone map and the exposure conditions support the rating results from Experiment 1 and 2, which suggested that participants prefer the better quality component image of the dichoptic pair over dichoptic viewing. The dichoptic pair in these comparisons was selected at a rate significantly lower than chance in both Experiment 4 and 5. Once again, this is perhaps not surprising for the detail visibility judgement on the images from Zhang et al., however, the fact that the same results were obtained for the overall preference judgement suggests that more work is needed to ensure that dichoptic tone maps boost subjective image quality beyond the components.

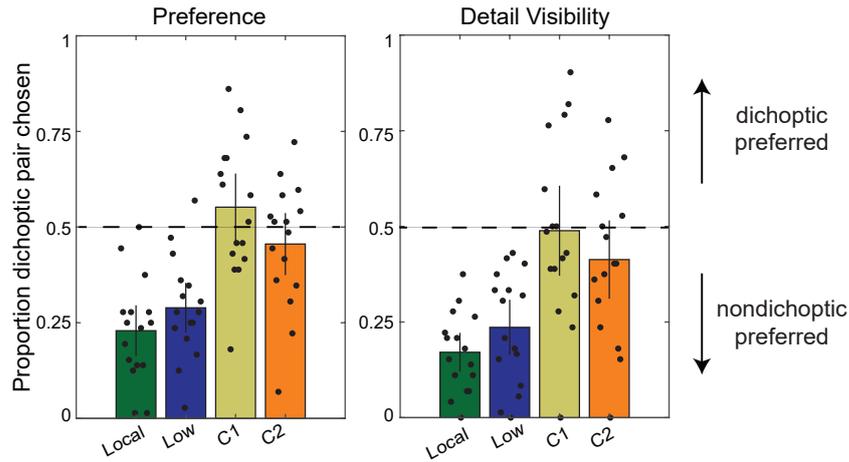


Figure 3.5: Results from 2AFC task in Experiments 4 (left) and 5 (right). Each black dot represents an individual participant’s responses, in terms of proportion of trials that the dichoptic condition was chosen over the nondichoptic component image (higher indicates that dichoptic is better). Dot locations are jittered for visibility. Each bar represents the overall mean proportion across all participants. Error bars indicate 95% confidence intervals. The black dotted horizontal line indicates chance level.

The results for the DiCE method are closer to chance on average and highly variable across participants. Statistical tests suggested that DiCE dichoptic images were chosen at levels that did not differ significantly from chance in both experiments. While a few participants had a consistent preference for DiCE over one or the other component, these effects were counterbalanced by other observers having a consistent preference against DiCE.

Table 3.6: Results of statistical tests for 2AFC task in Experiments 4 and 5. For each dichoptic to nondichoptic comparison, a two-tailed, single sample  $t$ -test was conducted, with the null hypothesis that the mean proportion was equal to chance (0.5) and a significance threshold of  $p < 0.05$ . Comparisons for which the mean was significantly different from 0.5 are marked with an asterisk.

	Preference				Detail Visibility			
	Mean	SD	$t$ -Stat	$p$	Mean	SD	$t$ -Stat	$p$
Local	0.23	0.13	-7.99	< 0.001*	0.17	0.10	-12.73	< 0.001*
Low	0.29	0.13	-6.35	< 0.001*	0.24	0.14	-7.17	< 0.001*
C1	0.55	0.17	1.16	0.26	0.49	0.23	-0.19	0.85
C2	0.46	0.16	-1.07	0.30	0.41	0.20	-1.66	0.12

In summary, the results of these 2AFC tasks (in concert with the ratings experiments) suggest that the study participants did not have a strong, consistent preference for dichoptically tone mapped images with respect to overall image quality or detail visibility as compared to the images tone mapped with conventional nondichoptic methods.

### Scene-based Analysis.

We were curious if different scenes might lend themselves more or less to dichoptic methods. For this exploratory analysis, we focused on the results obtained from the DiCE method because this approach produced a mixture of preferences that were greater than and less than chance across different participants (Fig. 3.5). We also focused on the detail visibility judgment, because this perceptual task is most closely related to contrast enhancement. When we analyzed the data separately for each scene, we found that different scenes produced consistently different preferences in the 2AFC task. Fig.

3.6A shows the proportion of trials on which the dichoptic pair was chosen for each scene when compared to the components (C1 and C2), averaged over all participants. Across different scenes, these averages varied from lower than chance to greater than chance. As shown in Fig. 3.6B, we observed a negative correlation between the responses when dichoptic pairs were compared to C1 and C2 ( $r = -0.94$ ). When the dichoptic pair was consistently chosen more than C1, it tended to be consistently chosen less than C2, and vice versa (upper left and lower right quadrants). No scenes were consistently preferred with the dichoptic method when compared to both components.

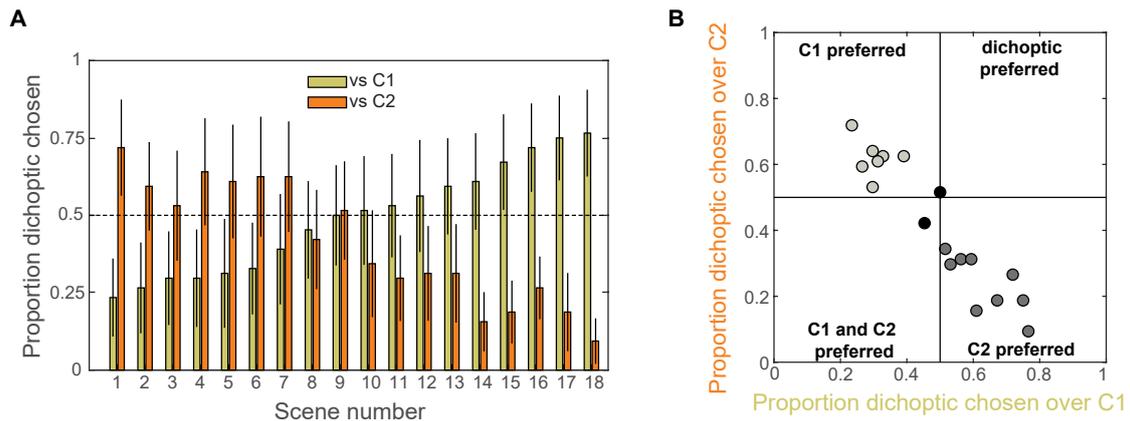


Figure 3.6: Results of exploratory scene-based analysis on Experiment 5. (A) We re-analyzed the results for the 2AFC local detail task on the DiCE-generated images separately for each scene. Each bar indicates the proportion of trials for which the dichoptic pair was chosen versus each component (C1, C2), averaged across participants. Error bars show the 95% confidence intervals. Data are sorted by the C1 comparisons. (B) Each point represents the dichoptic preference of an individual scene when compared to C1 (x-axis) and C2 (y-axis).

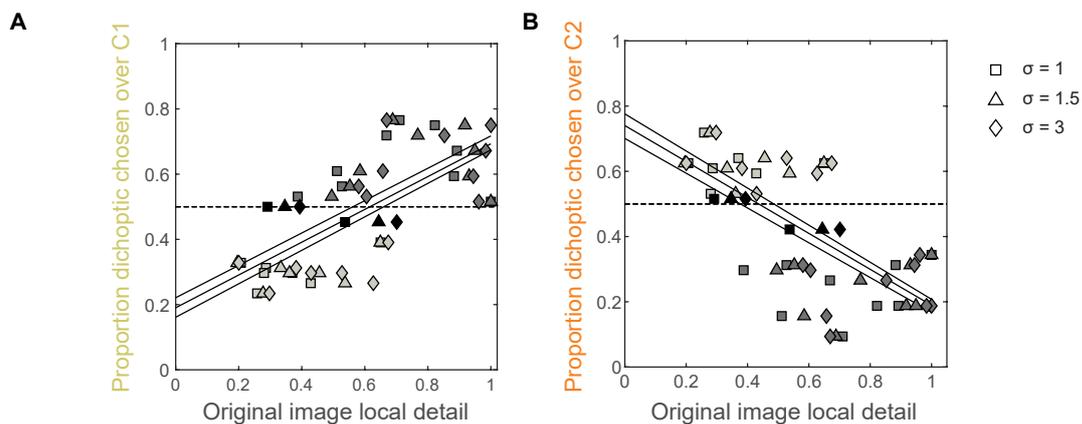


Figure 3.7: The relationship between local detail quantified using rotationally symmetric Laplacian of Gaussian filter (x-axis) and the proportion of time that the dichoptic pair was chosen over either C1 (A) or C2 (B) component images (y-axis). The filter standard deviations ( $\sigma$ ) were 1, 1.5, and 3 pixels, represented by different symbols. Black lines show best-fit linear regressions for each filter size. For the scatter plot, ranges were normalized for each filter size to allow plotting on the same axes.

We next asked whether this pattern could be explained by any low level features of the images. We found that the participant preferences were strongly correlated with the amount of local detail in the original images (Fig. 3.7, original image before applying DiCE tone maps). Local detail was quantified as the mean absolute response of a Laplacian of Gaussian filter, and this relationship was consistent over a range of filter sizes (see Figure

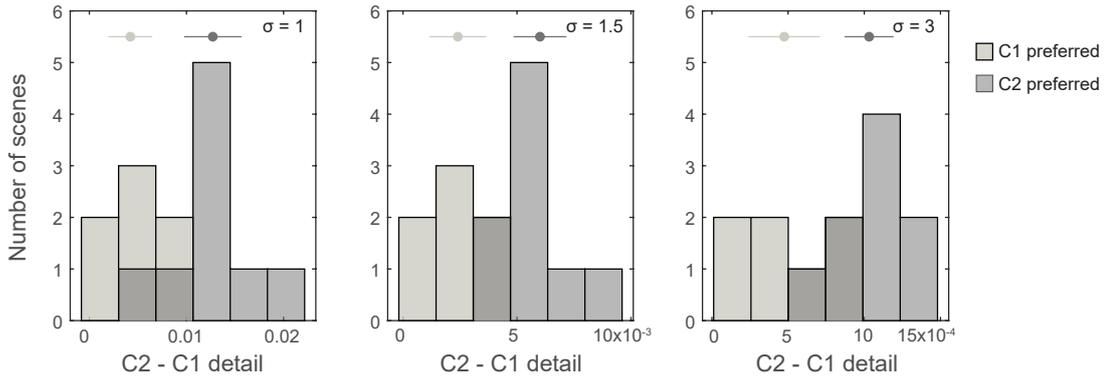


Figure 3.8: For each filter size, each plot shows a histogram of the difference between the C2 and C1 image local detail, separated out by C1-preferred and C2-preferred images. Above the bars, the mean and 95% confidence intervals are shown for the C1-preferred and C2-preferred scenes.

caption for details and Table 3.7 for statistical analysis). Higher levels of local detail in the original image were associated with a tendency for participants to prefer the dichoptic pair versus C1 (Fig. 3.7A), but also with a tendency for participants to prefer the C2 over the dichoptic pair (Fig. 3.7B).

Then, we categorized the scenes as either C1-preferred or C2-preferred. Two scenes without consistent preferences were excluded from this analysis. For each scene category, we calculated the difference in the local detail of the C2 and C1 images (Fig. 3.8). We found that C2-preferred scenes (dark gray marker/bar) tended to have higher detail in the C2 images than C1-preferred images (light gray marker/bar); C1-preferred scenes tended to have similar levels of detail between C1 and C2 (Table 3.7). These results suggest that low level image properties can reliably predict which nondichoptic component image would be preferred over the dichoptic pair, but they do not yet tell us why (see Discussion). This is the first analysis that we are aware of that examines image-based predictors for dichoptic tone mapping preferences based on user study data. Moving forward, image-based analyses of larger data sets that cover a diverse range of scenes may be able to help guide further algorithmic development for dichoptic tone mapping.

Table 3.7: Results of statistical tests for DiCE scene-based analysis shown in Fig. 3.6 Panels C & D. Left, Middle: For each filter size ( $\sigma$  in pixels), the correlation coefficient with 2AFC responses ( $r$ ), 95% confidence intervals (CI), and  $p$  values are reported. Significant correlations ( $p \leq 0.05$ ) are marked with an asterisk. Right: For each filter size, a two-tailed, unpaired  $t$ -test was conducted to compare C1-preferred scenes and C2-preferred scenes, with the null hypothesis that the mean local detail difference between the two scene categories were not different from each other and a significance threshold of  $p \leq 0.05$ . Comparisons for which the means were significantly different between C1- and C2-preferred scenes are marked with an asterisk.

DiCE vs C1			DiCE vs C2			C2 vs C1 preferred		
$\sigma$	$r$ [95% CI]	$p$	$\sigma$	$r$ [95% CI]	$p$	$\sigma$	$t$	$p$
1	0.70 [0.35 0.88]	0.001*	1	-0.67 [-0.87 -0.29]	0.002*	1	-4.26	0.001*
1.5	0.71 [0.36 0.88]	0.001*	1.5	-0.69 [-0.87 -0.33]	0.002*	1.5	-4.13	0.001*
3	0.71 [0.36 0.88]	0.001*	3	-0.70 [-0.88 -0.35]	0.001*	3	-3.97	0.001*

### 3.4.4 Experiment 6: Objective Detail Visibility

The potential use cases of VR are varied, and extend beyond situations in which simply enhancing subjective image quality is the goal. For example, if dichoptic methods can

objectively increase visible detail in a scene (without subjective improvement), they may be quite useful for applications such as remote guidance.

Objective perception of visibility in natural images is challenging to predict, and there exists no standard technique for characterizing how detail visibility varies across a natural image. For example, it was recently shown that detection of small targets embedded in natural images depends on several scene properties, such as luminance, contrast, and pattern similarity [143]. We thus designed a novel exploratory task that we predicted would be easier to perform if details were more visible in natural images. Specifically, we showed people images and cued them to look at a specific region. We then presented them with two probes, which were taken from a small patch in that region. One probe was consistent with the original scene, and the other probe was mirror-reversed about a vertical axis. We asked participants to respond which one matched the original scene. Our reasoning was that if a particular condition resulted in better perceived visual detail, then participants should be able to better discriminate the original patch from a mirrored one. Note that this reasoning is only valid if the patches have at least some visual detail within them, and if they are not perfectly left/right mirror symmetric.

### Participants.

Twenty-four adults participated (19 F, 5 M, ages 18 to 34).

### Stimuli.

In this experiment, we used the same eight tone mapping and exposure conditions as in Experiments 1-3. Usually highlight and low-light areas are the most difficult to reproduce on a conventional display, so we focused on these areas for our stimuli. We defined highlights and low-lights broadly as the top and bottom 15% of pixel values according to their intensity in grayscale in the original HDR images (illustrated in Fig. 3.9). We then segmented these images into 35 by 35 pixel patches and selected eight highlight-dominant and eight low-light-dominant patches from each scene. Highlight and low-light patches were defined as patches in which at least 75% of the content was made up by high- or low-light pixels. In cases where a scene contained more than 16 patches that met these criteria, we manually chose a subset with minimal mirror symmetry. The bottom panel in Fig. 3.9 shows examples of what these patches looked like when each of the tone maps used in this experiment were applied to the HDR image.

For each patch, we next created a *probe* version to test performance on the orientation task. This probe was created with a custom, purely localized tone map applied to the HDR values within the specific intensity range of each patch, using Matlab's *imadjust* function (Fig. 3.9, right column). This approach provided a way to test performance with the same probe that differed from all the specific tone maps being assessed. Visualizations of all patches and probes are available upon request.

*Procedure.* The procedure for one trial is illustrated in Fig 3.10. The participant was first presented with a red box cuing a target area on the screen ( $\sim 1$  deg visual angle) for 1 second. Then the cue disappeared and an image (dichoptic or nondichoptic) from a particular scene was shown for 4 seconds. The participant was asked to look at the cued location in the scene and study that area very well. Then the scene disappeared and a fixation cross was shown for 1 second at the center of the screen. After the fixation cross

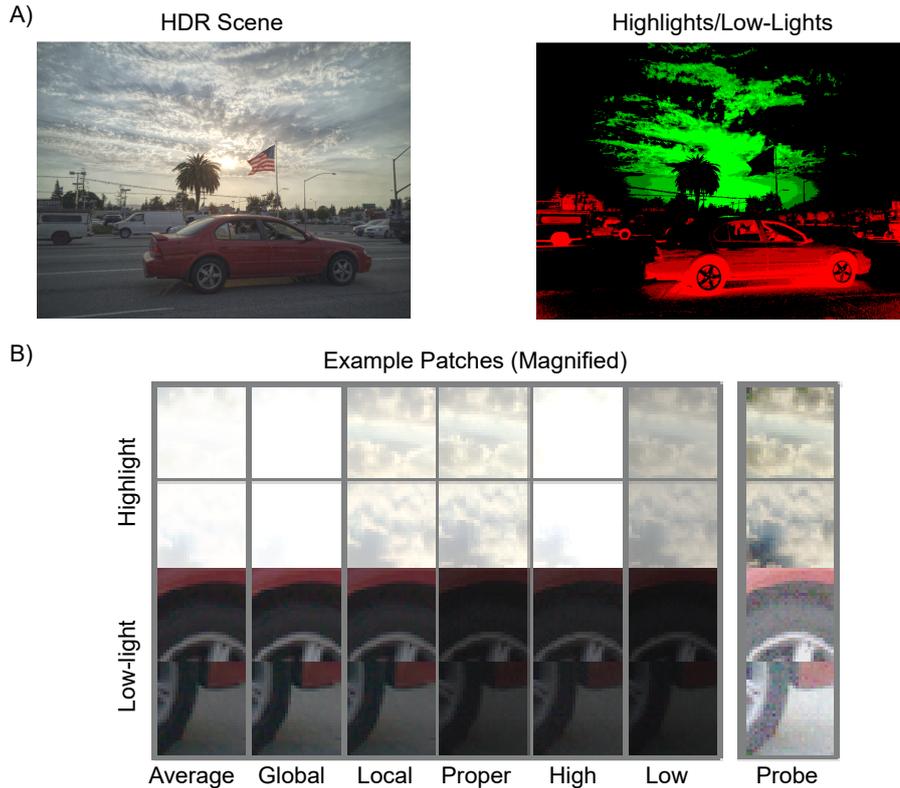


Figure 3.9: A) An example HDR image ([64]) and its intensity map with highlight and low-light areas color coded in green (e.g., sky) and red (e.g., the car) respectively. Color saturation reflect the strength of the high/low-light. B) Examples of test patches for two highlight and two low-light areas are shown below. The first 6 columns show how the patches are rendered in the original scene under different conditions. The last column shows the probe patch used, for which a local tone map has been applied.

disappeared, two probe patches were shown consecutively for 2.5 seconds each. One probe was always in the original orientation and one probe was mirror reversed, in random order. The participants were told that the probe patches could have different contrast, brightness and color compared to what they saw in the original full scene, so they were instructed to focus on visual patterns within the location. The patches were shown on a mean luminance background against a screen of mid-grey ( $83 \text{ cd/m}^2$ ). Due to a slight offset in the alignment of some images obtained from Zhang et al. [182], which was discovered after completing data collection, the cued location was slightly lower than the selected patch in some of these conditions. This offset was small (about 5 pixels) and is not expected to influence the results.

Following a Latin Square design, each participant saw a predetermined set of scene and condition combinations, where each scene was only viewed in one condition. This design prevents participants from remembering the scene content that may be visible in one condition but not another, while still randomizing the assignment of scenes to conditions across all participants. We used the same eight conditions as in Experiments 1-3. Specifically, two dichoptic methods (Zhang et al. dichoptic tone mapping [182] and dichoptic exposure) were included, along with nondichoptic versions of the component images, and two additional nondichoptic standards (average tone map and proper exposure). Each particular scene/condition combination was seen by three different participants. For example, Participants 1, 9, and 17 all saw Scene 1 in the average tone map condition only, and Scene 2 in the global tone map condition only, etc. For each scene, there were eight

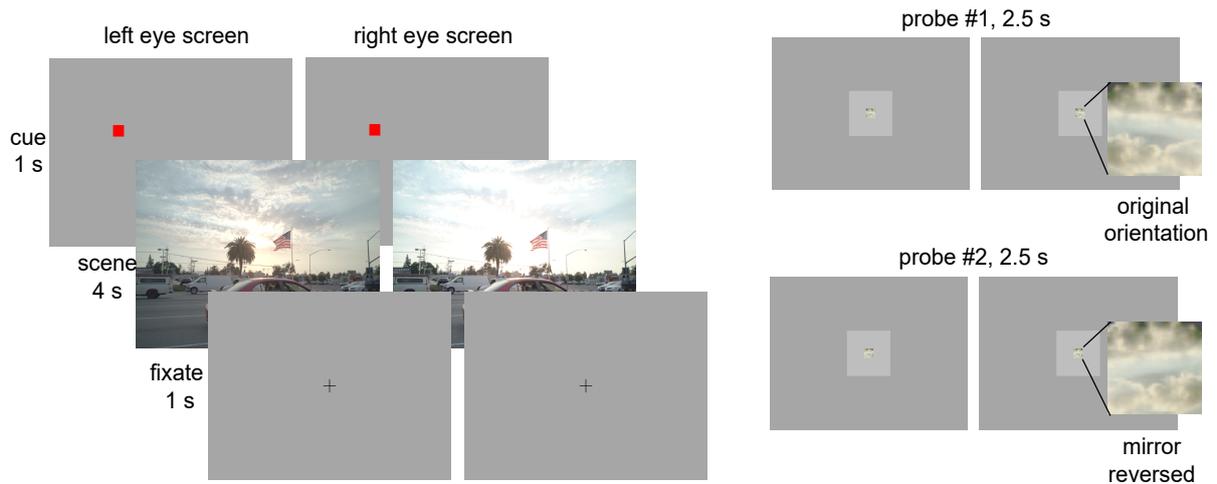


Figure 3.10: A schematic of a trial in the objective detail visibility task in Experiment 6. Participant goes through the cuing stage, the scene, and fixation (left) and then were presented with two probes (right). Magnified views of each probe are also shown. In this example, probe 1 matches the orientation in the original scene and probe 2 is mirrored.

low-light and eight highlight trials, creating a total of 128 trials.

## Results.

Fig. 3.11 shows the performance on the orientation discrimination task calculated over all participants in terms of the proportion of correct responses. Performance was above chance in all conditions, suggesting that the task was challenging, but not impossible. Qualitatively, we did not observe any substantial differences across conditions.

To examine potential statistical differences between conditions, we fit the trial-by-trial data using a logistic regression model. First, we focus on the results for the tone mapping conditions (Fig. 3.11, left panel). We modeled the four viewing conditions as categorical predictors and modeled both participant and scene as random effects. As in the analysis of the subjective data, we focus our analysis on differences between the dichoptic condition and the other three conditions. In the upper panel of Table 3.8, we report the coefficient estimates, odds ratios, and p-values associated with the model. Odds ratios were determined by exponentiating the coefficient estimates, and can be interpreted as the increase or decrease in the probability of a correct response associated with these nondichoptic conditions, as compared to the dichoptic condition. Values less than 1 indicate that a given condition resulted in fewer correct responses than the dichoptic condition, and visa versa. The p-values show that there were no statistically significant differences with the dichoptic condition. However, the performance when viewing the global tone map approaches being significantly worse than performance when viewing the dichoptic tone map. This makes sense, given that the global tone mapping algorithm is optimized for preserving overall global, rather than local, contrast [45].

The results for the different exposure conditions (Fig. 3.11, right panel) were analyzed in the same manner (Table 3.8, lower panel). We did not observe any statistically significant differences in performance between the dichoptic exposure condition and any of the nondichoptic conditions.

The premise for this experiment is that the performance on the patch orientation

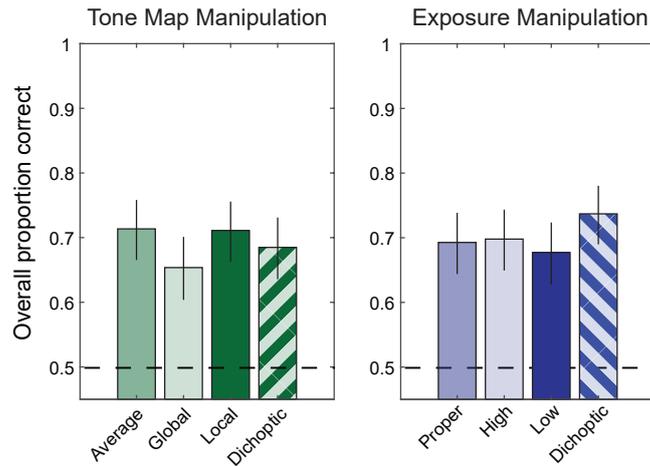


Figure 3.11: Results of the objective detail visibility task in Experiment 6. Each bar represents the proportion correct across all participants for each condition (labeled on the abscissa). Individual data are omitted because each participant only saw a subset of scenes in each condition. Error bars indicate 95% confidence intervals calculated from the binomial distribution. The dashed line indicates chance performance.

judgement task should be modulated by perceived contrast. Specifically, higher perceived contrast should lead to better performance because observers are better able to distinguish between the original and mirrored patches. In the limit, the task should be impossible if a patch is perceptually uniform. However, it is important to evaluate this premise on the specific stimulus set used. With this goal, we ran a post hoc analysis on the data. The details of the analysis are reported in the Supplementary Materials. The results suggest that, while contrast was correlated with performance on this task, the relationship was highly nonlinear. Specifically, higher contrast was associated with better performance for low contrast patches, but above a certain threshold this relationship was no longer reliable. At these supra-threshold levels, other factors of the stimulus, such as the horizontal mirror symmetry or the distinctiveness from the surrounding area, likely contribute to variability in performance. A definitive answer to the question of whether dichoptic tone mapping increases objective detail visibility will likely require a more challenging task that is consistently modulated by stimulus contrast across a broad range.

Table 3.8: Results of logistic regression model fit to data from Experiment 6. For each fixed effect (conditions), coefficients and odds ratios reflect the change in probability of correct responses associated with each nondichoptic condition relative to its dichoptic comparison. No coefficients were statistically significant at the threshold of  $p < 0.05$ . Contributions of scene and participant were modeled as random effects. There was moderate variability among different scenes and high variability among participants as shown by the standard deviations, which indicates how much each participant or scene deviates from the estimated average effects of the conditions.

Tone Map Conditions			
Condition	Coefficient Estimate	Odds Ratio	$p$
Average	0.114	1.30	0.25
Global	-0.18	0.66	0.06
Local	0.10	1.26	0.32
Random Effects	Standard Deviation	Lower Bound	Upper Bound
Scene	0.23	0.11	0.51
Participant	0.51	0.35	0.75

Exposure Conditions			
Condition	Coefficient Estimate	Odds Ratio	$p$
Proper	-0.04	0.91	0.68
High	-0.12	0.76	0.22
Low	-0.02	0.96	0.88
Random Effects	Standard Deviation	Lower Bound	Upper Bound
Scene	0.35	0.19	0.65
Participant	0.52	0.35	0.76

## 3.5 Discussion

### 3.5.1 Relationship to Prior Work

The current experiments did not include all published dichoptic methods (e.g., [51, 174, 181]) and should be interpreted in combination with the results from other studies [174, 181, 182, 183]. The efficacy of a dichoptic method is always judged relative to the standard used for comparison. Our study used nondichoptic standards that are different from most of the prior work since we were interested in how well dichoptic methods perform compared to their component images. The most similar data from prior work were obtained in the study by Zhong et al., in which the authors assessed Zhang et al.’s dichoptic method against the local tone map [183]. Our result is consistent with their finding that the dichoptic tone map was not chosen over the local tone map for overall preference. However, contrary to their findings, our data also suggest that the dichoptic tone map is not consistently preferred when judging detail visibility (in their user study, the dichoptic tone map was preferred over the local tone map in terms of contrast). While related, prompts to judge based on detail or contrast may be interpreted differently by participants, which may contribute to these different results. The other work that compared a dichoptic tone map against one component image is by Yang et al., but we did not include their tone mapping approach in our studies [174]. They found that the dichoptic method was consistently preferred based on “visual richness and content clarity”. Based on the pattern of results in our study, it is possible that subjective assessments of

visual richness may incorporate effects of dichoptic tone mapping on 3D impression in addition to perceived contrast and detail.

In addition to different nondichoptic standards and response prompts, there are a range of other experimental factors such as image sets and stimulus presentation time that differ across published studies and could affect the judgements of participants. In practice, the choice to adopt a dichoptic tone mapping algorithm may depend on balancing the range of desired perceptual effects and the tolerance for introducing artifacts such as rivalry. Taken together, these results emphasize that the choice of the baseline method for comparison and the perceptual criterion may influence the conclusions drawn from user studies. An assessment including both component images can provide a more complete picture for judging a dichoptic methods' performance. For example, the results from Experiment 3 show a compelling case in which the percept of the dichoptic pair exceeds the quality of both components individually.

### 3.5.2 Binocular Contrast Perception in Natural Images

Given the psychophysical work on binocular contrast combination, why might our study suggest that dichoptic tone mapping methods perform inconsistently? Psychophysical experiments on binocular contrast combination, which support a nearly winner-take-all model, tend to use simplified visual stimuli, such as sine wave gratings [103]. In natural imagery, it is possible that interactions between different spatial frequencies and other contextual information has an effect on binocular perceived contrast. For example, a recent study reported that binocular contrast perception can be dominated by the eye that sees the lower contrast stimulus (loser-take-all) when accompanied by local boundary information similar to what might occur at object boundaries in natural images [61]. This result suggests that context can strongly modulate binocular contrast perception.

Images of natural scenes also vary in terms of their local spatiotemporal properties. It is thus important to consider whether user study results could be specific to the selected images. For example, scenes that lack a high dynamic range may not require more than one tone map to capture visible contrast. While we do not have an absolute measure of luminance levels in the dataset that we used [64], the scenes were chosen to be diverse in content and dynamic range, with outdoor scenes that include both sky and shadows [171]. A follow up analysis of the variability of responses across scenes suggests that the overall conclusions of Experiments 4 and 5 are consistent even with smaller image samples (see Supplementary Material). Furthermore, our analysis based on image-computed local detail suggests that a large amount of variability across images is predictable: much of the variability in our data is accounted for by the local detail of the original image. The two DiCE tone mapping curves differentially allocate emphasis on local detail across different portions of the image histogram (low-lights or highlights). Thus, the current results support the idea that simple image-computable measures of visual contrast can be predictive of perceptual judgements of natural images. However, the image measures that need to be optimized for dichoptic imagery are still not well understood.

Moving forward, it would be interesting to examine the rules of contrast combination in stimuli of intermediate complexity (for example, small patches of natural images), in which the effects of image content and context can be studied in more detail (for example, see [42]). A better understanding of how image features relate to perceived contrast might yield insights for using these features to effectively modulate a dichoptic tone mapping

operator. For example, people may select particular salient features in a scene to make their decision about image quality. People may also tend to place different levels of emphasis on details in low-lights or highlights. Details in the shadows may be considered less important for taking in the content of a scene. Studying dichoptic tone mapping of smaller patches of natural images may help clarify some of these issues by reducing the number of different features in each unique stimulus.

### 3.5.3 3D Imagery

The stimuli used in our experiments were not stereoscopic. In this regard, our stimuli were not representative of typical imagery viewed on VR and stereoscopic displays, in which binocular disparities are leveraged to create a richer 3D percept. This choice was made to match the methods of prior work as closely as possible, but it also poses a potentially important limitation for inferring the effects of dichoptic tone mapping in VR. We can, however, draw some inferences from prior work about how binocular disparities might interact with dichoptic tone mapping in natural imagery. To our knowledge, the only dichoptic tone mapping study that included stereoscopic imagery was the study by Zhong et al. [183]. Their assessment of perceived contrast and 3D impressions in stereoscopic VR was consistent with their findings with non-stereo content. From the vision science literature, simple gratings have been used to study how binocular offsets (i.e., phase shifts between gratings shown to the two eyes) affect binocular contrast perception [39]. According to this work, adding a phase shift between gratings presented to the two eyes does not substantially change binocular contrast perception unless stimuli have low contrast and short presentation times (around 100 ms).

Based on these previous studies, it seems unlikely that the current results of our study for detail visibility would be qualitatively different if stereoscopic images were used. However, this assumption should be tested. Interestingly, binocular contrast differences between gratings do have a robust effect on binocular phase perception (i.e., the perceived phase of the binocularly viewing sine wave stimulus) [43]. This finding suggests that dichoptic tone mapping could have complex effects on perceived 3D shape, location and layout. A characterization of how dichoptic contrast differences interact with stereo depth in natural images is a valuable topic for future research.

### 3.5.4 Objective Visual Performance-Based Metric

Lastly, to probe the impact of dichoptic methods on perceived detail, we introduced an exploratory performance-based task. However, this task did not reveal substantial differences in perceived detail between the dichoptic conditions and the nondichoptic conditions. It is important to consider that this negative result could be due to the task itself. While we were attempting to design a simple task that could objectively assess detail visibility, it is worth noting that people were likely able to take into account the surrounding context of the patches to aid in their judgement. Indeed, all participants reported using some kind of continuation of lines or gradient of light and dark to make their judgement. Objective assessment has strong value in helping us better understand how people perceive visual content and potentially interact with it. However, particularly with natural images, it can be challenging to isolate the information of interest that people use to complete a task. Future work could explore developing a standard contrast perception task for natural scenes with a larger sample of users and image content, perhaps by

embedding standardized targets as in Sebastian et al. [143]. Furthermore, since dichoptic tone mapping methods are readily implementable in near-eye stereoscopic displays, it would be useful to further explore performance-based measures of tasks relevant for VR applications that incorporate dynamic content and stereoscopic viewing.

## 3.6 Conclusion

Despite many advances in imaging devices, display technology, and computational algorithms, digital scene reproduction remains a challenging process. With the rise of more commonplace stereoscopic displays for VR applications, it is appealing to consider how binocular vision can be leveraged to improve or augment existing reproduction pipelines beyond conventional stereoscopic imaging. Future work on developing dichoptic methods may benefit from examining what differs between scenes that elicit more or less improvement in subjective image quality and contrast, as well as focusing on gaining a better understanding of how these methods affect and enhance 3D impressions.

## 3.7 Supplemental Material

### 3.7.1 Binocular Rivalry and Eye Dominance (Experiments 1, 2, & 3)

When presented with different stimuli to the two eyes, the visual system needs to reconcile these stimuli to form a single percept. For dichoptic tone mapping to work as intended, the visual system should merge the two eyes' images in a way that maximizes information, regardless of channel that the information came in (e.g., left or right eye). However, people can have a perceptually dominant eye, such that the image shown to that eye is favored or contributes more to their percept. We wanted to know whether there were any differences in people's subjective ratings of dichoptic conditions that can be attributed to eye dominance. To do so, we conducted a *sensory eye dominance* test based on binocular rivalry for each of the participants in Experiments 1-3. We performed post hoc analyses to ask whether the participants' eye dominance status might be predictive of their rating results.

#### Stimuli and Procedure.

Two orthogonal sine wave gratings were presented to the eyes simultaneously via the haploscope. This is a standard stimulus for binocular rivalry tests [60]. Participants were instructed to press down buttons continuously to indicate whether they saw the grating predominately tilted top left or top right, and to not press any key if they saw an equal mixture of both orientations. The procedure took about 3 minutes. Participants had a short 1 minute practice before the actual test.

#### Results.

The dominant eye of each participant was determined by calculating the proportion of time that the participant reported seeing the grating presented to each eye. We also calculated the proportion of time each participant saw a mixture of both eye's stimuli.

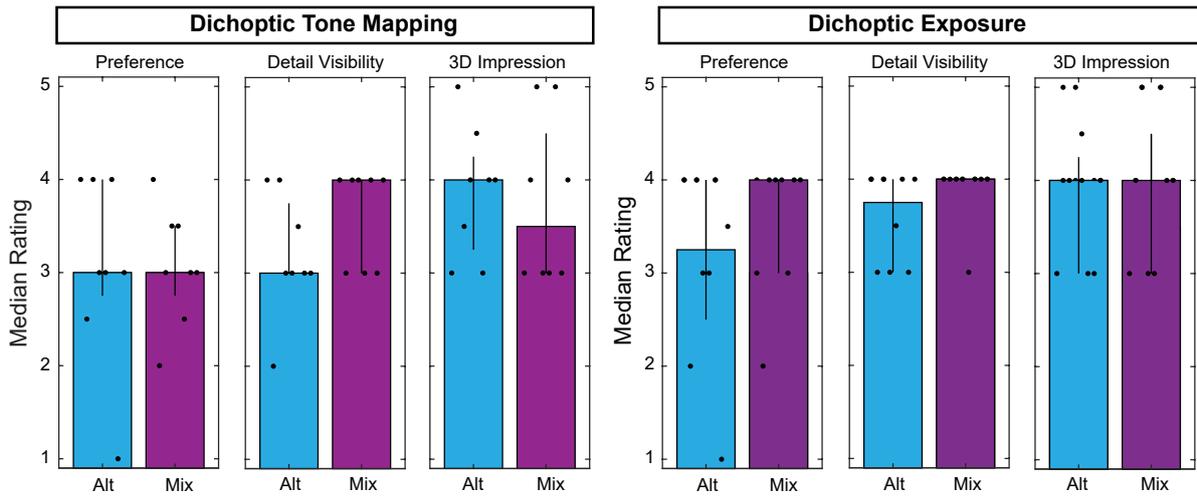


Figure 3.12: Results of post hoc analysis of the relationship between eye dominance and median ratings. For each panel, the dichoptic condition is split up between participants who were defined as ‘Alternators’ (Alt) and ‘Mixers’ (Mix). The top row shows the results for the three ratings tasks for the dichoptic tone mapping condition, the bottom row shows the same results for the dichoptic exposure condition. The black dots indicate individual participants, the bars indicate the median across participants, and the black lines indicate the 25<sup>th</sup> and 75<sup>th</sup> percentiles.

We categorized each participant as a ‘Mixer’ if their proportion of time seeing a mixture was greater than the median across all participants (52 seconds), or an ‘Alternator’ if their proportion of time seeing a mixture was less than the median. The rating results for the three experiments are replotted in Fig. 3.12, with data shown separately for Alternators and Mixers. There were no statistically significant differences between Alternator’s and Mixer’s ratings on dichoptic viewing conditions across all experiments (Table 3.9, upper panel). Descriptively, Mixers tended to rate the dichoptic tone map condition higher for detail visibility, and the dichoptic exposure condition higher for preference, but these differences were not statistically significant.

Since the dichoptic trials were repeated with the component images switched between the two eyes, we also explored if there was a difference in whether the higher-rated component image in the dichoptic pair was seen by the dominant eye or the non-dominant eye. User ratings tended to be similar but not identical between these two repeats (average Spearman correlation of 0.66, 0.70, and 0.58 for Experiments 1, 2, and 3 respectively). Some of this variability may be due to inconsistent preferences, but some may be due to the different images seen by the dominant eye. The results of this analysis are plotted in Fig. 3.13 and statistical comparisons are reported in Table 3.9 (lower panel). We found no consistent or statistically significant improvements when the higher-rated component image was seen by the dominant eye.

In summary, we did not find evidence for a relationship between the person’s sensory eye dominance and their responses in our user studies. This may indicate that the perceptual outcome is driven more by the content of the stimuli (i.e., the better component image) than by which eye sees which stimulus. This result is in line with results from Yang et al., in which they switched the presentation of the dichoptic pair for each eye, and found no difference in 2AFC responses [174]. The DiCE study also included a supplementary study on eye dominance in which they tested *sighting* eye dominance, and found no systematic relationship with contrast enhancement [183]. It is important to note that the existing

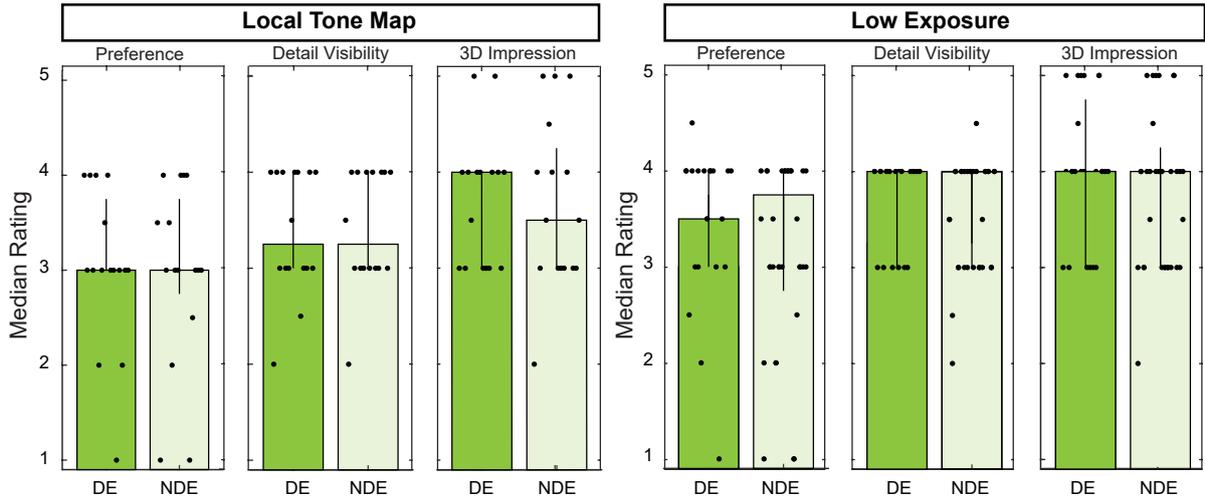


Figure 3.13: Results of post hoc analysis of the which eye saw the higher rated component image (i.e., the local tone map or the low exposure) during dichoptic conditions. For each panel, the median ratings of a dichoptic condition are shown, split into trials for which the higher rated component image was seen by dominant eye (DE) or non-dominant eye (NDE). The top row shows the results for the dichoptic tone map condition, the bottom row shows the same results for the dichoptic exposure condition.

Table 3.9: Results of Wilcoxon statistical tests examining differences between the responses of participants who were Alternators and Mixers (top) and differences between responses when the preferred component image was presented to the dominant versus non-dominant eye (bottom). We used z-statistics to assess significantly different ratings with a p-value threshold of 0.05. Each row indicates the results for a particular perceptual quality. For unpaired samples (top), the rank sum test was used. For paired samples (bottom), the signed-rank test was used.

Tone Map			Exposure		
Alternator vs. Mixer Rating					
	<i>z-Stat</i>	<i>p</i>		<i>z-Stat</i>	<i>p</i>
Preference	0.22	0.83	Preference	-0.85	0.40
Detail	-1.32	0.19	Detail	-1.42	0.16
3D	0.38	0.70	3D	-0.23	0.82
Dominant vs. Non-Dominant Eye Rating					
	<i>z-Stat</i>	<i>p</i>		<i>z-Stat</i>	<i>p</i>
Preference	0.00	1.00	Preference	0.28	0.78
Detail	-0.28	0.78	Detail	-0.52	0.60
3D	0.32	0.75	3D	0.38	0.71

analyses of eye dominance effects, including our own, have used relatively small sample sizes for analyses of individual differences. As such, there may still be robust individual differences in binocular combination that contribute to the variability of preferences for dichoptic methods, which may be explored in future work.

### 3.7.2 Scene Sample Analysis (Experiments 4 & 5)

It is important to consider whether our user study results could be specific to the selected scenes, rather than a generalizable observation about dichoptic tone mapping. This is a challenging question to answer, but we can gain some insight by examining the variability within the data we have. To examine the relationship between the number of natural images tested and the conclusions drawn from the user studies, we performed a post hoc resampling analysis. We asked how variable the results for Experiments 4 and

5 would be for a range of smaller image samples from 3 images to 15 images. Our logic was that, for a given sample size, a large amount of variability suggests that conclusions from a given study would be less generalizable. Thus, we repeatedly sampled our user response data from a subset of the original 18 scenes randomly and recalculated the average proportion dichoptic chosen. For example, for a simulation using 3 scenes, we might randomly chose Scenes 1,2,3 in the first simulation, but Scene 2,3,10 might be chosen for the next simulation. We performed each sub-sampling (3, 6, 9, 12, and 15 scenes) for 100 simulations to obtain the average and standard deviation. The results are shown in Fig. 3.14. As expected, lower numbers of scenes result in more variability across simulations. For the local tone map and low exposure comparison conditions, the bulk of the results showed the nondichoptic images being consistently preferred even for very small sample sizes. For the DiCE comparison conditions, small sample sizes could result in consistent above or below chance preferences for the dichoptic tone maps. But once at least 10 images were included, the range of results was highly consistent with the results with the full 18 image set. While this technique only allows us to sample the variability within the current images, it provides some support to the notion that the conclusions are not specific to particular images.

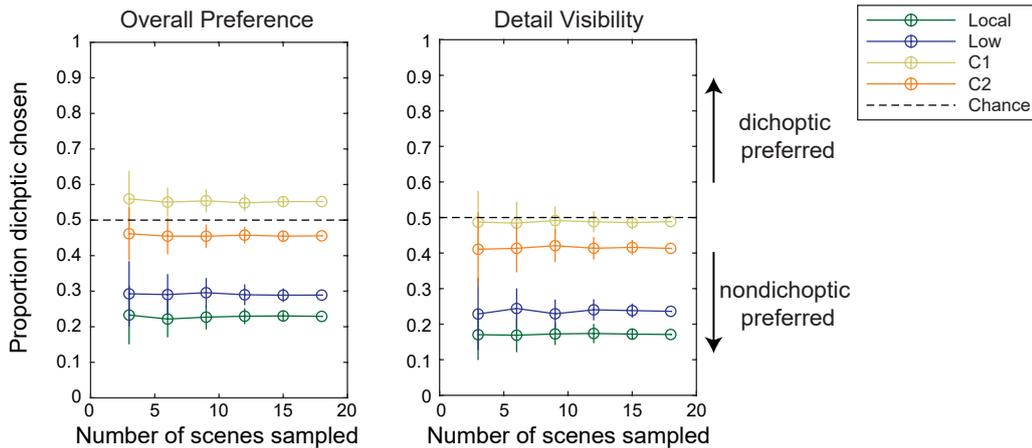


Figure 3.14: Results for simulating Experiment 4 and 5 using the data from a subset of scenes. Each line represents one condition’s mean and vertical bars indicate +/- one standard deviation. The x-axis indicates how many scenes were randomly sampled, equally spaced from 3 to the actual number of scenes that we tested, which was 18.

### 3.7.3 Effect of Contrast on Objective Task Performance (Experiment 6)

We conducted a post hoc analysis of the results from Experiment 6 to examine whether patch contrast was predictive of task performance. This analysis focused on the trials with nondichoptic presentation, because on these trials we have a reasonable model for perceived contrast that does not rely on assumptions about binocular combination. For each unique stimulus presented with a nondichoptic tone map ( $n = 768$ ), we computed the normalized contrast of the patch by dividing the standard deviation of pixel gray-scale values by the mean gray-scale value. This approach calculates contrast in units that are approximately proportionate to the overall brightness of the patch, reflecting Weber’s Law for contrast detection [129]. Due to the limited number of user responses available to compute accuracy for each individual stimulus (in the Latin Square design, a given

patch and tone map combination was only seen by three users), we used a sliding window (of size  $n/10$ ) to compute the average proportion correct responses as a function the normalized contrast. The results are plotted in Fig. 3.15. Note that the contrast values on the abscissa are plotted on a log scale. At very low contrasts, the users' performance was positively related to contrast, with average performance increasing rapidly by  $\sim 10\%$  as contrast increased above the minimum. However, above a certain normalized contrast level ( $\approx 0.002$ ), the performance no longer increased systematically with contrast. These results suggest that visible contrast is important for enabling performance on the task, but that above a relatively low threshold the contrast becomes sufficient and no longer limits performance.

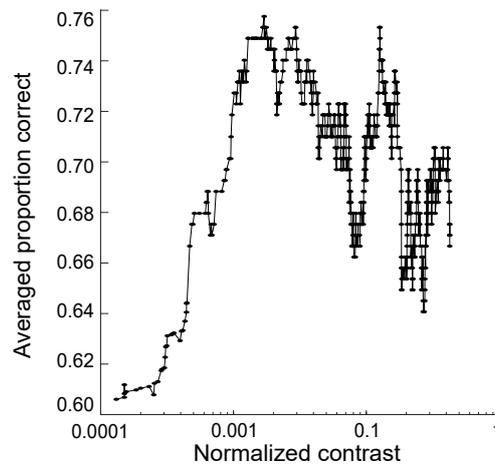


Figure 3.15: Analysis of the effect of stimulus contrast on the objective task performance in Experiment 6. The x-axis is the normalized contrast on a log scale and the y-axis is the average performance. Data were smoothed with a moving average filter with a span of 77 samples. Edge cases that exceeded the half-width of the filter were cropped.

# CHAPTER 4

## Perceptual guidelines for optimizing field of view in stereoscopic augmented reality displays

### 4.1 Introduction

Near-eye display systems for augmented reality (AR) aim to seamlessly merge virtual content with the user's view of the real world. A substantial limitation of current systems is that they only present virtual content over a limited portion of the user's natural field of view (FOV). This limitation reduces the immersion and utility of these systems. Thus, it is essential to quantify FOV coverage in AR systems and understand how to maximize it. It is straightforward to determine the FOV coverage for monocular AR systems based on the system architecture. However, stereoscopic AR systems that present 3D virtual content create a more complicated scenario because the two eyes' views do not always completely overlap. The introduction of partial binocular overlap in stereoscopic systems can potentially expand the perceived horizontal FOV coverage, but it can also introduce perceptual nonuniformity artifacts. In this paper, we first review the principles of binocular FOV overlap for natural vision and for stereoscopic display systems. We report the results of a set of perceptual studies that examine how different amounts and types of horizontal binocular overlap in stereoscopic AR systems influence the perception of nonuniformity across the FOV. We then describe how to quantify the horizontal FOV in stereoscopic AR when taking 3D content into account. We show that all stereoscopic AR systems result in a variable horizontal FOV coverage and variable amounts of binocular overlap depending on fixation distance. Taken together, these results provide a framework for optimizing perceived FOV coverage and minimizing perceptual artifacts in stereoscopic AR systems for different use cases.

Digital displays have become essential tools for education, work, healthcare, and entertainment. While conventional displays present imagery on an opaque panel, emerging augmented reality (AR) display systems aim to create mixtures of real and virtual content that are visually immersive. These AR systems often rely on stereoscopic near-eye displays that optically combine the user's natural vision with 3D virtual content directed to the viewer's eyes from a pair of micro-displays or other light sources (Figure 4.1).

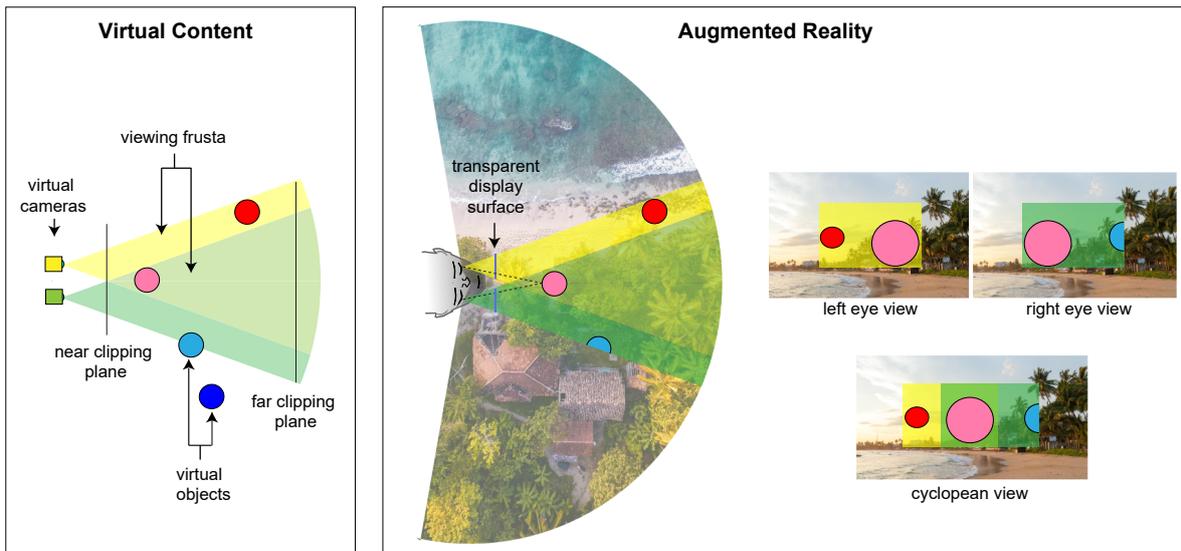


Figure 4.1: Illustration of the virtual field of view (FOV) and the natural FOV in augmented reality systems. Virtual cameras capture virtual content within their viewing frusta (left), and the resulting images are presented on a stereoscopic display that merges this content with a portion of each eye’s natural FOV (right). When the eyes’ images are fused, the “cyclopean view” can capture a greater horizontal extent than either monocular FOV, with some regions seen binocularly, and others seen monocularly. Image credits: Unsplash (Oliver Sjöström, Rowan Heuvel)

Many factors can influence the immersive nature of AR experiences. For example, physical realism can be limited by the resolution, contrast, and depth information provided by an AR system [97, 179]. Optical elements can interfere with display visibility and lead to distracting visual artifacts as users look around a scene [26, 31]. Importantly, the limited coverage of the user’s natural field of view (FOV) by the display can impair the immersive experience [108] and can also affect performance on a variety of tasks [49, 123, 141, 155]. As such, consumer AR devices often aim to maximize the FOV covered by the displays.

There are important engineering trade-offs, however, between a near-eye display’s FOV coverage and other factors such as device weight, display resolution, and eyebox size [93, 96]. For example, many current near-eye displays for AR use waveguides to optically combine virtual imagery with the natural FOV. Waveguides work by total-internal-reflection: the permissible direction of light that will be propagated through a waveguide and the coupling grating set limits on the achievable FOV [36, 97, 179]. In addition to optical factors, the physical requirements for covering a large region of the FOV with high spatial and temporal resolution are difficult to achieve without adding weight and size to the system, impacting the form factor [72]. New optical architectures are being developed to overcome these constraints (e.g., [23, 145, 173]), but are not yet mature or practical to manufacture. Currently, near-eye displays for AR still have quite limited FOV compared to the capacity of natural vision. Thus, finding ways to increase FOV coverage without sacrificing other important design factors is a high priority. Here, we describe in detail the concept of FOV for both natural human vision and stereoscopic AR display systems. We then report the results of user studies designed to develop updated perceptual guidelines for optimizing horizontal FOV in AR displays with minimal perceptual artifacts (specifically, visual nonuniformity). Our primary contributions are:

1. We clarify the importance of considering binocular overlap when quantifying FOV, both for natural vision and for stereoscopic AR systems.

2. We conduct user studies to evaluate two key design factors – the amount of binocular overlap and whether the overlap is convergent or divergent – that are thought to influence perceived FOV in AR. Our results suggest that increasing the amount of binocular overlap effectively reduces perceptual nonuniformity across the horizontal FOV. Contrary to prior work using simple stimuli, we find that divergent configurations are generally better than convergent configurations at the content distance.

3. While divergent configurations and large binocular overlap are preferred, we show that these properties of near-eye stereoscopic displays change when users look around a 3D scene. Using a simplified display model and combining this model with the user study results, we provide a guide to assist with determining the best display configuration for a given system.

## 4.2 Related Work

### 4.2.1 The Natural Human Field of View

Knowing people’s natural FOV, and how this FOV changes with eye movements, is important for creating technologies that aim to augment natural vision. The natural FOV limits the visual space that is available to the viewer at any given moment, and showing virtual content beyond this limit is excessive because the content will not be seen. On the other hand, if the FOV provided by AR is smaller than the natural FOV, this can compromise the immersive experience that these technologies seek to deliver.

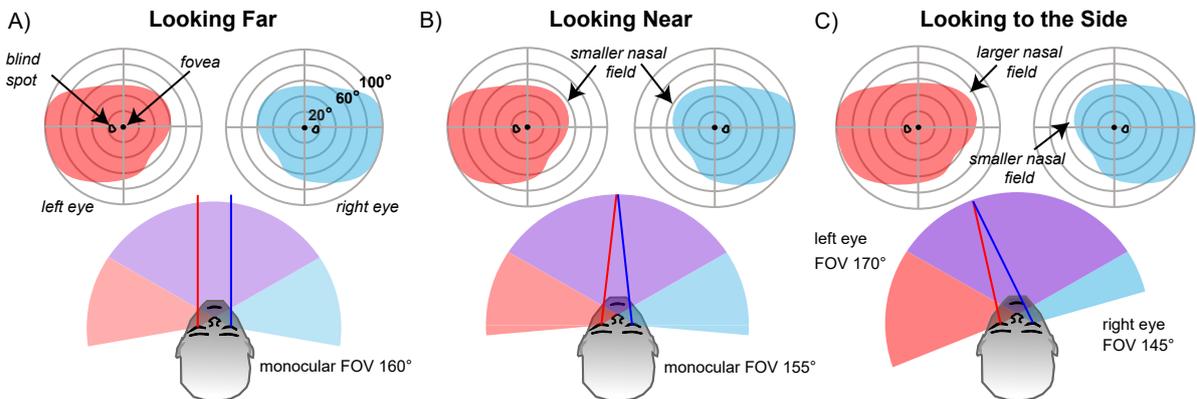


Figure 4.2: The natural FOV of human vision is illustrated, including the top-down view of the horizontal FOV (bottom) and a direct depiction of the left and right eye’s angular visual field (top). A) When the eyes are looking far and straight ahead, each eye’s temporal field subtends approximately 100° out from the fovea towards the temple, and the nasal field subtends approximately 60° towards the nose. B) When the eyes converge to a near distance, more of the nasal fields are blocked by the nose. As a result, both the monocular and the cyclopean FOV decrease. However, the amount of binocular overlap remains unchanged. C) When the eyes are looking to the side, the nasal limit of the FOV changes, expanding the FOV in one eye while shrinking in the other eye. The amount of expanding and shrinking is not necessarily equal between the two eyes, for example, if the eyes are also converged.

The natural FOV for human vision is defined as the angular region of visibility for each eye. These angles, together with the direction of gaze, determine the volume of visual

space that is visible at a given point in time. Different parts of this FOV are processed differently in the human visual system – for example, the fovea is a small region (about  $5^\circ$  [91]) of the retina with high resolution vision, so when people look around a scene, they direct the foveas of both eyes to examine the object of interest (fixation point) and this changes the volume of visible space. Beyond the foveas, each eye sees a large monocular FOV with a shape determined by various factors that are fixed to the anatomy of the eye and the surrounding facial structures [165]. For most people, the upper and nasal sides of the monocular FOVs are head-fixed (that is, they are limited by facial anatomy such as the brow and nose bridge), while the lower and temporal sides are retina-fixed (that is, they are limited by the edge of retina). For each eye, this monocular FOV extends approximately  $60^\circ$  upward,  $75^\circ$  downward,  $60^\circ$  nasally, and  $100^\circ$  temporally when the eyes look straight ahead [150].

The FOVs of the two eyes do not completely overlap each other. Certain regions of visual space are seen by both eyes (binocular) and other regions are seen by one eye only (monocular), as shown in the top-down views in Figure 4.2. The visual system takes these partially-overlapping retinal images from the two eyes and creates a *cyclopean* view of the world as if looking from a single viewpoint between the two eyes (as a cyclops in Greek mythology would). Despite having both monocular and binocular regions within this natural cyclopean FOV, the subjective experience is a seamless, singular view of the world. Together, both eyes provide a natural cyclopean FOV that extends to around  $\pm 100^\circ$  horizontally from the midsagittal plane when the eyes are gazing straight ahead, bounded by the temporal margin of each eye’s visual field (Figure 4.2A). The binocular overlap region (purple), which extends to around  $\pm 60^\circ$ , plays an important role in depth perception, allowing for stereopsis and precise depth discrimination.

The natural FOV is also dynamic due to eye movements (Figure 4.2B, C). For example, when people fixate on objects at near distances, their eyes rotate in opposite horizontal directions (called a vergence eye movement) and the horizontal sizes of each monocular FOV and the cyclopean FOV change. Specifically, when the eyes fixate a point that is near to the face, each eye rotates nasally (Figure 4.2B) to converge so that the near point falls on the foveas of both eyes. The nasal side of the monocular FOV of each eye shrinks (because the nose is head-fixed) and the cyclopean FOV also shrinks. On the other hand, the size of the binocular overlap region is the same because the visual angle taken up by the nose in the nasal field is compensated by the temporal field. The eyes also make conjugate horizontal and vertical movements to explore content within the same or different depth planes (Figure 4.2C). With horizontal eye movements, the binocular overlap region remains the same size and in the same position relative to the head, but the monocular and cyclopean FOVs change. Specifically, the monocular portions of the cyclopean FOV change such that one side increases relative to the midsagittal plane, and the other side decreases.

## 4.2.2 Field of View in Stereoscopic Near-Eye Displays

In AR, a portion of the natural FOV is superimposed with content from digital displays. To avoid occluding natural vision with the display surface, AR devices typically redirect the virtual image of a display to each eye. The physical display is located away from the natural FOV, whereas the virtual images are presented near the center of each eye’s natural FOV. The size, distance, and magnification of each virtual image determines

the angular size of the FOV that can be stimulated for each eye (the virtual monocular FOV) and their position in each eye’s field determines the virtual cyclopean FOV and the amount of binocular overlap.

An ideal AR system would allow the delivery of virtual content anywhere within the natural FOV of the viewer. Because the natural FOV is dynamic (it expands, shrinks, and reorients as described in the previous section), this would require each display to subtend a visual angle that is larger than the natural monocular FOV at any given point in time, or it would require a moving display that fills the instantaneous monocular FOV and moves with the eyes. However, existing systems cannot yet achieve this ideal. Commercially-available AR systems typically subtend around 30°-40° horizontally and vertically in each eye’s FOV, creating a rectangular region in which content can be presented [95].

The development of wider FOV near-eye stereoscopic displays for AR has been a topic of research for several decades. In the 1980’s, early near-eye displays were being developed for use in the military, and concerns were raised that a restricted FOV could be detrimental for certain operations such as target detection during flight (e.g., [166]). One consideration that emerged prominently during this period was the amount of full or partial binocular overlap in the virtual FOV [88]. In a full binocular overlap scenario, the cyclopean FOV coverage is the same as the monocular FOV coverage (Figure 4.3, left). However, in a partial overlap scenario, the total horizontal cyclopean FOV over which virtual content can be displayed is increased by horizontally displacing the physical displays. When both eyes view a virtual scene, a larger horizontal portion of the natural FOV is then filled with virtual content, with some regions of the virtual scene seen by both eyes, and other regions are seen by only one eye. This design yields a binocular overlap region flanked by two monocular regions, which is similar to the natural FOV except that the total coverage is still small enough to fit fully within the binocular overlap region of natural vision for most eye movements. Unlike the natural FOV, there are two possible configurations for partial overlap displays: convergent and divergent (Figure 4.3, middle and right). When looking at convergent displays, the monocular regions of the cyclopean FOV occur in the nasal field of each eye, while divergent displays present these monocular regions in the temporal field of each eye. This concept can be used for both VR and AR technologies. In either modality, it is clear that the horizontal FOV coverage in stereoscopic near-eye systems has an important degree of freedom that can modify both the total cyclopean coverage, and the amount that is binocularly-visible.

### 4.2.3 Nonuniformity Artifacts from Partial Field of View Overlap

The partially overlapping views illustrated in Figure 4.3 pose a problem for the visual system because the viewer sees monocular content within the binocular region of natural vision. Early work on partial overlap displays identified a range of perceptual artifacts associated with these designs [5, 58, 77, 84, 85, 86, 87, 88, 89, 90, 116, 119, 120, 131, 137]. First, a perceptual fading (sometimes called *luning*) of the content around the monocular-binocular border (blue lines in Figure 4.3) was noted [90, 119, 120]. The monocular-binocular border was also found to be associated with elevated detection thresholds for presented targets [89]. Lastly, partial overlap was shown to lead to perceived fragmentation, in which the monocular regions appear to break up from the binocular region in some aspect of appearance (e.g., different depth, brightness) [87]. We refer to these collectively

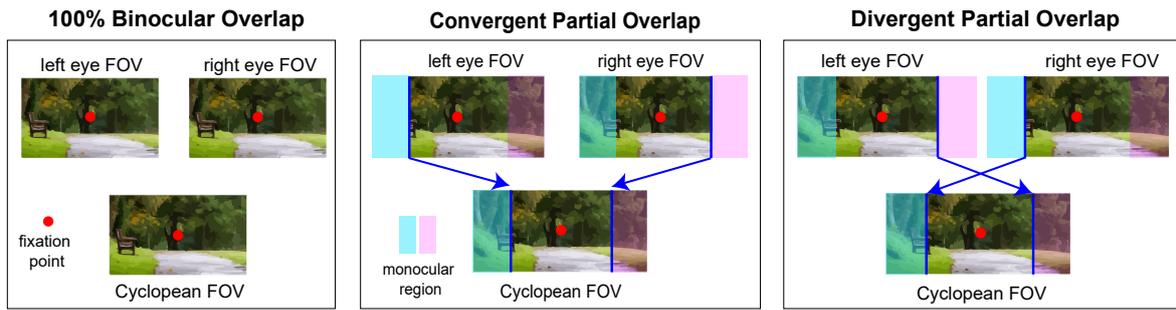


Figure 4.3: Illustrations of the monocular and cyclopean FOV subtended when a user fixates at the same location in space (thus the same vergence eye position) for different binocular configurations. Examples are shown for complete binocular overlap (left) and partial overlap imagery: convergent (middle) and a divergent (right). The monocular regions at corresponding locations in each eye are shaded with cyan and magenta. Similarly, monocular edges of the display (blue lines) are present in one eye but not the other, creating monocular-binocular borders in the cyclopean view. Image credit: Unsplash (Gary Ellis)

as nonuniformity artifacts.

These nonuniformity artifacts interfere with the percept of a continuous virtual FOV. The underlying cause of such artifacts is thought to be the fact that corresponding points in the two eyes are stimulated by highly discrepant stimuli. For example, in Figure 4.3, the cyan region of the left eye corresponds with the cyan region of the right eye (same for the magenta regions), but only one eye sees content in this region (in this illustration, the other eye simply sees nothing). This type of discrepancy in binocular inputs can lead to interocular suppression, in which the content seen by one eye is perceptually suppressed, or binocular rivalry, in which the content seen by the two eyes appears to alternate in time [22]. It is worth noting that monocular regions do occur in the binocular region of natural vision as well, so their existence is not wholly unnatural [63]. For example, when a foreground object occludes a background, often one eye sees more of the background at the occlusion boundary (a so-called partial occlusion).

Extensive perceptual studies sought to understand and reduce the nonuniformity artifacts associated with partial binocular overlap, resulting in a set of guidelines for how to maximize the horizontal FOV in wearable displays with minimal artifacts [87, 89, 90, 119]. Several different stimulus factors were found to reduce, but not eliminate, these artifacts and create a more coherent cyclopean view. These factors include adding a smooth luminance fall off towards the monocular-binocular border [119, 120], adjusting the relative luminance of the monocular regions [90], increasing the amount of binocular overlap [87], and adopting a convergent rather than a divergent display configuration (see [86, 88] for review). For example, one study measured how often people detected fading artifacts over the duration of half a minute trial, and showed that viewers reported seeing less fading over time with convergent overlap as compared to divergent overlap [90]. In another study, viewers chose convergent views to be better in terms of perceived uniformity across the display when both configurations were shown together at the same time on the screen to the viewer [87]. However, there are currently barriers to implementing these guidelines in modern stereoscopic AR displays. These prior studies largely used stimuli that do not reflect the visual appearance of current AR systems: they used simple gray scale images in which the entire FOV was limited to the virtual display with no other visual information outside of the display’s FOV [87, 89, 90]. In AR, a smaller display FOV is superimposed over the larger natural FOV of human vision. In addition, most

modern AR systems use additive light, making virtual content often semi-transparent. As such, both eyes will share more similar visual information in the monocular region, which may reduce the ability to detect nonuniformity artifacts. Thus, it is not clear that the strategies for mitigating artifacts will be similarly effective in AR.

## 4.3 Perceptual Studies

We conducted two perceptual studies to examine how the amount of binocular overlap and the display configuration (convergent and divergent) influence the perceived quality of the FOV in AR. We focused on using visuals that are more similar to AR applications than those found in the existing literature on partial binocular overlap. We also chose one type of nonuniformity artifact to focus on: the fading of content near the binocular-monocular border. If the results of prior studies extend to AR visuals, we would expect users to experience less fading when viewing convergent configurations and configurations with more binocular overlap.

### 4.3.1 Methods.

#### Participants.

In Experiment 1, twenty adults (ages 20 - 30 years, 1M 19F) participated. In Experiment 2, a different group of twenty adults (ages 19 - 27 years, 5M 15F) participated. One participant in Experiment 2 indicated that they could not fuse the fixation target, so their data were discarded and an additional participant was recruited. All participants had normal or corrected-to-normal visual acuity in both eyes, and normal stereo vision assessed with Randot Stereo Test. All participants were naïve to the study hypotheses, were compensated for their time, and gave informed consent for their participation. The experimental procedures were approved by the Institutional Review Board.

#### Display System.

Stimuli were presented on a desk-mounted mirror stereoscope with two LCD displays (LG-32UD99-W, maximum luminance of 138 cd/m<sup>2</sup> measured by a PR-650 photometer) as shown in Figure 4.4. The viewing distance from the participant’s eyes to each display was approximately 57 cm, resulting in a FOV for each display of 63° horizontally and 38° vertically. Each display was 3840 by 2160 pixels, resulting in a resolution of ~55 pixels per degree (~27.5 cycles per degree). During the study, the participants rested their chin on a chin rest while sitting in a dark room.

#### Stimuli.

We used three types of visual stimuli: a simple oval, a simple rectangle, and simulated AR. The oval and rectangle stimuli consisted of uniform white shapes presented on a black background (Figure 4.4, simple (oval) and simple (rectangle)), emulating stimuli that have been used to study perceptual artifacts from partial overlap in prior literature [84, 85, 87, 89, 90]. Thus, we expected to replicate previous findings with these stimuli (i.e., fewer artifacts with larger binocular overlap regions and fewer artifacts with a convergent configuration compared to divergent).

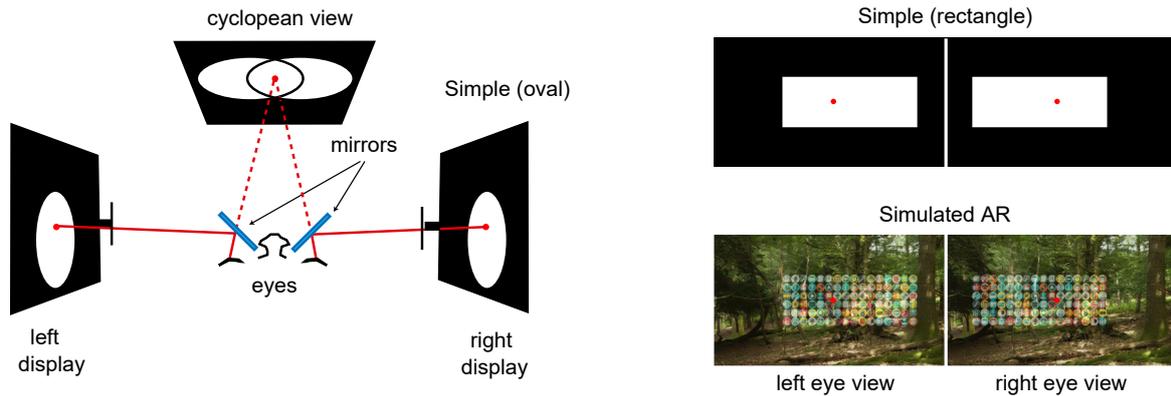


Figure 4.4: Left: A top-down view of the stereoscope setup (not drawn to scale). Participants viewed a pair of desk-mounted displays through a pair of mirrors and fixated on a red dot to fuse the left and right eye’s image. The cyclopean view illustrates the monocular-binocular borders where perceptual artifacts tend to occur, resulting in a nonuniform appearance of the white shape. Right: in addition to the simple oval shape shown on the left, two additional types of stimuli were used in the studies: simple rectangle and simulated AR. The monocular FOV could be either  $30^\circ$  or  $40^\circ$  wide and the monocular region could be either  $4.5^\circ$  or  $9^\circ$  wide. The red fixation dot is exaggerated for visibility.

A typical use case of AR is showing application icons against the real environment. Thus, for the simulated AR stimuli, we used icon arrays for mobile applications [1] and tiled them on a virtual display that was superimposed over a stereoscopic natural background (Figure 4.4, simulated AR). For the natural backgrounds, we selected scenes from the SYNS natural stereo image dataset [2]. Adjacent views were stitched together to create a wider FOV image to fill our display [24]. These scenes included both outdoor and indoor environments, captured with stereo cameras with 6.3 cm separation. To combine the icons and the backgrounds, the background intensity was first normalized from 0 to 1 for all scenes, and then all pixel values were reduced by 66.6%. The icon’s pixel intensities were normalized to range from 0 to 33.3% and then added to the background image. These percentages were selected in order to produce imagery in which the AR icons were clearly visible but also appeared semi-transparent as in most optical see-through AR systems. In reality, the amount of perceived contrast of AR content relative to the real background depends on both the luminance level of the environment and the settings of the display.

We varied several properties of the stimuli and examined the effect on nonuniformity artifacts. First, we examined whether the binocular overlap region size influenced nonuniformity artifacts for each stimulus type. To simulate binocular and monocular regions that may be typical of current AR systems, we tested two horizontal monocular FOV sizes for the virtual content:  $30^\circ$  and  $40^\circ$ , and two monocular region sizes:  $4.5^\circ$  and  $9^\circ$ . This resulted in binocular overlap region sizes that ranged from  $21^\circ$  to  $35.5^\circ$  horizontally, with horizontal cyclopean FOVs ranging from  $34.5^\circ$  to  $49^\circ$ . All stimuli were  $15^\circ$  tall vertically. We examined convergent versus divergent partial overlap for each of these binocular region sizes, resulting in eight conditions total for each stimulus type (oval/rectangle/AR).

The AR icons and the simple shapes were rendered at a fixed vergence distance of 1.5 m based on an interpupillary distance (IPD) of 6 cm, such that the icon arrays appeared to float in front of the background. An IPD of 6 cm was chosen for rendering since it allowed a majority of the viewers to fuse the stimuli without having to adjust the stereoscope. Individual IPDs were not measured due to social distancing protocols. To aid fusion, a

red fixation dot was presented to match the vergence angle needed to fuse the binocular overlap regions in the two eyes' views.

### 4.3.2 Experiment 1.

#### Procedure.

In this experiment, participants were instructed to continuously indicate when they saw any fading of the stimuli (e.g., luning, fragmentation). They held down one keyboard key when they saw fading, and a different key when they did not see fading. On each trial, a particular combination of stimulus type (oval/rectangle/AR), binocular overlap region size, and convergent or divergent overlap was shown in pseudo-random order. Each trial was 30 seconds long. Participants were instructed to look at the fixation dot at the center of the screen for the duration of each trial so that within a trial, the amount of partial overlap was always fixed and determined by the stimulus design. Between trials, a uniform gray screen was shown for five seconds. Five unique AR scenes were used for this experiment. There were total of 56 trials.

#### Analysis.

For each trial, we calculated the proportion of time that participants indicated that they saw fading, excluding the first five seconds to account for delays in starting to respond. Trials were included in analysis if at least 90% of the recorded key presses were valid responses (i.e., the participant pressed one of the two response keys). Based on this threshold, 2.6% of trials were omitted. When looking at the data with different thresholds, the results were very similar. For the valid trials, the proportion of fading time was calculated by dividing the duration that the participant indicated fading over the total duration of the responses.

To examine the effects of the stimulus type, binocular region size, and convergent/divergence configuration on perceived fading, we fit the response data with a logistic regression model. Because the response data were bimodally distributed (many, but not all, proportions were near 0 or 1), prior to fitting the model we re-coded the responses into a binary variable indicating whether or not more than half of each trial had fading. The initial model included three main effects and all two-way interactions. In this initial model, we included the oval and rectangle shapes as separate categorical predictors, however, we did not observe an effect of the rectangle compared the oval so in the final analyses we combine these into a single variable that we call “simple stimuli”, which was compared to the simulated AR. Individual participants were modelled as random intercepts. Follow up logistic regression models were used to examine the significant interactions. For all statistical tests, a significance threshold of  $p < 0.05$  was used.

#### Results.

The results are shown in Figure 4.5A, with the simple stimuli on the left and the AR stimuli on the right. The x-axis shows the size of the binocular region. The convergent and divergent configurations are plotted separately. Recall that previous studies found that a larger binocular region corresponded with fewer perceived artifacts [87]. Qualitatively, our data is consistent with prior work: there is a negative trend between the binocular region size and the proportion of time that participants reported fading. However, this

effect seems to be stronger for simple stimuli compared to AR stimuli. Also consistent with prior literature, we see that the convergent configuration was associated with less fading compared to the divergent configuration for the simple stimuli. However, this did not appear to be the case for the AR stimuli, for which the convergent and divergent stimuli elicited similar responses.

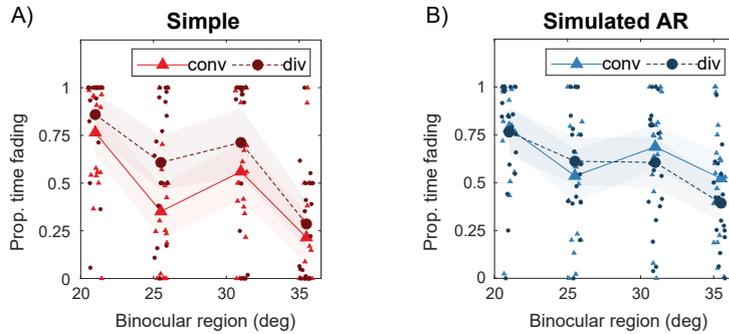


Figure 4.5: Experiment 1 results for A) simple stimuli and B) simulated AR stimuli. The mean and 95% confidence intervals for the average proportion of time that participants reported fading during a trial are represented by the large symbols and the shaded regions. Different colors indicate different stimulus configurations. Individual subject's proportion data are plotted as small dots. The binocular region size from small to large corresponds to the four different monocular FOV and monocular region combinations: 30° and 9°, 30° and 4.5°, 40° and 9°, 40° and 4.5°

The logistic regression model was consistent with this interpretation of the data. The model indicated that the coefficients for stimulus type, stimulus configuration, and binocular region size were all statistically significant. In addition, there were significant interactions between stimulus type and stimulus configuration, and stimulus type and binocular region size. The model accounted for 58% of the variation in the fitted data and the results are shown in Table 4.1. For the categorical predictors (stimulus type and configuration), we modelled the *simple* stimuli condition and the *convergent* condition as the intercept, so regression coefficients reflect the relative effects of the *AR* and *divergent* conditions. To better understand the interactions, we conducted a series of follow up analyses by fitting models to subsets of the data, based on the different categories of the main effects, and compared among them (Table 4.2). First, we examined the effect of configuration (convergent/divergent) separately for the simple stimuli and for the AR stimuli. For the simple stimuli, the divergent configuration was associated with an increase in fading compared to convergent configuration. Qualitatively, this increase is associated most strongly with the binocular region size of 25°. There was no significant difference between the configurations for the AR stimuli. Next, we examined the effect of binocular region size separately for the simple stimuli and for the AR stimuli. Both the simple stimuli and AR stimuli were associated with a significant decrease in fading as binocular region size increased, but the magnitude of the decrease was larger for the simple stimuli.

### Exploratory Analysis.

In our main analysis, we chose to focus on binocular region size as a predictor because prior work suggested that this is a more reliable predictor of fading than monocular FOV, monocular region size, and total cyclopean FOV alone [87]. However, in the limit, binocular region size is unlikely to explain all of the variance in fading because it does not take the monocular region size into account at all. In an exploratory analysis, we

Table 4.1: Logistic regression model fit to Experiment 1 data. For each predictor, the coefficient reflects an increase or decrease in the probability that fading was perceived for more than half of each trial. Positive values indicate more fading, and negative values indicate less fading. Coefficients that are significantly different from zero based on the  $t$ -statistics are marked with asterisks (\*).

Variable	Coefficient (95% CI)	$t$	$p$
Type (AR)	-2.16 (-3.91, -0.40)	-2.42	0.02*
Config. (Divergent)	2.05 (0.46, 3.64)	2.53	0.01*
Bino. region	-0.17 (-0.23, -0.11)	-5.78	< 0.001*
AR*Divergent	-0.97 (-1.60, -0.35)	-3.05	0.002*
AR*Bino. region	0.10 (0.04, 0.16)	3.38	< 0.001*
Divergent*Bino. region	-0.04 (-0.09, -0.01)	-1.65	0.10
Intercept	4.61 (2.86, 6.36)	5.18	< 0.001*

Table 4.2: Follow up tests examining the interaction terms in the main model for Experiment 1. Data are split into subsets based on stimulus type. Statistically significant coefficients are marked with asterisks (\*).

Variable	Coefficient (95% CI)	$t$	$p$
Simple: Convergent vs. Divergent			
Divergent	0.64 (0.17, 1.11)	2.65	0.01*
Intercept	-0.12 (-0.62, 0.38)	-0.48	0.63
AR: Convergent vs. Divergent			
Divergent	-0.17 (-0.49, 0.14)	-1.06	0.29
Intercept	0.62 (0.04, 1.20)	2.11	0.04*
Simple: Bino. Region			
Bino. region	-0.19 (-0.24, -0.13)	-7.05	< 0.001*
Intercept	5.49 (3.90, 7.07)	6.82	< 0.001*
AR: Bino. Region			
Bino. region	-0.09 (-0.12, -0.06)	-5.84	< 0.001*
Intercept	3.11 (2.06, 4.16)	5.81	< 0.001*

looked at whether different ways of characterizing the virtual FOV correlate better with the proportion of fading time in our data. In addition to the binocular region size as suggested by previous work, we plotted our average AR results against the total cyclopean FOV and the proportion of each eye’s FOV that is only monocularly-visible (i.e., the size of the monocular region divided by the size of FOV in one eye) (Figure 4.6). We fitted a simple linear regression line to these data. For these data, we found that using the monocular proportion explains more of the variance in responses ( $R^2 = 0.85$ ) compared to the original binocular region size metric ( $R^2 = 0.61$ ). The total cyclopean FOV metric was not a strong predictor of fading in these data ( $R^2 = 0.02$ ). While by no means definitive, this strong trend suggests that as the ratio between the monocular region and the FOV increases, there may be a roughly linear increase in the fading time, at least for the combinations of parameters tested in our experiment. We will return to this measure as a possible predictor of perceptual artifacts in stereoscopic AR systems in the Discussion.

## Summary.

Experiment 1 replicated previous findings for simple stimuli, but suggests that the impact of binocular region size and convergent/divergent configuration differ for stimuli more closely approximating AR. Specifically, we did not find evidence that convergent configurations produce fewer perceptual nonuniformity artifacts in AR, suggesting that there is no need to favor systems that create convergent overlap over divergent overlap as previously thought—an important potential opportunity to relax design constraints when

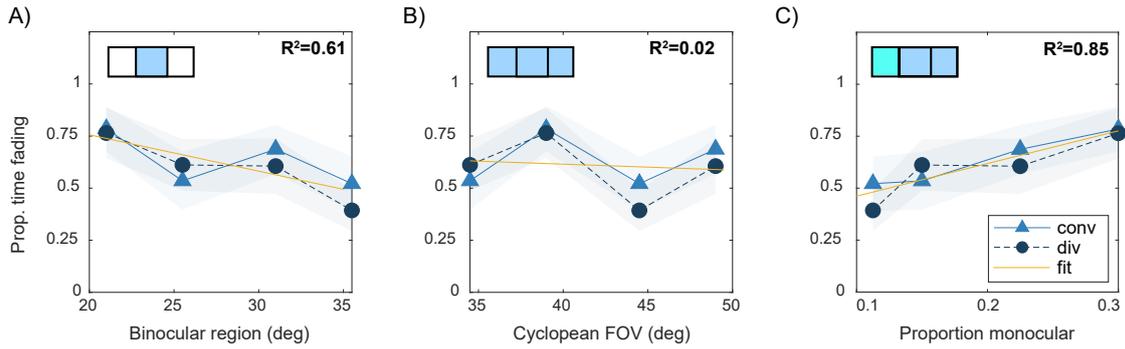


Figure 4.6: Exploratory analysis looking at the proportion fading of the AR stimuli as a function of: A) binocular region size, with monocular FOV and monocular region sizes of  $30^\circ$  and  $9^\circ$ ,  $30^\circ$  and  $4.5^\circ$ ,  $40^\circ$  and  $9^\circ$ ,  $40^\circ$  and  $4.5^\circ$  from left to right B) cyclopean FOV, with monocular FOV and monocular region sizes of  $30^\circ$  and  $4.5^\circ$ ,  $30^\circ$  and  $9^\circ$ ,  $40^\circ$  and  $4.5^\circ$ ,  $40^\circ$  and  $9^\circ$  from left to right, and C) proportion of each eye’s FOV that is only monocularly-visible, with monocular FOV and monocular region sizes of  $40^\circ$  and  $4.5^\circ$ ,  $30^\circ$  and  $4.5^\circ$ ,  $40^\circ$  and  $9^\circ$ ,  $30^\circ$  and  $4.5^\circ$  from left to right,. The  $R^2$  values for the linear regressions are shown.

FOV is being optimized along with other factors. In the next experiment, we further examined the difference between convergent and divergent partial overlap.

### 4.3.3 Experiment 2.

#### Procedure.

In this experiment, participants were asked to directly compare two stimuli that were identical (the same stimulus type and binocular region size) except that one was in a convergent configuration and one was in a divergent configuration. Their task was to select the one that had “a wider field-of-view with minimum fading of the content”. Because the FOV was actually identical for both stimuli on each trial, these instructions served to resolve ambiguous cases in which fading was so strong that the FOV actually appeared smaller for one stimulus. On each trial, participants could toggle back and forth between the two stimuli without any time limit. Twenty AR scenes were used for this experiment and each was repeated 4 times with a different icon set. The simple stimuli were each repeated 10 times. There were 400 trials total. As in Experiment 1, participants were instructed to fixate a point in the middle of the screen for the duration of the trial.

#### Analysis.

For each unique condition, we calculated the proportion of trials that each participant chose the convergent stimulus over the divergent stimulus. We also performed a logistic regression on the trial-by-trial data, as described for Experiment 1. In this case, each response was coded with 0 if the divergent stimulus was chosen and 1 if the convergent stimulus was chosen.

#### Results.

In this experiment, for both stimulus types (simple and AR), participants had an overall preference for divergent over convergent stimuli. In Figure 4.7A, the results are plotted with the x-axis indicating the size of the binocular region and the y-axis indicating

the proportion of trials for which the convergent stimulus was preferred. Values greater than 0.5 indicate a preference for convergent stimuli, and values less than 0.5 indicate a preference for divergent stimuli. For comparison, in Figure 4.7B, we replot the data from Experiment 1, calculating the difference between convergent and divergent trials for the same stimulus, where a value greater than zero indicates that convergent trials had less fading, and less than zero indicates that divergent trials had less fading. From this comparison, it is clear that the forced-choice task in Experiment 2 resulted in a greater preference for the divergent configuration.

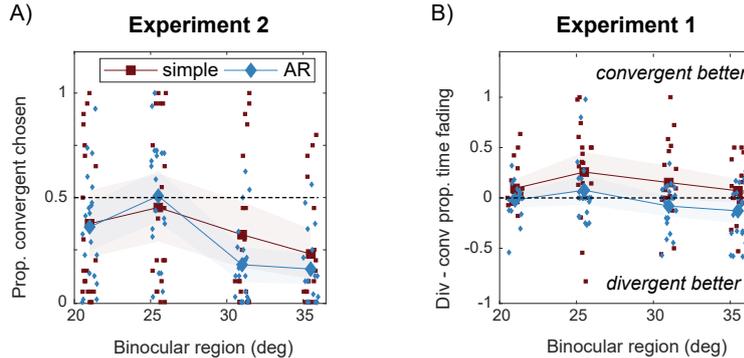


Figure 4.7: A) For Experiment 2, the proportion of trials that the convergent configuration was chosen over the divergent configuration. B) For Experiment 1, the difference in proportion of fading time (divergent - convergent). In both panels, the dashed line represents the point of equality for convergent and divergent. Above the line means the convergent stimulus was preferred, and below the line means divergent was preferred. Data are otherwise plotted the same as Figure 4.5.

The results of the logistic regression for Experiment 2 are shown in Table 4.3. The main effects of stimulus type and binocular region size were both statistically significant, as was their interaction. In a follow up analysis, we performed two logistic regressions by categorizing the binocular region sizes into two levels: small ( $< 30^\circ$ ) and large ( $> 30^\circ$ ), and ran the model with each subset of the data to compare simple stimuli and AR stimuli. We grouped the binocular region sizes into two levels rather than running separate models for all four levels for ease of interpretation. These results are shown in Table 4.4. The results suggest that binocular region size modulated the effect of stimulus type on convergent preference. For small binocular region sizes, there was no difference between the AR and simple stimuli. For larger binocular regions, AR stimuli were associated with a stronger divergent preference than the simple stimuli. However, there was also a large amount of variation in the data across participants, and full model only accounted for 32% of the variance in the fitted data.

Table 4.3: Logistic regression model fit to Experiment 2 data. For each variable, coefficients reflect the change in probability of choosing convergent over divergent stimuli (positive values indicate more likely to choose convergent, and negative values indicate less likely). Coefficients that are significantly different from zero based on the  $t$ -statistics are marked with asterisks (\*).

Variable	Coefficient (95% CI)	$t$	$p$
Type (AR)	1.24 (0.56, 1.94)	3.54	$< 0.001^*$
Bino. region	-0.07 (-0.09, -0.05)	-6.07	$< 0.001^*$
AR*Bino. region	-0.05 (-0.08, -0.03)	-4.39	$< 0.001^*$
Intercept	1.02 (0.20, 1.83)	2.44	$0.01^*$

This regression analysis indicates how likely convergent stimuli were to be chosen across different conditions. However, it does not answer whether the tendency to select the

Table 4.4: Follow up analysis on the interaction between binocular region size and stimulus types on the probability that convergent was chosen. Statistically significant coefficients are marked with asterisks (\*).

Variable	Coefficient (95% CI)	$t$	$p$
Small Binocular Region: Simple vs. AR			
AR	0.11 (-0.07, 0.29)	1.17	0.24
Intercept	-0.44 (-1.05, 0.17)	-1.43	0.15
Large Binocular Region: Simple vs. AR			
AR	-0.83 (-1.04, -0.62)	-7.78	< 0.001*
Intercept	-1.46 (-2.23, -0.70)	-3.74	< 0.001*

Table 4.5: Chi-square test results for whether or not the number of people preferring convergent or divergent differed significantly based on stimulus type and binocular region size in Experiment 2. Statistically significant results are marked with asterisks (\*).

Bino. Region	Num. Conv.	Num. Div.	$\chi^2$	$p$
Simple				
21°	7	13	1.8	0.18
25.5°	8	12	0.8	0.37
31°	5	15	5	0.03*
35.5°	4	16	7.2	0.007*
AR				
21°	5	15	5	0.03*
25.5°	10	10	0	1
31°	2	18	12.8	0.001*
35.5°	1	19	16.2	0.001*

convergent or divergent stimuli was significantly different from chance (i.e., that both are equally likely to be chosen). Thus, to investigate whether or not there was a significant preference for convergent or divergent stimuli in Experiment 2, we performed a chi-square goodness-of-fit test and asked whether the proportion of people preferring convergent and divergent deviated significantly from the expected values if half of the participants preferred convergent and half preferred divergent. For each condition, we coded each participant as having a preference for convergent if they chose convergent more than 50% of the time and as having a preference for divergent if they chose convergent less than or equal to 50% of the time. If there was equal preference for convergent and divergent stimuli, we would expect that 10 participants would prefer one over the other. The results are shown in Table 4.5. For larger binocular regions ( $\geq 30^\circ$ ), we see that most people preferred divergent for both shape and AR stimuli. For smaller binocular region sizes, the results are more mixed, and the only significant effect was a divergent preference for the AR stimuli with the smallest binocular region. Thus, contrary to previous work [87], we found no significant convergent preference across all conditions in this experiment.

## Summary.

When participants were asked to make a direct comparison of convergent and divergent configurations in Experiment 2, we again observed no consistent preference for convergent AR stimuli. However, unlike Experiment 1, we observed a tendency for participants to prefer the divergent configuration for both stimulus types, especially for large binocular region sizes. These results suggest that there may be task-dependent or timing-dependent differences in perceived nonuniformity, which we will take up in the Discussion.

## 4.4 3D Field of View in Augmented Reality

In the previous section, we treated binocular region size and convergent/divergent overlap as static properties of an AR system, similar to prior literature. However, this is not accurate for AR systems in which people view and interact with content at both near and far distances. When eyes converge and diverge, the display’s monocular FOV limits stay fixed in the world but they are not fixed on the retina, which can alter the sign and amount of partial overlap. The binocular overlap, the total FOV, and configuration can change quite substantially depending on where the viewer is looking. In this section, we provide a simplified model for determining the cyclopean FOV and binocular overlap when different fixation distances are taken into account (see [4] for a similar analysis based on VR systems). This model, combined with the perceptual study results, allows for maximizing the cyclopean FOV over particular distances depending on the use case and display configuration.

### 4.4.1 Viewing Geometry and Camera Frusta

To create 3D AR experiences on a near-eye display, a pair of stereoscopic images needs to be generated and displayed according to the appropriate 3D viewing geometry. Specifically, two images should be constructed such that the directions of the light rays entering each eye’s pupil from the displays match the intended locations of virtual objects in the 3D space. Typically, virtual cameras that are horizontally offset from each other are used to project a 3D virtual world into a pair of images. Each virtual camera will capture a left or a right view, which are then presented to the corresponding eye of the user. To achieve the correct 3D viewing, the horizontal offset of the cameras needs to match the user’s interpupillary distance (IPD) [67], the dimensions of each camera’s viewing frustum need to match the visual angle subtended by each display, and the frusta positions need to align with the visual direction of the displays relative to the viewer’s eyes (see Figure 4.1). In the following analyses, we assume that this viewing geometry is followed, but see [13] for a more systematic description of possible configurations to achieve correct stereoscopic viewing geometry. Because we assume correct viewing geometry, the camera’s FOV is equivalent to the FOV subtended by the AR content for the user and we will visualize the virtual camera frusta in front of the eyes of the observer, corresponding to the region of the virtual world that can be presented and merged with the natural FOV.

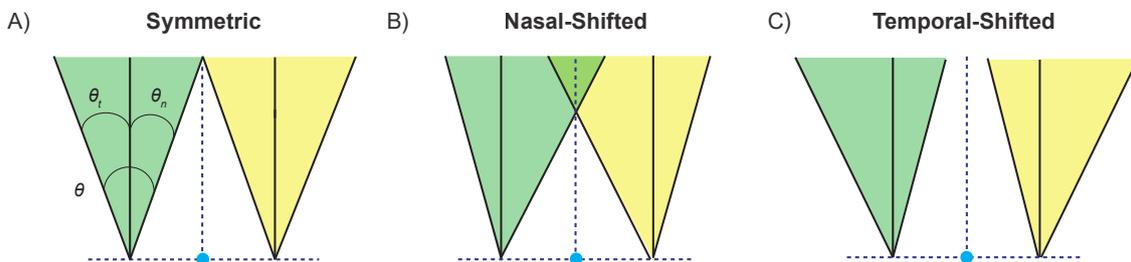


Figure 4.8: Examples of three types of camera frusta, showing the horizontal monocular FOV ( $\theta$ ), nasal field ( $\theta_n$ ), and temporal field ( $\theta_t$ ). Each camera has a  $\theta$  of  $40^\circ$  horizontally. A) Symmetric frusta,  $|\theta_n| = |\theta_t|$ . B) Nasal-shifted frusta,  $|\theta_n| > |\theta_t|$ . C) Temporal-shifted frusta,  $|\theta_n| < |\theta_t|$ . The dashed lines indicate a cyclopean coordinate system, with the blue dot indicating the origin.

In this setup, each virtual camera’s horizontal frustum determines the horizontal monocular FOV angle ( $\theta$ ) covered by the display for each eye. A standard camera frustum

is symmetrical about the camera’s optical axis – that is, the nasal field ( $\theta_n$ ) and the temporal field ( $\theta_t$ ) are equal (Figure 4.8A). But other arrangements are possible in which the nasal field and the temporal field are not equal (Figure 4.8B, C). These arrangements are called asymmetric frusta. There are two types of asymmetric frusta, which we will call nasal-shifted and temporal-shifted to avoid confusion with the convergent and divergent terminology that has been used historically for partial overlap displays. With nasal-shifted frusta, both cameras’ nasal fields are extended and the temporal fields are reduced (Figure 4.8B). The left camera captures more content in the right field, and the right camera captures more content in the left field. For temporal-shifted frusta, the temporal fields are extended and the nasal fields are reduced, so the left camera captures more of the left field and vice versa for the right camera (Figure 4.8C).

#### 4.4.2 Qualitative Horizontal Field-of-View Analysis

In 3D AR, the angular cyclopean FOV (which we will refer to as  $\gamma$ ) and the angular size of the binocular overlap region ( $\gamma_b$ ) measured from the midpoint between the two eyes depend on the size of the individual camera frusta (monocular FOV,  $\theta$ ), the frustum configuration (nasal-shifted or temporal-shifted), the camera separation ( $a$  — in this case equal to the user’s IPD), and the distance to the content that is being fixated ( $d$ ). Nasal-shifted and temporal-shifted frusta are generally associated with the convergent and divergent partial overlap described in previous work, respectively. But we will show that there is not a one-to-one relationship between the binocular overlap considered in previous work and frustum asymmetry when considering 3D AR content.

We start here with a geometric demonstration of how two variables — fixation distance and frustum asymmetry — interact. A visual comparison of the different camera frusta is shown in Figure 4.9 when the eyes are looking at a far object (10 m) and a near object (0.3 m). As illustrated, each camera configuration does not have a fixed cyclopean FOV ( $\gamma$ ) and binocular overlap ( $\gamma_b$ ), instead, these values change dynamically depending on the distance that is being fixated.

For far fixation distances with a symmetric configuration (Figure 4.9A, top), the two monocular frusta fully overlap each other, so the cyclopean FOV is roughly equal to each camera’s monocular FOV and the cyclopean FOV is (almost) fully binocular. At far fixation distances, nasal-shifted and temporal-shifted frusta result in a larger cyclopean FOV but smaller binocular region (Figure 4.9B, C top). The nasal-shifted frusta create convergent partial overlap, and the temporal-shifted frusta create divergent overlap at far distances.

At nearer fixation distances, the eyes converge and fixate on points that are shifted rightward on the left display and leftward on the right display. For symmetric and temporal-shifted frusta, this results in a view that now has divergent partial overlap (Figure 4.9A, C bottom). For nasal-shifted frusta, the overlap direction and amount are both distance-dependent: eye convergence may result in convergent partial overlap, divergent partial overlap, or full overlap of content (Figure 4.9B bottom, see next section). Thus, asymmetric frusta achieve a wider FOV compared to symmetric frusta for most, but not for all, distances. In short, the preferred frustum configuration to achieve a certain FOV varies with fixation distance.

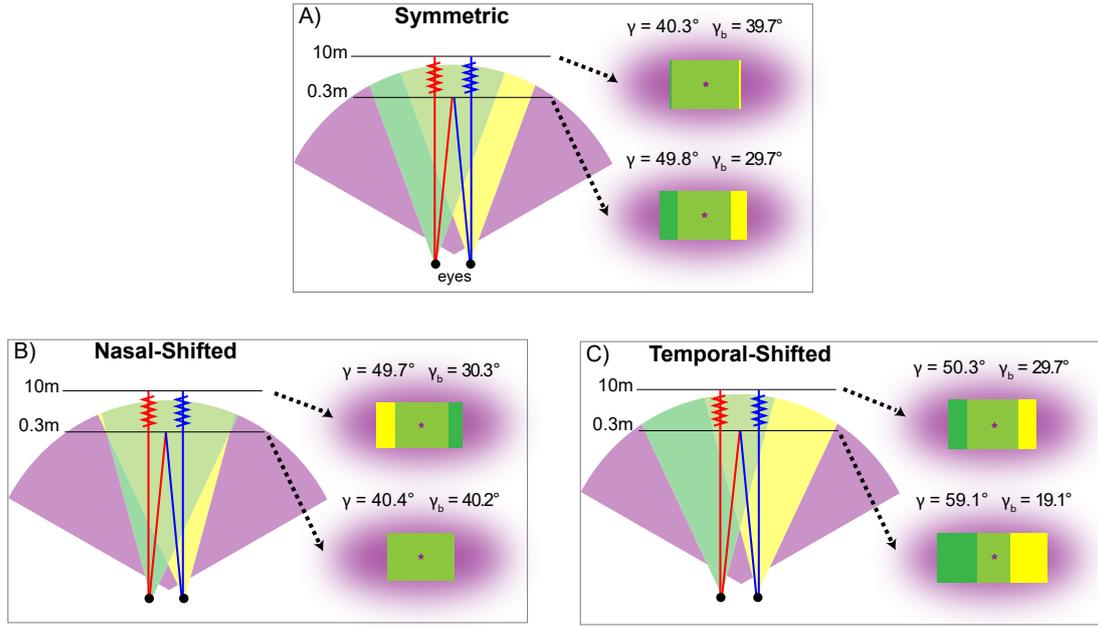


Figure 4.9: Comparison of the cyclopean FOV ( $\gamma$ ) and binocular overlap region ( $\gamma_b$ ) for A) symmetric camera frusta, B) asymmetric nasal-shifted camera frusta, and C) asymmetric temporal-shifted camera frusta at two different planes of fixation. Fixation distance is not drawn to scale. For each panel, monocular FOV  $\theta = 40^\circ$ , asymmetry  $s = 0^\circ, +10^\circ$  or  $-10^\circ$ , and interpupillary distance  $a = 6$  cm

### 4.4.3 Quantitative Horizontal Field-of-View Analysis

Here we describe how to calculate the horizontal cyclopean FOV and binocular overlap for a given configuration. We adopt a 2D Cartesian coordinate system with the cyclopean eye at the origin (midpoint between the two eyes), the x-axis co-linear to the interocular axis (positive rightward), and the z-axis co-linear to the midsagittal plane (positive forward). Clockwise angles in this system are positive.

The angular horizontal extent (in radians) for each camera frustum ( $\theta$ ) is defined as the sum of the magnitude of the nasal field,  $\theta_n$  and the temporal field,  $\theta_t$ :

$$\theta = |\theta_n| + |\theta_t|. \quad (4.1)$$

We define the amount of asymmetry ( $s$ ) as the difference in magnitude of these two angles:

$$s = |\theta_n| - |\theta_t|. \quad (4.2)$$

When  $|\theta_n| = |\theta_t|$ , then  $s = 0$  and the cameras have symmetric frusta. To make asymmetric frusta but maintain the same monocular FOV, an equal angle is added to  $|\theta_n|$  and subtracted from  $|\theta_t|$  for nasal-shifted ( $s > 0$ ), and vice versa for temporal-shifted ( $s < 0$ ). We assume the shift is small such that the angular size,  $\theta$ , is relatively constant for a given display size.

To indicate a distance at which to calculate the FOV, we define a line within this coordinate system at distance  $d$  in front of the eyes, parallel to the interocular axis. The bounds of the frusta intersect this line at four points:  $\ell_L, \ell_R, r_L, r_R$ , where  $\ell$  and  $r$  denote

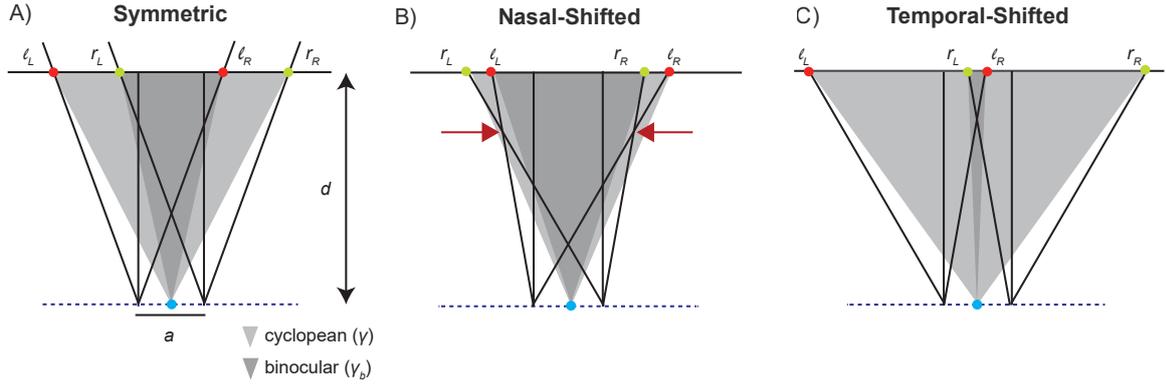


Figure 4.10: Examples of three types of camera frusta with the same distance ( $d$ ) to the depth plane of interest. The points of intersection bounding the binocular region ( $\gamma_b$ ) and the cyclopean FOV ( $\gamma$ ) are illustrated on this depth plane. A) Symmetric frusta. B) Nasal-shifted frusta. The red arrows indicate the plane at which the two frusta completely overlap each other. C) Temporal-shifted frusta. For these frusta, there is no distance with complete overlap.

the left and right cameras, and L and R denote the left and right bounds of the frusta (Figure 4.10). The x-coordinates of these points are:

$$\begin{aligned}
 \ell_L &= d \tan(-|\theta_t|) - \frac{a}{2} \\
 r_L &= d \tan(-|\theta_n|) + \frac{a}{2} \\
 \ell_R &= d \tan(|\theta_n|) - \frac{a}{2} \\
 r_R &= d \tan(|\theta_t|) + \frac{a}{2}.
 \end{aligned} \tag{4.3}$$

Note, the sign of the angles differs depending on the reference eye because the temporal field is clockwise from the right eye, but counterclockwise from the left eye. Because the two cameras' frusta are mirror-symmetric about the origin, all calculations can be done by considering only the right-side edges of the frusta ( $\ell_R, r_R$ ) and multiplying by a factor of two. These points will define the angular cyclopean FOV ( $\gamma$ ) and binocular region size ( $\gamma_b$ ) for that distance. The point with the smaller x-value ( $\min(\ell_R, r_R)$ ) will bound the binocular region ( $\gamma_b$ ) on the right side, and the point with the larger x-value ( $\max(\ell_R, r_R)$ ) will bound the cyclopean FOV ( $\gamma$ ):

$$\gamma_b = 2 \tan^{-1} \left( \frac{\min(\ell_R, r_R)}{d} \right) \tag{4.4}$$

$$\gamma = 2 \tan^{-1} \left( \frac{\max(\ell_R, r_R)}{d} \right). \tag{4.5}$$

Note that when  $\gamma_b$  is negative, there is no binocular overlap for content at that distance.

We can now quantitatively examine  $\gamma$  and  $\gamma_b$  for different display configurations and viewing situations. Figure 4.11A shows examples of the horizontal cyclopean FOV ( $\gamma$ , solid lines) and binocular region ( $\gamma_b$ , dotted lines) for different camera configurations at different depth planes for a fixed  $\theta$  of  $40^\circ$  and  $a$  of 6 cm. All angles have been converted to degrees. For symmetric frusta (purple lines), the horizontal offset created by  $a$  is negligible

at large distances and the two frusta have essentially 100% binocular overlap with  $\gamma$  equal to  $\theta$ . At close distances,  $\gamma$  increases, but  $\gamma_b$  decreases, due to increasing partial (divergent) overlap of the two eyes' views.

For asymmetric frusta, we plot the results for an asymmetry ( $s$ ) of  $10^\circ$  in either direction. The cyclopean FOV ( $\gamma$ ) at far distances is larger than for symmetric frusta, but the binocular region ( $\gamma_b$ ) is smaller. With temporal-shifted frusta (dark green lines),  $\gamma$  increases at nearer distances and  $\gamma_b$  decreases (Figure 4.9), becoming quite small for fixation distances less than 50 cm. At all distances, the binocular overlap is divergent. For nasal-shifted frusta (light green lines), the effect of fixating nearer is quite different: fixating at nearer distances decreases  $\gamma$  and increases  $\gamma_b$  (Figure 4.9). There is a distance where 100% overlap occurs, shown here at 30 cm. For distances nearer than this, the line slope reverses. This 100% overlap distance reflects a transition from convergent to divergent overlap. Thus, nasal-shifted and temporal-shifted frusta will result in quite different patterns of binocular overlap as a user looks around a 3D scene. Increasing the amount of nasal or temporal asymmetry ( $s$ ) effectively shifts the lines defining  $\gamma$  upwards, and shifts the lines defining  $\gamma_b$  downwards. By eye, it appears that nasal-shifted frusta provide a better compromise between expanding the cyclopean FOV across a range of distances, without producing overly small binocular overlap regions at any distance.

For nasal-shifted frusta, the distance that results in 100% binocular overlap ( $d_0$ ), can be calculated as follows:

$$d_0 = \frac{a}{\tan(|\theta_n|) - \tan(|\theta_t|)}. \quad (4.6)$$

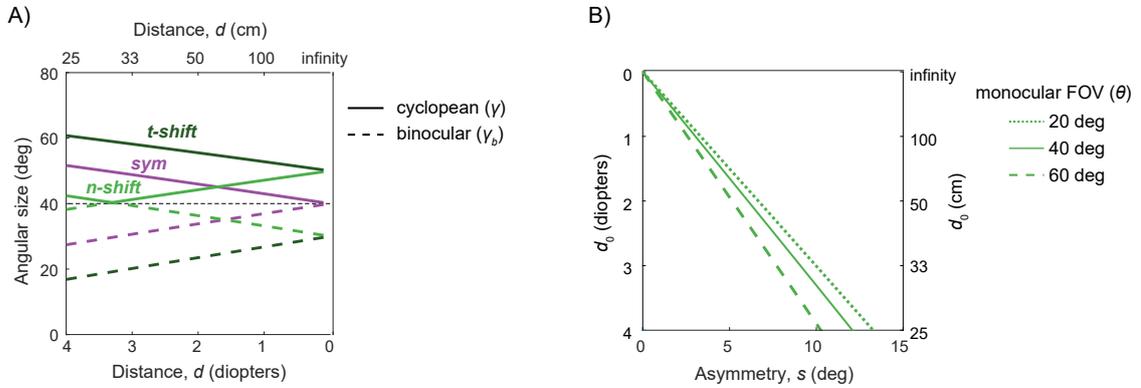


Figure 4.11: A) The cyclopean FOV and binocular overlap region size for the three types of frusta ( $sym$  = symmetric;  $n$ -shift = nasal-shifted,  $t$ -shift = temporal shifted). Distances are reported in both diopters and centimeters for fixed parameters of:  $\theta=40^\circ$ ,  $a = 6$  cm, and  $s = \pm 10^\circ$ . B) For different values of  $\theta$ , the content distance that contains complete overlap ( $d_0$ ) varies by the amount of asymmetry ( $s$ ) of the nasal-shifted frustum.

In Figure 4.11B, we plot  $d_0$  as a function of the nasal-shifted asymmetry angles ( $s$ ) for three different monocular FOV ( $\theta$ ) values. As  $s$  increases,  $d_0$  moves to closer and closer depth planes. However, the slope of these lines depends on the frustum size, such that larger values of  $\theta$  also result in complete binocular overlap at closer distances.

Lastly, it is notable that for one distance, the nasal-shifted frusta have the same  $\gamma$  and  $\gamma_b$  as the symmetric frusta (intersections of solid purple and light green lines in Figure 4.11A at 59 cm). When this happens, the symmetric frusta have produced divergent

overlap for the user, while the nasal-shifted frusta have produced convergent overlap. For temporal-shifted frusta, there is no distance at which the frusta have 100% overlap and no distance that has the same  $\gamma$  and  $\gamma_b$  as the symmetric frusta.

The provided equations can be used to determine and customize the cyclopean FOV and binocular overlap region for different AR systems and use cases (e.g., near viewing versus far viewing). For example, a system could be optimized to have 90% convergent binocular overlap at a working distance of 1 m using a nasal-shifted design, and the implications for FOV at other distances could then be examined.

## 4.5 Discussion

The question of FOV coverage in stereoscopic AR systems is deceptively complex. All stereoscopic AR systems can have variable horizontal FOV coverage and variable binocular overlap, whether they are explicitly designed to be “partial overlap systems” or not. The variability emerges from the impact of fixation distance on the overlap in the FOV coverage of the two eyes. In situations where binocular overlap is partial, the implications for perceptual nonuniformity across the visual field must be considered. A geometric analysis of system design can predict the amount of FOV covered for an AR system at different distances, but cannot determine the visibility of these artifacts. By combining perceptual results with a geometric analysis, we hope to provide a new toolkit for optimizing the FOV of stereoscopic near-eye display systems for AR. Here, we discuss the implications of the perceptual study as well as how to incorporate these insights into designing AR FOV.

### 4.5.1 Convergent and Divergent Configurations

A key question that arises when considering FOV in stereoscopic AR is whether convergent or divergent partial overlap is preferable. Prior work using simple stimuli strongly suggested that convergent partial overlap produces fewer nonuniformity artifacts than divergent partial overlap [87, 90]. Our results using simulated AR, however, suggest that this finding does not generalize well to modern stereoscopic AR systems. In Experiment 1, our results showed no strong difference between the two configurations for AR stimuli. In Experiment 2, we asked participants to directly compare convergent and divergent configurations and we found that both simple and AR stimuli had better visual quality (i.e., a wider FOV with less fading of content) with a divergent configuration than a convergent configuration, and the difference was more pronounced as binocular region size increased.

For AR stimuli, our results suggest no need to prioritize convergent overlap to minimize nonuniformity artifacts per se. However, we are left with a challenge to understand why our results with simple shapes in Experiment 2 differ from the canon of prior work. One potential explanation is that we used larger binocular region sizes. For example, in one prior study that used a similar comparison paradigm [87], convergent configurations were found to have a more uniform FOV compared to divergent configurations 93.3% of the time. However, that study used stimuli with a much smaller binocular overlap region (15.6° monocular FOV, 7.8° monocular region, 7.8° binocular region). If we look at the trend in our data from Experiment 2 (Figure 4.7), we see that smaller binocular region

sizes were associated with a weaker preference for divergent configurations. Indeed, if we use the logistic regression model fit to these data and predict the probability that the convergent configurations was preferred for a binocular region of this size ( $7.8^\circ$ ), for simple stimuli we obtain 0.62 and for AR stimuli we obtain 0.79. That is, our model would indeed predict a convergent preference for this smaller binocular region. Importantly, the range of FOVs used in the current studies is more reflective of current AR systems than previous work.

While the different FOV regime might explain differences in our results from previous work, it does not explain why Experiments 1 and 2 suggested slightly different biases. One important difference between these paradigms is the timing. Experiment 1 involved a continuous response over 30 seconds, while viewing times for Experiment 2 were self-controlled and presumably much shorter. While Experiment 2 was designed to directly compare the two configurations, we suggest that the results from Experiment 1 might be more reflective of artifact visibility during typical use of AR systems, when the amount of overlap would not rapidly alternate unless someone was possibly alternating their fixation depth back and forth. In future work, it would be fruitful to examine artifact detection using more naturalistic paradigms, such as performing target detection or reading tasks in AR.

#### 4.5.2 Predicting Fading in 3D AR Experiences

There is an inherent trade off between the amount of binocular overlap and the cyclopean FOV for a fixed monocular FOV, as demonstrated by the geometric analysis. With the perceptual studies, we can now also add predicted nonuniformity artifacts across various fixation distances into the mix to better understand the perceptual implications of specific design decisions.

By way of example, we focus here on the AR stimulus from Experiment 1. We combine the data from the divergent and convergent configurations since they were highly similar. We can then fit the average proportion fading time ( $F$ ) data as a function of the proportion of each eye's FOV that was only monocularly visible as mentioned earlier in Section 3.2.4. Here, we use ( $p_m$ ) to denote this monocular proportion. We chose this variable because it was the strongest FOV-based predictor of the fading time. In order to extrapolate beyond the ratios included in the experiment, we assert that the amount of fading should be 0 when this variable is 0 (i.e., complete binocular overlap). When adding the point (0,0) to the fitting, a straight line can no longer capture the data. We found the best fitting non-linear function of the form  $b\sqrt{p_m}$ , where  $b$  denotes the only free parameter. Because  $p_m$  should not exceed 1, we obtain the following piece-wise non-linear function:

$$F = \begin{cases} 1.405\sqrt{p_m}, & 0 \leq p_m \leq 0.50 \\ 1, & p_m > 0.50. \end{cases} \quad (4.7)$$

This fit is plotted along with the average data in Figure 4.12A. The first case in this equation can be solved for  $p_m$  as a function of  $F$  to calculate largest acceptable ratio for a given threshold on fading:

$$p_m = \left( \frac{F}{1.405} \right)^2. \quad (4.8)$$

For example, if a fading proportion of 0.50 is the maximum acceptable level ( $F = 0.50$ ),

that would suggest the portion to each eye’s FOV that is monocularly-visible should be 0.13 or less.

Of course, it is unlikely for any 3D AR experience to have a fixed monocular proportion if users are looking around a scene at different depths. To add 3D viewing into consideration, we can combine this analysis with the geometric analysis. Specifically, we can apply Eq.4.7 to calculate the predicted amount of fading for an arbitrary frustum and display setup. For example, given the binocular region size and the total cyclopean FOV in Figure 4.11A for a monocular FOV of  $40^\circ$ , we can determine the monocular region size at various distances for the three types of frusta and input the ratio into Eq.4.7. The predicted proportion fading times as a function of fixation distance are shown in Figure 4.12B. The important thing to note is that across all distances the nasal-shifted frusta produce less fading than the temporal-shifted frusta (although they also produce a smaller cyclopean FOV). Recall that the nasal shift produces a mixture of convergent and divergent overlap. We also see that the symmetric frusta may have fewer artifacts over time than nasal-shifted frusta for use cases with a relatively far fixation distance (about 2 meters and farther). However, more perceptual data using stimuli with different FOV sizes are needed to validate the function plotted in Figure 4.12A. Validating and expanding on these guidelines with more diverse visual stimuli will be an important direction for future work. It is also important to keep in mind that a large monocular region not only induces nonuniformity artifacts, but also lacks stereo cues for depth. Indeed, a recent analysis of virtual reality screen placement and typical fixation distances during virtual reality tasks produced guidelines advising a small nasal shift so as to achieve near 100% binocular overlap and maximize stereo cues [4]. Given the current differences in monocular FOV in VR and AR systems and their different use cases, it is likely that different trade offs should be prioritized for these systems.

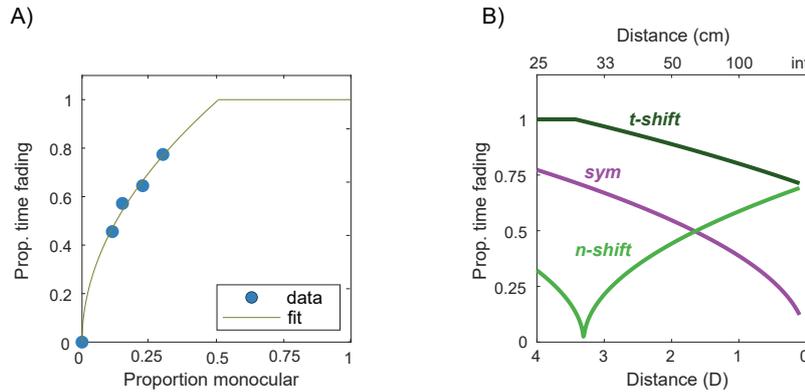


Figure 4.12: A) A piece-wise function was fit to the AR data from Experiment 1, including the (0,0) theoretical point. The x-axis is the ratio between monocular region and monocular FOV shown as a proportion. The y-axis is the predicted fading proportion and is capped at 1. B) Predictions for the proportion fading time across various fixation distances for the three types of camera frusta described in the geometric analysis (parameters are the same as in Figure 4.11A).

### 4.5.3 Monocular Regions in Natural Vision

At the core of these nonuniformity artifacts is the issue of AR systems that create monocularly-visible regions within the natural binocular FOV. However, we know that monocularly-visible regions are not inherently problematic. For example, the blind spots

in the two eyes create two monocularly-visible regions, which are not detected during most of daily life. In addition to the blind spots, monocularly-visible regions often result at occlusion boundaries in 3D scenes. However, these regions tend to serve as a source of depth information instead of causing nonuniformity artifacts [63], possibly because the size of naturally occurring monocularly-visible regions tend to be small compared to what is generated by partial overlap displays. These naturally-occurring monocular regions also have statistical regularities that may aid the visual system in detecting and integrating them into the natural binocular percept without producing interocular competition [16].

Indeed, previous work argued that convergent partial overlap configurations are better because they simulate a more natural occlusion geometry. With the convergent configuration, the location of the monocular regions simulates situations in which the viewer looks through an aperture, whereas the divergent configuration simulates a situation when the eyes are fixating on a near surface that occludes a background [9, 87, 90, 120]. Importantly, because the AR stimuli are spatially contiguous, for divergent stimuli this means the visual pattern on the near surface (binocular region) and the occluded background (monocular region) would match, which may be unlikely in natural scenes. However, it is worth remembering that the natural FOV is also divergent, with temporal flanking monocular regions in the far periphery[58]. Perhaps a deeper understanding of monocular regions in natural vision can yield new insights into how these features can be designed to mitigate artifacts in AR.

#### 4.5.4 Vertical Partial Overlap

In this work, we focused on the horizontal FOV and horizontal partial overlap. Of course, horizontal partial overlap has no effect on extending the vertical FOV. In principle, a vertical partial overlap could be created with one frustum/display shifted up and the other one down. However, unless there is 100% horizontal binocular overlap for all fixation distances, the resulting vertical monocular regions would be offset to the left and right as well. It seems unlikely that this strategy would produce a large increase in vertical FOV without introducing other issues, but small vertical offsets may be interesting to explore.

#### 4.5.5 Limitations and Future Work

While our stimuli made an effort to replicate modern AR use cases, there are still limitations that should be taken into consideration when translating our results into more complex viewing situations. For example, during real AR experiences, there may be different motion in the AR content and the natural background, which could influence the strength of nonuniformity artifacts. In addition, the presence of blur in the background due to different focal distances of the AR content and the background may affect the visibility of virtual content and detection of fading artifacts.

Indeed, integrating natural eye movements will be an important direction for future work. In the current perceptual experiments, the eyes were always fixating at the center of the binocular overlap region, but when users look directly at the monocular region there may be differences in how well they can detect the artifacts. In addition, changes in fixation distance will result in monocular regions that shrink and expand dynamically. These movements could suppress or enhance fading and warrant further investigation. If eye

movements make fading more visible, it would be interesting to consider integrating partial overlap with an actuated display and eye tracking that could maintain binocular overlap at the fixation point as users look around a scene. In addition, it has been speculated that the binocular region need not to be greater than  $40^\circ$  relative to fixation since binocular interactions are diminished beyond this eccentricity [58]. Our desk-mounted experiment setup could not achieve a wide enough FOV to test this prediction, but it could be explored with a virtual reality (VR) headset.

## 4.6 Conclusion

AR systems aim to create immersive mixtures of real and virtual content, but limited FOV coverage remains a persistent barrier to realizing this goal. In this work, we reviewed the importance of considering partial binocular overlap when analyzing FOV coverage in AR. We highlighted the need for a set of perceptual guidelines to optimize the horizontal FOV in AR systems, and conducted two perceptual studies to facilitate the development of these guidelines. Our results suggest that a large binocular overlap region and divergent configuration may reduce perceptual nonuniformity artifacts when viewing partial overlap imagery in AR, allowing more effective FOV expansion. We provide a model that can be used for optimizing FOV in AR headsets while taking into account the dynamic FOV during 3D AR viewing. By better understanding the factors that affect perceived FOV size and quality during dynamic 3D interactions, we hope to facilitate the development of systems with sufficient FOV coverage that FOV is no longer a limiting factor in AR experiences.

# Discussion

In this dissertation, I evaluated the perceptual appearance of dichoptic stimuli with various degrees of complexity. In many cases, I used stimuli intended to emulate viewing scenarios that have important implications for stereoscopic display design. As mentioned in Chapter 2, 3, and 4, there are many technical constraints that limit the ability of emerging displays to provide a good visual experience and integrate well with natural vision. While simple perceptual guidelines for display practitioners were provided based on the experimental results, our holistic understanding of binocular integration during natural vision remains incomplete. This limitation is especially highlighted by the work covered in Chapter 1 that showed stark differences in the perceptual appearance associated with different stimulus patterns. I addressed this limitation in Chapters 2, 3, and 4 by adopting stimuli that emulate the appearance of each target application to generate display guidelines. The long-term goal, however, should be an image-computable model of human binocular vision that is able to take an arbitrary pair of dichoptic images as inputs and predict the perceptual outcome. Such image-computable models of human vision are rather rare, let alone one that focuses on binocular vision. In the following discussion, I elaborate on some of the key ideas that arose from the main chapters that I think warrant further research to better predict binocular appearance.

## 4.6.1 Binocular Contrast Combination Depends on Spatial Structure

The spatial structure of a stimulus can greatly influence the perceptual outcome of binocular contrast combination. For example, one factor that modulates the appearance of dichoptic stimuli is the presence of visual contours. In Chapter 1, I showed that the visual system is influenced by contour information differently for simple and complex stimuli. However, these results do not tell the whole story about how different types and intensities of contours affect binocular vision. Future research should explore the effect of contour information and how different types of contour cues interact. For example, Figure 4.13 shows the original stimuli (top pair) used in the Chapter 1 experiments and modified versions (bottom pair). The original stimuli resulted in 'loser-take-all' binocular combination for the grating, but 'winner-take-all' for the noise. If we add an additional line contour (black lines in the bottom pairs), does that drive binocular combination in the opposite direction for these stimuli? Anecdotally, adding a black contour line did not make the noise pattern completely behave like the grating or vice versa. However, it is worth pointing out that the binocular interaction is nonetheless altered by the manipulation, since personal observation has noted more rivalry with the monocular black line contour, which is in agreement with Chapter 2 that monocular features tend to elicit rivalry.

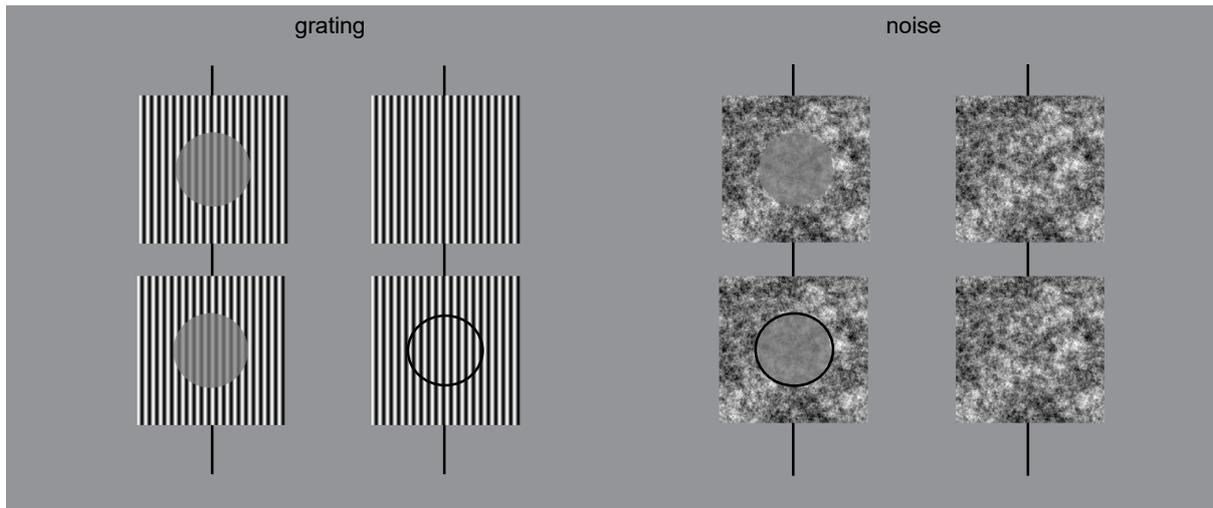


Figure 4.13: Examples of the original low surround grating and noise stimuli, along with a version that include an outline contour to drive binocular appearance.

Why might contour information bias binocular combination? Considering that the visual system is tasked with parsing objects in the complex visual world, it is not surprising that the visual system is highly sensitive to contours as they usually signal object boundaries. Importantly, many sources of information can lead to the perception of contour as shown by Figure 4.14, and the visual system is likely to utilize all these sources to construct object boundaries. I speculate that this process may be thought of as a cue combination problem, where different cues informing about an object boundary are integrated together. To quantify contour information, one potential area to explore is leverage work done with camouflage detection, which in essence is about how the brain determines object segregation. For example, unpublished work has shown that computing contour strength using local gradient methods can predict human performance on detection of  $1/f$  noise patterns embedded in another  $1/f$  noise pattern [33]. However, the computation of contour is not only a local process. Instead, the brain can fill-in missing contour information based on surrounding context (i.e., illusory contour) [6, 175]. The contour information, able to override image contrast when it signifies an object boundary, should improve the prediction of each eye's contribution in binocular contrast combination.

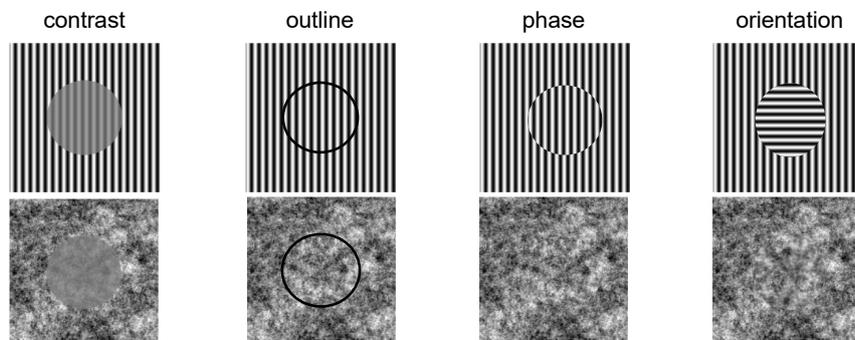


Figure 4.14: Different types of contours created by: contrast difference, outline, phase difference, and orientation difference.

## 4.6.2 Predicting Secondary Perceptual Effects

The possibility of multiple perceptual effects occurring during the viewing of dichoptic stimuli has great implications for developing rendering algorithms and hardware for displays. For example, in an image-computable model of binocular combination, it would be useful to also generate a predicted likelihood map for each perceptual effect based on interocular difference and visualize the hot spots where perceptual effects are likely. As an initial step, it may be worthwhile to treat the different perceptual effects as separate processes, and provide a model prediction for each effect separately via different computations. We can check whether these effects are independent from each other or not by doing an additional post hoc analysis on the data from Chapter 2. If the occurrence of two effects, A and B, are independent, then the following is true:

$$P(A \cap B) = P(A) * P(B).$$

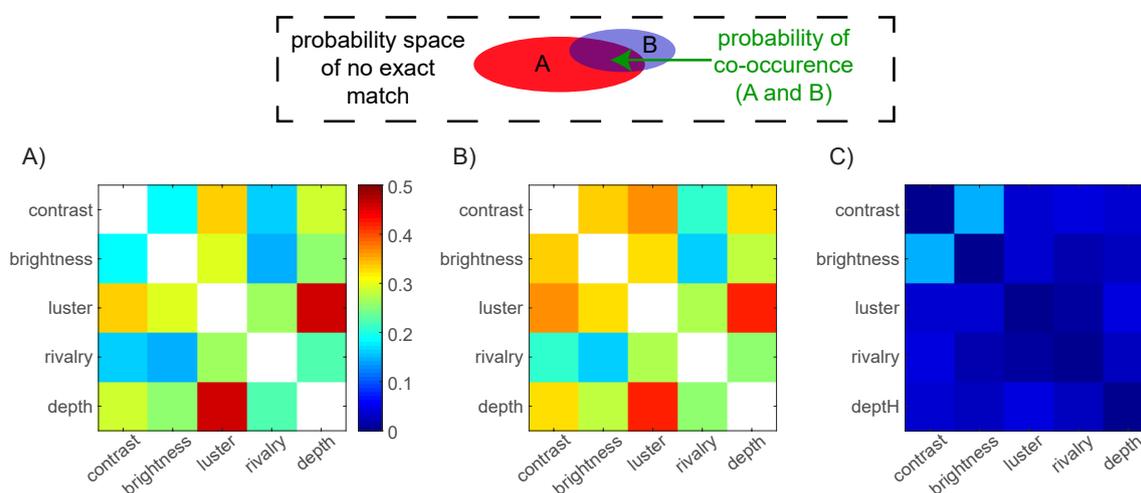


Figure 4.15: Post hoc analysis for Chapter 2 evaluating the independence of effect occurrence. The heatmap color shows: A) the predicted co-occurrence of effect A ( $x$  label) and B ( $y$  label) if A and B were independent events, B) the actual observed co-occurrence of A and B, and C) the absolute value of the difference between Panel A and B. A larger difference indicates a dependence of the effects.

Figure 4.15 shows the comparison between the predicted co-occurrence of two perceptual effects and the actual observed co-occurrence. The result suggests that only the brightness and contrast effects are deviating substantially from being independent from each other. This means that for these two perceptual effects, the likelihood of co-occurrence exceeds the expectations based on independence. However, for any other combination of perceptual effects, the likelihood of one does not notably correlated with another. This observation suggests that if we want to know whether a particular stimulus will be associated with a particular perceptual effect, we need to measure them separately since we cannot predict one effect better than chance by knowing the occurrence of the other.

## 4.6.3 Eye Dominance

One thing that is often asked by display designers is whether or not the observer would perceive better image quality if good quality image is only shown to the dominant eye. The notion of eye dominance comes up often since it can lead to a clear display design

decision. For example a monocular display may be designed to fit over the dominant eye, or a binocular display may use a simple calibration test to determine the dominant eye and prioritize image quality for that eye. In addition, if eye dominance plays a significant role in dichoptic appearance, then it would be worth incorporating into binocular appearance models to account for individual differences.

There are multiple ways of defining and measuring eye dominance. In this dissertation, I discussed two measures for eye dominance. One measure was derived post hoc based on the matching task result (Chapter 1) and the other one was a common measure based on rivalry (sensory) dominance (Chapter 3). In Chapter 3, rivalry eye dominance did not reveal any consistent difference for the subjective preference of dichoptic tonemapping imagery. In Chapter 1, I found some eye dominance effects for the contrast matching result. To further look into the effect of eye dominance for AR display, I ran additional post hoc analyses for the AR stimuli in Chapter 2 (Figure 4.16A). However, this analysis suggests no meaningful differences when the images are swapped between the two eyes on average. That is, the matching result was very similar when the dominant eye saw the higher contrast target (left panel) and when the non-dominant eye saw it (right panel).

We can define eye dominance based on other criteria as well. For example, there may be fewer perceptual effects like luster and rivalry when the dominant eye sees the higher contrast target (i.e., the dominant eye leads to better fusion). Figure 4.16B shows the absolute value for the change in frequency of the perceptual responses when the stimulus images were swapped between the two eyes. If which eye sees the high contrast image does not matter, then the amount of change would be close to zero, whereas a large amount of change would indicate an effect of eye swap. This calculation was done separately for each perceptual judgement prompt. The dominant eye status based on the six prompts was not consistent within each observer. For example, a participant could have five out of six of these measures indicating that there was a lower chance of having a no match, contrast, brightness, luster, and depth effects when the left eye saw the higher contrast target, but a lower chance of rivalry when the right eye saw the higher contrast. This implies that the contribution of each eye might be different for each effect; some effects are more likely to occur when one eye sees the higher contrast image but other effects are more likely with the other eye seeing the higher contrast image. This suggests potentially different neural processes underlying these effects. It is worth pointing out that while most people did not show any significant eye dominance effect, there were two participants who were clearly the outliers based on the rivalry measure; they show a large eye effect that is likely to be reliable. These exploratory results should be assessed further with dedicated studies because for these observers, switching which eye saw the higher contrast image makes a difference for their ability to maintain a stable perception when viewing dichoptic imagery. While the frequency of people like them are low (2 out of 31) in this sample, the impact will be scaled as binocular displays reach a larger population of consumers.

Current models of binocular combination extract stimulus properties based on the physical image rather than the retinal image. However, it is important to keep in mind that the human eye is far from a perfect imaging system, thus the stimulus is not a good representation of what the visual system actually processes. As we have seen in Chapter 1, presenting the higher contrast stimuli to the dominant eye versus the non-dominant eye did not reveal a difference for most stimuli, but did show a difference for the 5cpd grating, and somewhat for the bandpassed noise as well. This leads me to speculate that

each spatial frequency may have a different binocular imbalance pattern, and certain stimuli are better at revealing eye dominance effects than others. The spatial frequency dependency has been found in people with amblyopia, with more imbalance for higher spatial frequencies [38, 41], but it is not well understood for people with normal vision. It is possible that there is some relationship between the eye dominance pattern and the eye's optics [30]. Previous literature has shown that pupil size and optical aberration can change the contrast sensitivity function in complex ways [151]. Current theory of binocular appearance is that the visual system is evaluating the amount of contrast in each eye and prioritize the higher contrast eye. If optical blur would systematically reduce image contrast in one eye, the binocular system should adapt by suppressing the effect of such blur to maintain good visual fidelity. However, the suppression may not be global as in the case of a complete suppression of one eye (i.e., amblyopia), but instead it is specific to the impact of the blur on different spatial frequency channels.

Adaptive optics could be a useful tool to tease apart the effect of retinal image quality and neural mechanism on binocular interaction by manipulating aberration in one eye while performing psychophysics measurements and qualitative measures to see how the manipulation affects binocular combination. For example, for the two outliers in Figure 4.16B by the rivalry measure, we can do two types of optical correction. Let us consider that both of them perceived more rivalry when the left eye was shown the higher contrast image than when the right eye saw the higher contrast image. Then, by definition, we call their right eye their dominant eye since it exerted more complete suppression on the left eye which leads to perceptual stability and less rivalry. If we were to correct their right eye optically, we might expect to see less rivalry because the suppression is enhanced, which could lead to fusion with the corrected eye dominating. If we were to correct their left eye, we might expect to observe more rivalry since we are increasing the amount of suppression that the left eye exert on the right eye. In this case, there may also be orientation specificity of each eye's contribution to the binocular percept if the observer has non-spherical aberrations. With future displays that could correct people's optical aberrations (perfectly or partially), it is worth examining the short term and long term perceptual effects and consequences of these corrections. An additional consideration about retinal image quality is that optical aberration is often corrected to be minimal at the fovea, but will increase dramatically away from the fovea due to a combination of the eye's aberrations and the shape of the eyeball [144]. The periphery would be interesting to probe for two reasons. First, there is greater optical aberration that affects the retinal image considerably. Second, the periphery possesses different neural processing which may reduce the influence of retinal image quality for binocular combination. I will briefly discuss binocular combination at larger eccentricities next.

#### 4.6.4 Binocular Combination at Large Eccentricities

Chapter 4 provide some insight about rivalry and suppression effects at larger eccentricities. In the fovea where the two fovea have roughly equal contribution, suppression observed during dichoptic viewing is mainly attributed to the image properties of the stimulus (i.e., contrast). On the other hand, there may be anatomical differences that bias each eye's contribution to binocular appearance in the periphery. The results hint at two potential differences in binocular combination away from the fovea. First, in Experiment 1, stimuli with binocular region size of  $25^\circ$  showed a deviation from the overall monotonic trend. Note that with this binocular region size, the monocular-binocular

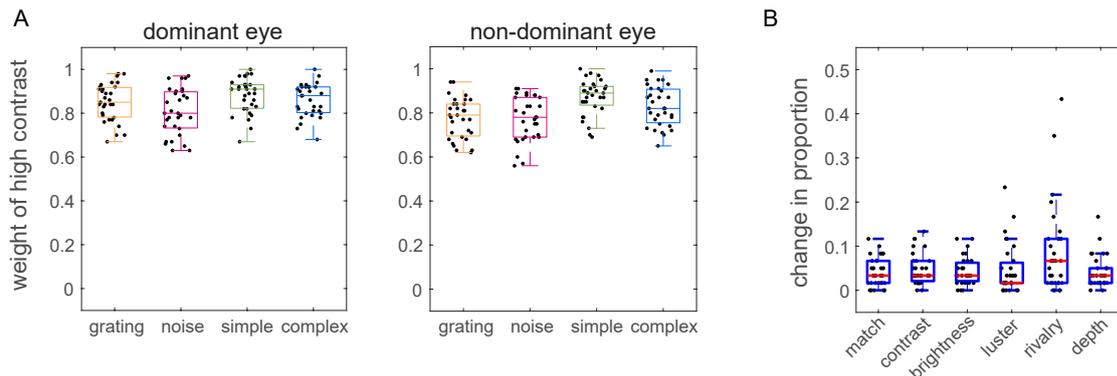


Figure 4.16: The boxplots show the median, 25<sup>th</sup> and 75<sup>th</sup> percentile, and the non-outlier range. The black dots represent each individual subject. A) The two panels show a comparison between the higher contrast image was shown to the dominant or non-dominant eye for the AR stimuli in Chapter 2, Experiment 2. The matching result is expressed as a weight for the higher contrast image similar to Chapter 1, Figure 3.13. B) The absolute value of change in the proportion of time that people responded to each prompt (x-axis) with image swap between eyes. A larger change would indicate more eye dominance effect.

border happens to situate around our natural blind spot in one eye (12° to 15° nasally away from the fovea). The result suggests that there are unique patterns of binocular interaction occurring around the blind spot. The reduction in rivalry detection (fading) for this stimulus may be explained by the fact that the unmatched border is in the blind spot and the corresponding location in the other eye suppresses it. The suppression was not complete and some fading was still detected since my stimulus was not restricted only to the blind spot (i.e., the border was 30° tall in the stimulus). It is also possible that the brain filled-in the border in the blind spot based on surrounding information around it which can still exert some suppression to the other eye [135, 153].

Second, Experiment 2 showed an effect of retinal location when directly comparing the perceived quality of stimulus presented to the nasal retina versus temporal retina. Content shown to the nasal retina was better at suppressing the competing image on the temporal retina in the other eye (divergent configuration). Importantly, this effect was larger at larger eccentricities (i.e., stimuli that have larger binocular region and their monocular region farther out in the periphery). Future work could explore how the visual system weights binocular information across the retina more directly by using a matching paradigm like the one in Chapter 1. When we are looking with both eyes, we do not notice that the temporal regions of our cyclopean view are monocular. The monocular regions, which stimulate the nasal retina, do not give a sensation of reduced brightness compared to the binocular region. My hypothesis is that the nasal retina contributes more to the cyclopean appearance away from the fovea gradually, rather than a dichotomy with a sharp transition when reaching the monocular temporal visual fields. This knowledge would be useful to incorporate into models that predict binocular combination across a larger field of view since it could be potentially leveraged for display design (i.e., nasal retina could be prioritized over temporal retina in addition to foveated rendering). In addition, understanding binocular interactions of different retinal locations also has important clinical relevance for people who do not have typical retinal correspondence. For example, someone with strabismus might have their fovea competing with the nasal or temporal retina depending on the type of strabismus [48].

## 4.6.5 Conclusion

In this dissertation, I used new stimuli that are more representative of real viewing scenarios and a new behavioral response paradigm to probe the binocular appearance of complex imagery. I also discussed the implications of binocular appearance for display design. I hope the work presented herein can help guide improvements for next-generation displays and support the development of better models of binocular appearance that are able to account for different eyes, different vision, and different content to create inclusive visual technology for everyone.

# Bibliography

- [1] Freepik mobile app icon images, 2019. <https://www.freepik.com/> [Accessed: April 19, 2019)].
- [2] ADAMS, W. J., ELDER, J. H., GRAF, E. W., LEYLAND, J., LUGTIGHEID, A. J., AND MURYY, A. The Southampton-York natural scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports* 6 (2016), 35805.
- [3] ADELSON, E. H. Perceptual organization and the judgment of brightness. *Science* 262, 5142 (1993), 2042–2044.
- [4] AIZENMAN, A. M., KOULIERIS, G. A., GIBALDI, A., SEHGAL, V., LEVI, D. M., AND BANKS, M. S. The statistics of eye movements and binocular disparities during VR gaming: Implications for Headset Design. *ACM Trans. Graph.* 42, 1 (2023).
- [5] ALAM, M. S., ZHENG, S. H., IFTEKHARUDDIN, K., AND KARIM, M. A. Study of field-of-view overlap for night vision applications. In *Proceedings of the IEEE 1992 National Aerospace and Electronics Conference* (1992), vol. 3, IEEE, pp. 1249–1255.
- [6] ALBERT, M. K. Cue interactions, border ownership and illusory contours. *Vision Research* 41, 22 (2001), 2827–2834.
- [7] ALBERTI, C. F., AND BEX, P. J. Binocular contrast summation and inhibition depends on spatial frequency, eccentricity and binocular disparity. *Ophthalmic and Physiological Optics* 38, 5 (2018), 525–537.
- [8] AN, J., LEE, S. H., KUK, J. G., AND CHO, N. I. A multi-exposure image fusion algorithm without ghost effect. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing* (2011), pp. 1565–1568.
- [9] ANDERSON, B. L., AND NAKAYAMA, K. Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psychological Review* 101, 3 (1994), 414–445.
- [10] BAKER, D. H., MEESE, T. S., AND GEORGESON, M. A. Binocular interaction: Contrast matching and contrast discrimination are predicted by the same model. *Spatial vision* 20, 5 (2007), 397–413.
- [11] BAKER, D. H., MEESE, T. S., AND SUMMERS, R. J. Psychophysical evidence for two routes to suppression before binocular summation of signals in human vision. *Neuroscience* 146, 1 (2007), 435–448.

- [12] BAKER, D. H., WALLIS, S. A., GEORGESON, M. A., AND MEESE, T. S. Nonlinearities in the binocular combination of luminance and contrast. *Vision Research* 56 (2012), 1–9.
- [13] BANKS, M. S., HOFFMAN, D. M., KIM, J., AND WETZSTEIN, G. 3D displays. *Annual Review of Vision Science* 2, 1 (2016), 397–435.
- [14] BARBONI, M. T. S., MANESCHG, O. A., NÉMETH, J., NAGY, Z. Z., VIDNYÁNSZKY, Z., AND BANKÓ, M. Dichoptic spatial contrast sensitivity reflects binocular balance in normal and stereoanomalous subjects. *Investigative Ophthalmology & Visual Science* 61, 11 (2020), 23–23.
- [15] BAŞGÖZE, Z., MACKAY, A. P., AND COOPER, E. A. Plasticity and adaptation in adult binocular vision. *Current Biology* 28, 24 (2018), R1406–R1413.
- [16] BAŞGÖZE, Z., WHITE, D. N., BURGE, J., AND COOPER, E. A. Natural statistics of depth edges modulate perceptual stability. *Journal of Vision* 20, 8 (2020), 10.
- [17] BEARSE, M. A., AND FREEMAN, R. D. Binocular summation in orientation discrimination depends on stimulus contrast and duration. *Vision Research* 34, 1 (1994), 19–29.
- [18] BENJAMINI, Y., AND HOCHBERG, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300.
- [19] BLAKE, R. A primer on binocular rivalry, including current controversies. *Brain and Mind* 2 (2001), 5–38.
- [20] BLAKE, R., AND BOOTHROYD, K. The precedence of binocular fusion over binocular rivalry. *Perception & Psychophysics* 37 (1985), 114–124.
- [21] BLAKE, R., AND FOX, R. The psychophysical inquiry into binocular summation. *Perception & Psychophysics* 14 (1973), 161–185.
- [22] BLAKE, R., AND WILSON, H. Binocular vision. *Vision Research* 51, 7 (2011), 754–770.
- [23] BORISOV, V. N., MURAVYEV, N. V., OKUN, R. A., ANGERVAKS, A. E., VOSTRIKOV, G. N., AND POPOV, M. V. A doe-based waveguide architecture of wide field of view display for augmented reality eyewear. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) II* (2021), vol. 11765, SPIE, pp. 94 – 102.
- [24] BRADSKI, G. The opencv library. *Dr. Dobb’s Journal of Software Tools* (2000).
- [25] CADÍK, M., WIMMER, M., NEUMANN, L., AND ARTUSI, A. Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics* 32, 3 (2008), 330–349.
- [26] ÇAKMAKCI, O., HOFFMAN, D. M., AND BALRAM, N. 31-4: Invited paper: 3D eyebox in augmented and virtual reality optics. *SID Symposium Digest of Technical Papers* 50, 1 (2019), 438–441.

- [27] CAMPBELL, F. W., AND GREEN, D. G. Monocular versus binocular visual acuity. *Nature* 208 (1965), 191–192.
- [28] CAO, R., PASTUKHOV, A., ALESHIN, S., MATTIA, M., AND BRAUN, J. Binocular rivalry reveals an out-of-equilibrium neural dynamics suited for decision-making. *eLife* 10 (2021), e61581.
- [29] CHEN, Y., JIANG, G., YU, M., YANG, Y., AND HO, Y.-S. Learning stereo high dynamic range imaging from a pair of cameras with different exposure parameters. *IEEE Transactions on Computational Imaging* 6 (2020), 1044–1058.
- [30] CHENG, C.-Y., YEN, M.-Y., LIN, H.-Y., HSIA, W.-W., AND HSU, W.-M. Association of ocular dominance and anisometropic myopia. *Investigative Ophthalmology & Visual Science* 45, 8 (2004), 2856–2860.
- [31] CHOLEWIAK, S. A., BAŞGÖZE, Z., ÇAKMAKCI, O., HOFFMAN, D. M., AND COOPER, E. A. A perceptual eyebox for near-eye displays. *Optics Express* 28, 25 (2020), 38008–38028.
- [32] CHUBB, C., SPERLING, G., AND SOLOMON, J. A. Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences* 86, 23 (1989), 9631–9635.
- [33] DAS, A. *Camouflage Detection & Signal Discrimination: Theory, Methods & Experiments*. PhD thesis, 2022.
- [34] DE WEERT, C. M., AND LEVELT, W. J. M. Binocular brightness combinations: Additive and nonadditive aspects. *Perception & Psychophysics* 15, 15 (1974), 551–562.
- [35] DEBEVEC, P. E., AND MALIK, J. Recovering high dynamic range radiance maps from photographs. SIGGRAPH '97, pp. 369–378.
- [36] DEHOOG, E., HOLMSTEDT, J., AND AYE, T. Field of view limitations in see-through HMD using geometric waveguides. *Applied Optics* 55, 22 (2016), 5924–5930.
- [37] DICARLO, J. M., AND WANDELL, B. A. Rendering high dynamic range images. In *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications* (2000), vol. 3965, pp. 392–401.
- [38] DING, J., KLEIN, S. A., AND LEVI, D. M. Binocular combination in abnormal binocular vision. *Journal of Vision* 13, 2 (2013), 14–14.
- [39] DING, J., KLEIN, S. A., AND LEVI, D. M. Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model. *Journal of Vision* 13, 2 (2013), 13–13.
- [40] DING, J., AND LEVI, D. M. Rebalancing binocular vision in amblyopia. *Ophthalmic and Physiological Optics* 34, 2 (2014), 199–213.
- [41] DING, J., AND LEVI, D. M. Rebalancing binocular vision in amblyopia. *Ophthalmic and Physiological Optics* 34 (2014), 199 – 213.
- [42] DING, J., AND LEVI, D. M. Binocular combination of luminance profiles. *Journal of Vision* 17, 13 (2017), 4–4.

- [43] DING, J., AND SPERLING, G. A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America* 103, 4 (2006), 1141–1146.
- [44] DRAGO, F., MARTENS, W. L., MYSZKOWSKI, K., AND SEIDEL, H.-P. Perceptual evaluation of tone mapping operators. In *ACM SIGGRAPH 2003 Sketches & Applications* (2003), SIGGRAPH '03, p. 1.
- [45] DURAND, F., AND DORSEY, J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics* 21, 3 (2002), 257–266.
- [46] DURGIN, F. H., TRIPATHY, S. P., AND LEVI, D. M. On the filling in of the visual blind spot: Some rules of thumb. *Perception* 24, 7 (1995), 827–840.
- [47] E. JAMIY, F., AND MARSH, R. Survey on depth perception in head mounted displays: Distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing* 13, 5 (2019), 707–712.
- [48] ECONOMIDES, J. R., ADAMS, D. L., AND HORTON, J. C. Perception via the deviated eye in strabismus. *Journal of Neuroscience* 32, 30 (2012), 10286–10295.
- [49] ELLIS, S. R., ADELSTEIN, B. D., REISMAN, R. J., SCHMIDT-OTT, J. R., GIPS, J., KROZEL, J., AND COHEN, M. Augmented reality in a simulated tower environment: Effect of field of view on aircraft detection. *NASA/TM-2002-211853* (2002).
- [50] FATTAL, R., LISCHINSKI, D., AND WERMAN, M. Gradient domain high dynamic range compression. *ACM Transactions on Graphics* 21, 3 (2002), 249–256.
- [51] FENG, M., AND LOEW, M. H. Video-level binocular tone-mapping framework based on temporal coherency algorithm. In *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* (2017), pp. 1–5.
- [52] FIELD, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 12 (1987), 2379–2394.
- [53] FORMANKIEWICZ, M. A., AND MOLLON, J. D. The psychophysics of detecting binocular discrepancies of luminance. *Vision Research* 49, 15 (2009), 1929–1938.
- [54] FRY, G. A., AND BARTLEY, S. H. The brilliance of an object seen binocularly. *American Journal of Ophthalmology* 16, 8 (1933), 687–693.
- [55] GEORGESON, M. A., WALLIS, S. A., MEESE, T. S., AND BAKER, D. H. Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Research* 129 (2016), 98–118.
- [56] GILCHRIST, J., AND MCIVER, C. Fechner’s paradox in binocular contrast sensitivity. *Vision Research* 25, 4 (1985), 609–613.
- [57] GRAHAM, D., SCHWARZ, B., CHATTERJEE, A., AND LEDER, H. Preference for luminance histogram regularities in natural scenes. *Vision Research* 120 (2016), 11–21.

- [58] GRIGSBY, S. S., AND TSOU, B. H. Visual processing and partial-overlap head-mounted displays. *Journal of the Society for Information Display* 2, 2 (1994), 69–74.
- [59] GUNDLACH, B. S., FRISING, M., SHAHSAFI, A., VERSHBOW, G., WAN, C., SALMAN, J., ROKERS, B., LESSARD, L., AND KATS, M. A. Design considerations for the enhancement of human color vision by breaking binocular redundancy. *Scientific Reports* 8 (2018), 11971.
- [60] HAN, C., HE, Z. J., AND OOI, T. L. On sensory eye dominance revealed by binocular integrative and binocular competitive stimuli. *Investigative Ophthalmology & Visual Science* 59 (2018), 5140–5148.
- [61] HAN, C., HUANG, W., HE, Z. J., AND OOI, T. L. The role of boundary contours in suprathreshold binocular perception of contrast and spatial phase. *Journal of Vision* 19, 10 (2019), 223–223.
- [62] HAN, C., HUANG, W., SU, Y. R., HE, Z. J., AND OOI, T. L. Evidence in support of the border-ownership neurons for representing textured figures. *iScience* 23, 8 (2020), 101394.
- [63] HARRIS, J. M., AND WILCOX, L. M. The role of monocularly visible regions in depth and surface perception. *Vision Research* 49, 22 (2009), 2666–2685.
- [64] HASINOFF, S. W., SHARLET, D., GEISS, R., ADAMS, A., BARRON, J. T., KAINZ, F., CHEN, J., AND LEVOY, M. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics* 35, 6 (2016).
- [65] HASSANI, N., AND MURDOCH, M. J. Investigating color appearance in optical see-through augmented reality. *Color Research & Application* 44, 4 (2019), 492–507.
- [66] HAUN, A. M., AND PELI, E. Perceived contrast in complex images. *Journal of Vision* 13, 13 (2013), 3–3.
- [67] HELD, R. T., AND BANKS, M. S. Misperceptions in stereoscopic displays: A vision science perspective. *ACM Transactions on Graphics* 2008 (2008), 23–32.
- [68] HIBBARD, P. B., AND ASHER, J. M. Robust natural depth for anticorrelated random dot stereogram for edge stimuli, but minimal reversed depth for embedded circular stimuli, irrespective of eccentricity. *PLoS ONE* 17, 9 (2022), 1–21.
- [69] HIBBARD, P. B., SCOTT-BROWN, K. C., HAIGH, E. C., AND ADRAIN, M. Depth perception not found in human observers for static or dynamic anti-correlated random dot stereograms. *PLoS ONE* 9 (2014), 1–9.
- [70] HOEFFLINGER, B. *The Eye and High-Dynamic-Range Vision*, vol. 26 of *Advanced Microelectronics*. Springer, Berlin, Heidelberg, 2007.
- [71] HOHWY, J., ROEPSTORFF, A., AND FRISTON, K. J. Predictive coding explains binocular rivalry: An epistemological review. *Cognition* 108 (2008), 687–701.
- [72] HOU, Q., WANG, Q., CHENG, D., AND WANG, Y. Geometrical waveguide in see-through head-mounted display: a review. In *Optical Design and Testing VII* (2016), vol. 10021, SPIE, pp. 52–59.

- [73] HOWARD, I. P. Depth from binocular rivalry without spatial disparity. *Perception* 24 (1995), 67 – 74.
- [74] HUANG, C.-B., ZHOU, J., ZHOU, Y., AND LU, Z.-L. Contrast and phase combination in binocular vision. *PLOS ONE* 5, 12 (12 2010), 1–6.
- [75] HUBEL, D. H., AND WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology* 160 (1962).
- [76] JENNINGS, B. J., AND KINGDOM, F. A. A. Unmasking the dichoptic mask. *Journal of Vision* 19, 6 (2019), 3–3.
- [77] JENNINGS, S., AND DION, M. An investigation of helmet-mounted display field-of-view and overlap tradeoffs. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (1997), vol. 41, pp. 32–36.
- [78] JULESZ, B. Binocular depth perception without familiarity cues. *Science* 145, 3630 (1964), 356–362.
- [79] KINGDOM, F. A., AND WANG, D. Dichoptic colour-saturation masking is unmasked by binocular luminance contrast. *Vision Research* 116 (2015), 45–52.
- [80] KINGDOM, F. A., AND WOESSNER, P. Binocular summation and efficient coding. *Vision Research* 179 (2021), 53–63.
- [81] KINGDOM, F. A. A., AND LIBENSON, L. Dichoptic color saturation mixture: binocular luminance contrast promotes perceptual averaging. *Journal of Vision* 15, 5 (2015), 2–2.
- [82] KINGDOM, F. A. A., AND LIBENSON, L. Dichoptic colour-saturation masking is unmasked by binocular luminance contrast. *Vision Research* 116 (2015), 45 – 52.
- [83] KLEINER, M., BRAINARD, D., PELLI, D., INGLING, A., MURRAY, R., AND BROUSSARD, C. What’s new in psychtoolbox-3. *Perception* 36, 14 (2007), 1–16.
- [84] KLYMENKO, V., HARDING, T. H., BEASLEY, H. H., MARTIN, J. S., AND RASH, C. E. The effect of helmet mounted display field-of-view configurations on target acquisition. *USAARL Report No. 99-19* (1999).
- [85] KLYMENKO, V., HARDING, T. H., BEASLEY, H. H., AND RASH, C. E. Visual search performance in hmds with partial overlapped binocular fields-of-view. *USAARL Report No. 2001-05* (2001).
- [86] KLYMENKO, V., VERONA, R. W., BEASLEY, H. H., AND MARTIN, J. S. Convergent and divergent viewing affect luning, visual thresholds, and field-of-view fragmentation in partial binocular overlap helmet-mounted displays. In *Helmet-and Head-Mounted Displays and Symbology Design Requirements* (1994), vol. 2218, SPIE, pp. 82–96.
- [87] KLYMENKO, V., VERONA, R. W., BEASLEY, H. H., MARTIN, J. S., AND MCLEAN, W. E. Factors affecting the visual fragmentation of the field-of-view in partial binocular overlap displays. *USAARL Report No. 94-29* (1994).

- [88] KLYMENKO, V., VERONA, R. W., BEASLEY, H. H., MARTIN, J. S., AND MCLEAN, W. E. Visual perception in the field-of-view of partial binocular overlap helmet-mounted displays. *USAARL Report No. 94-40* (1994).
- [89] KLYMENKO, V., VERONA, R. W., MARTIN, J. S., BEASLEY, H. H., AND MCLEAN, W. E. The effect of binocular overlap mode on contrast thresholds across the field-of-view as a function of spatial and temporal frequency. *USAARL Report No. 94-49* (1994).
- [90] KLYMENKO, V., VERONA, R. W., MARTIN, J. S., BEASLEY, H. H., AND MCLEAN, W. E. Factors affecting the perception of luning in monocular regions of partial binocular overlap displays. *USAARL Report No. 94-47* (1994).
- [91] KOLB, H. Facts and figures concerning the human retina. In *Webvision: The Organization of the Retina and Visual System*. 2007.
- [92] KOOI, F. L., AND TOET, A. Visual comfort of binocular and 3D displays. *Displays* 25, 2 (2004), 99–108.
- [93] KOULIERIS, G. A., AKŞIT, K., STENGEL, M., MANTIUK, R. K., MANIA, K., AND RICHARDT, C. Near-eye display and tracking technologies for virtual and augmented reality. *Computer Graphics Forum* 38, 2 (2019), 493–519.
- [94] KRESS, B. C. Digital optical elements and technologies (EDO19): Applications to AR/VR/MR. In *Digital Optical Technologies 2019* (2019), vol. 11062, pp. 343 – 355.
- [95] KRESS, B. C. Digital optical elements and technologies (edo19): applications to ar/vr/mr. In *Digital Optical Technologies 2019* (2019), vol. 11062, SPIE, pp. 343 – 355.
- [96] KRESS, B. C. Optical Waveguide Combiners for AR Headsets: Features and Limitations. In *Digital Optical Technologies 2019* (2019), vol. 11062, SPIE, pp. 75 – 100.
- [97] KRESS, B. C., AND CHATTERJEE, I. Waveguide combiners for mixed reality headsets: a nanophotonics design perspective. *Nanophotonics* 10, 1 (2021), 41–74.
- [98] KUANG, J., YAMAGUCHI, H., LIU, C., JOHNSON, G. M., AND FAIRCHILD, M. D. Evaluating hdr rendering algorithms. *ACM Transactions on Applied Perception* 4, 2 (2007), 9–36.
- [99] LE VAY, S., WIESEL, T. N., AND HUBEL, D. H. The development of ocular dominance columns in normal and visually deprived monkeys. *Journal of Comparative Neurology* 191, 1 (1980), 1–51.
- [100] LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. Evaluation of tone mapping operators using a high dynamic range display. *ACM Transactions on Graphics* 24, 3 (2005), 640–648.
- [101] LEE, C.-K., GI PARK, S., MOON, S., HONG, J.-Y., AND LEE, B. Compact multi-projection 3D display system with light-guide projection. *Optical Express* 23, 22 (2015), 28945–28959.

- [102] LEGGE, G. Binocular contrast summation —i. detection and discrimination. *Vision Research* 24, 4 (1984), 373–83.
- [103] LEGGE, G. E., AND RUBIN, G. S. Binocular interactions in suprathreshold contrast perception. *Perception & Psychophysics* 30, 1 (1981), 49–61.
- [104] LEVELT, W. J. Binocular brightness averaging and contour information. *British Journal of Psychology* 56, 1 (1965), 1–13.
- [105] LEVELT, W. J. On binocular rivalry. 1–118.
- [106] LI, H.-H., CARRASCO, M., AND HEEGER, D. J. Deconstructing interocular suppression: Attention and divisive normalization. *PLoS Computational Biology* 11, 10 (2015), 1–26.
- [107] LI, S., AND KANG, X. Fast multi-exposure image fusion with median filter and recursive filter. *IEEE Transactions on Consumer Electronics* 58, 2 (2012), 626–632.
- [108] LIN, J. J., DUH, H. B., PARKER, D. E., ABI-RACHED, H., AND FURNESS, T. A. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings IEEE Virtual Reality 2002* (2002), pp. 164–171.
- [109] LIU, Y., LONG, X., ZHOU, D., ZENG, X., WU, Y., AND ZHANG, Y. A novel method for high dynamic range with binocular cameras. In *Tenth International Conference on Digital Image Processing (ICDIP 2018)* (2018), vol. 10806, pp. 1731 – 1737.
- [110] MALKOC, G., AND KINGDOM, F. A. Dichoptic difference thresholds for chromatic stimuli. *Vision Research* 62 (2012), 75–83.
- [111] MANN, S., AND PICARD, R. W. Being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of Imaging Science & Technology* (1995), vol. 48, pp. 442–448.
- [112] MANTIUK, R., LEWANDOWSKA, A., AND MANTIUK, R. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum* 31, 8 (2012), 2478–2491.
- [113] MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception* 3, 3 (2006), 286–308.
- [114] MATSUMIYA, K., HOWARD, I. P., AND KANEKO, H. Perceived depth in the ‘sieve effect’ and exclusive binocular rivalry. *Perception* 36 (2007), 1002 – 990.
- [115] MAUSFELD, R., WENDT, G., AND GOLZ, J. Lustrous material appearances: Internal and external constraints on triggering conditions for binocular lustre. *i-Perception* 5 (2014), 1–19.
- [116] MCLEAN, B., AND SMITH, S. Developing a wide field of view hmd for simulators. In *Display System Optics* (1987), vol. 0778, pp. 79–82.
- [117] MEESE, T. S., BAKER, D. H., AND SUMMERS, R. J. Perception of global image contrast involves transparent spatial filtering and the integration and suppression of local contrasts (not rms contrast). *Royal Society Open Science* 4, 9 (2017), 170285.

- [118] MEESE, T. S., GEORGESON, M. A., AND BAKER, D. H. Binocular contrast vision at and above threshold. *Journal of Vision* 6, 11 (2006), 7–7.
- [119] MELZER, J. E., AND MOFFITT, K. Partial binocular-overlap in helmet-mounted displays. In *Display System Optics II* (1989), vol. 1117, SPIE, pp. 56 – 62.
- [120] MELZER, J. E., AND MOFFITT, K. W. Ecological approach to partial binocular overlap. In *Large Screen Projection, Avionic, and Helmet-Mounted Displays* (1991), vol. 1456, SPIE, pp. 124 – 131.
- [121] MORGAN, M. J., AND THOMPSON, P. Apparent motion and the pulfrich effect. *Perception* 4, 1 (1975), 3–18.
- [122] MORI, Y., TANAHASHI, K., AND TSUJI, S. Quantitative evaluation of visual performance of liquid crystal displays. In *Algorithms and Systems for Optical Information Processing IV* (2000), vol. 4113, pp. 242 – 249.
- [123] NI, T., BOWMAN, D. A., AND CHEN, J. Increased display size and resolution improve task performance in information-rich virtual environments. In *Proceedings of Graphics Interface 2006* (2006), Canadian Information Processing Society, p. 139–146.
- [124] OLMOS, A., AND KINGDOM, F. A. A. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* 33, 12 (2004), 1463–1473.
- [125] OOI, T. L., AND HE, Z. J. Binocular rivalry and surface-boundary processing. *Perception* 35, 5 (2006), 581–603.
- [126] OOI, T. L., AND HE, Z. J. Sensory eye dominance: Relationship between eye and brain. *Eye and brain* 2020, 12 (2020), 25–31.
- [127] O’SHEA, R. P., AND BLAKE, R. Depth without disparity in random-dot stereograms. *Perception & Psychophysics* 42 (1987), 205–214.
- [128] PAMIR, Z., AND BOYACI, H. Context-dependent lightness affects perceived contrast. *Vision Research* 124 (2016), 24–33.
- [129] PATEL, A. S., AND JONES, R. W. Increment and decrement visual thresholds. *Journal of the Optical Society of America* 58, 5 (1968), 696–9.
- [130] PATTERSON, R., WINTERBOTTOM, M., PIERCE, B., AND FOX, R. Binocular rivalry and head-worn displays. *Human Factors* 49, 6 (2007), 1083–1096.
- [131] PATTERSON, R., WINTERBOTTOM, M. D., AND PIERCE, B. J. Perceptual issues in the use of head-mounted visual displays. *Human Factors* 48, 3 (2006), 555–573.
- [132] PELLI, D. G. The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* 10, 4 (1997), 437 – 442.
- [133] PIEPER, W., AND LUDWIG, I. Binocular vision: Rivalry, stereoscopic lustre, and sieve effect. *Perception (suppl)* 30 (2001), 1–1.
- [134] POLTORATSKI, S., AND TONG, F. Resolving the spatial profile of figure enhancement in human v1 through population receptive field modeling. *Journal of Neuroscience* 40, 16 (2020), 3292–3303.

- [135] QIAN, C. S., BRASCAMP, J. W., AND LIU, T. On the functional order of binocular rivalry and blind spot filling-in. *Vision Research* 136 (2017), 15–20.
- [136] QIU, S. X., CALDWELL, C. L., YOU, J. Y., AND MENDOLA, J. D. Binocular rivalry from luminance and contrast. *Vision Research* 175 (2020), 41–50.
- [137] RASH, C. E., MCLEAN, W. E., MOZO, B. T., LICINA, J. R., AND MCENTIRE, J. B. Human factors and performance concerns for the design of helmet-mounted displays. *USAARL Report No. 99-08* (1999).
- [138] READ, J. C., AND EAGLE, R. A. Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Research* 40 (2000), 3345–3358.
- [139] REICH, L. N., LEVI, D. M., AND FRISHMAN, L. J. Dynamic random noise shrinks the twinkling aftereffect induced by artificial scotomas. *Vision Research* 40, 7 (2000), 805–816.
- [140] REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. Photographic tone reproduction for digital images. *ACM Transactions on Graphics* 21, 3 (2002), 267–276.
- [141] REN, D., GOLDSCHWENDT, T., CHANG, Y., AND HÖLLERER, T. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In *2016 IEEE Virtual Reality (VR)* (2016), IEEE, pp. 93–102.
- [142] SCHALLMO, M.-P., AND MURRAY, S. O. Identifying separate components of surround suppression. *Journal of Vision* 16, 1 (2016), 2–2.
- [143] SEBASTIAN, S., ABRAMS, J., AND GEISLER, W. S. Constrained sampling experiments reveal principles of detection in natural scenes. *Proceedings of the National Academy of Sciences* 114, 28 (2017), E5731–E5740.
- [144] SEIDEMANN, A., SCHAEFFEL, F., GUIRAO, A., LÓPEZ-GIL, N., AND ARTAL, P. Peripheral refractive errors in myopic, emmetropic, and hyperopic young subjects. *Journal of the Optical Society of America. A, Optics, image science, and vision* 19 (2002), 2363–73.
- [145] SHI, Z., CHEN, W. T., AND CAPASSO, F. Wide field-of-view waveguide displays enabled by polarization-dependent metagratings. In *Digital Optics for Immersive Displays* (2018), vol. 10676, SPIE, pp. 272–277.
- [146] SIMONCELLI, E. P., AND OLSHAUSEN, B. A. Natural image statistics and neural representation. *Annual Review of Neuroscience* 24, 1 (2001), 1193–1216.
- [147] SKERSWETAT, J., AND BEX, P. J. Inform (indicate-follow-replay-me): A novel method to measure perceptual multistability dynamics using continuous data tracking and validated estimates of visual introspection. *Consciousness and Cognition* 107 (2023), 103437.
- [148] SKERSWETAT, J., FORMANKIEWICZ, M. A., AND WAUGH, S. J. Levelt’s laws do not predict perception when luminance- and contrast-modulated stimuli compete during binocular rivalry. *Scientific Reports* 8, 1 (2018), 14432.

- [149] SNOWDEN, R. J., AND HAMMETT, S. T. The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination. *Vision Research* 38, 13 (1998), 1935–1945.
- [150] SPECTOR, R. H. Visual fields. In *Visual Fields – Clinical Methods: The History, Physical, and Laboratory Examinations*. Butterworth Publishers, 1990, ch. 116.
- [151] STRANG, N. C., ATCHISON, D. A., AND WOODS, R. L. Effects of defocus and pupil size on human contrast sensitivity. *Ophthalmic and Physiological Optics* 19, 5 (1999), 415–426.
- [152] SU, Y. R., HE, Z. J., AND OOI, T. L. Seeing grating-textured surface begins at the border. *Journal of Vision* 11, 1 (2011), 14–14.
- [153] TONG, F., AND ENGEL, S. A. Interocular rivalry revealed in the human cortical blind-spot representation. *Nature* 411 (2001), 195–199.
- [154] TONG, J., ALLISON, R. S., AND WILCOX, L. M. Optical distortions in VR bias the perceived slant of moving surfaces. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2020), pp. 73–79.
- [155] TREPKOWSKI, C., EIBICH, D., MAIERO, J., MARQUARDT, A., KRUIJFF, E., AND FEINER, S. The effect of narrow field of view and information density on visual search performance in augmented reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), pp. 575–584.
- [156] TRIPATHY, S. P., AND LEVI, D. M. Looking behind a pathological blind spot in human retina. *Vision Research* 39, 11 (1999), 1917–1925.
- [157] TRIPATHY, S. P., LEVI, D. M., AND OGMEN, H. Two-dot alignment across the physiological blind spot. *Vision Research* 36, 11 (1996), 1585–1596.
- [158] TSAI, J. J. P., AND VICTOR, J. D. Neither occlusion constraint nor binocular disparity accounts for the perceived depth in the ‘sieve effect’. *Vision Research* 40, 17 (2000), 2265 – 2276.
- [159] VERGHESE, P., AND GHAHGHAEI, S. Predicting stereopsis in macular degeneration. *Journal of Neuroscience* 40, 28 (2020), 5465–5470.
- [160] VON HELMHOLTZ, H. L. F. *Helmholtz’s treatise on physiological optics (Translated from the 3rd German ed.)*. The Optical Society of America, Rochester, New York, 1924.
- [161] WANG, M., AND COOPER, E. A. A re-examination of dichoptic tone mapping. *ACM Transactions on Graphics* 40, 2 (2021), 1–15.
- [162] WANG, M., AND COOPER, E. A. Perceptual guidelines for optimizing field of view in stereoscopic augmented reality displays. *ACM Trans. Appl. Percept.* 19, 4 (2022).
- [163] WANG, M., DING, J., LEVI, D. M., AND COOPER, E. A. The effect of spatial structure on binocular contrast perception. *Journal of Vision* 22, 12 (2022), 7–7.
- [164] WANG, Y., HE, Z., LIANG, Y., CHEN, Y., GONG, L., MAO, Y., CHEN, X., YAO, Z., SPIEGEL, D. P., QU, J., LU, F., ZHOU, J., AND HESS, R. F. The

- binocular balance at high spatial frequencies as revealed by the binocular orientation combination task. *Frontiers in Human Neuroscience* 13 (2019).
- [165] WATSON, A. B. The field of view, the field of resolution, and the field of contrast sensitivity. *Electronic Imaging 2018*, 14 (2018), 1–11.
- [166] WELLS, M. J., VENTURINO, M., AND OSGOOD, R. K. The effect of field-of-view size on performance at a simple simulated air-to-air mission. In *Helmet-Mounted Displays* (1989), vol. 1116, SPIE, pp. 126 – 137.
- [167] WENDT, G., AND FAUL, F. Differences in stereoscopic luster evoked by static and dynamic stimuli. *i-Perception* 10 (2019), 2041669519846133.
- [168] WENDT, G., AND FAUL, F. The role of contrast polarities in binocular luster: Low-level and high-level processes. *Vision Research* 176 (2020), 141–155.
- [169] WENDT, G., AND FAUL, F. Binocular luster – a review. *Vision Research* 194 (2022), 108008.
- [170] WIESEL, T. N., AND HUBEL, D. H. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology* 26, 6 (1963), 1003–1017.
- [171] XIAO, F., DICARLO, J. M., CATRYSSSE, P. B., AND WANDELL, B. A. High dynamic range imaging of natural scenes. In *Color Imaging Conference* (2002), pp. 337–342.
- [172] XING, J., AND HEEGER, D. J. Measurement and modeling of center-surround suppression and enhancement. *Vision Research* 41, 5 (2001), 571–583.
- [173] XIONG, J., TAN, G., ZHAN, T., AND WU, S.-T. Breaking the field-of-view limit in augmented reality with a scanning waveguide display. *OSA Continuum* 3, 10 (2020), 2730–2740.
- [174] YANG, X. S., ZHANG, L., WONG, T.-T., AND HENG, P.-A. Binocular tone mapping. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)* 31, 4 (2012), 1–10.
- [175] YANKELOVICH, A., AND SPITZER, H. Predicting illusory contours without extracting special image features. *Frontiers in Computational Neuroscience* 12 (2019).
- [176] YEHEZKEL, O., DING, J., STERKIN, A., POLAT, U., AND LEVI, D. M. Binocular combination of stimulus orientation. *Royal Society Open Science* 3, 11 (2016), 160534.
- [177] YOSHIDA, A., BLANZ, V., MYZKOWSKI, K., AND SEIDEL, H.-P. Perceptual evaluation of tone mapping operators with real-world scenes. In *Human Vision and Electronic Imaging* (2005), vol. 5666, pp. 192–03.
- [178] YOSHIDA, A., MANTIUK, R., MYZKOWSKI, K., AND SEIDEL, H.-P. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum* 25, 3 (2006), 415–426.
- [179] ZHAN, T., YIN, K., XIONG, J., HE, Z., AND WU, S.-T. Augmented reality and virtual reality displays: Perspectives and challenges. *iScience* 23, 8 (2020), 101397.

- [180] ZHANG, L., MURDOCH, M. J., AND BACHY, R. Color appearance shift in augmented reality metamerism matching. *Journal of the Optical Society of America A* 38, 5 (2021), 701–710.
- [181] ZHANG, Z., HAN, C., HE, S., LIU, X., ZHU, H., HU, X., AND WONG, T.-T. Deep binocular tone mapping. *The Visual Computer* 35, 6 (2019), 997–1011.
- [182] ZHANG, Z., HU, X., LIU, X., AND WONG, T.-T. Binocular tone mapping with improved overall contrast and local details. *Computer Graphics Forum* 37, 7 (2018), 433–442.
- [183] ZHONG, F., KOULIERIS, G. A., DRETTAKIS, G., BANKS, M. S., CHAMBE, M., DURAND, F., AND MANTIUK, R. K. DiCE: dichoptic contrast enhancement for VR and stereo displays. *ACM Transactions on Graphics (SIGGRAPH Asia Conference Proceedings)* 38, 6 (2019), 1–13.