Designing Mixed Reality to Improve Spatial Perception and Accessibility

by

Haley Adams

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Computer Science

May 10, 2024
Nashville, Tennessee

Approved by

Bobby Bodenheimer, Ph.D.
Sarah Creem-Regehr, Ph.D.
Maithilee Kunda, Ph.D.
Shilo Anders, Ph.D.
Tonia Rex, Ph.D.

To the cherished memory of Jereline 'Memaw' Adams. While the final chapter arrived without you by my side, your spirit whispered encouragement with every turn of the page

Contents

Listing of tables

Listing of figures

Part I

The Preamble

The struggle itself towards the heights is enough to fill a man's heart. One must imagine Sisyphus happy.

Albert Camus

1 Introduction

1.1 Research Problem

A formidable challenge for immersive technology is the challenge of creating accurate spatial perception. Although humans adeptly perceive and act upon physical spaces, there is evidence that in mixed reality (MR)—which refers to a broad spectrum of devices that integrate computer-generated and physical sensory information into a single percept —people's perception of depth is distorted [94, 121, 132, 368]. Within the gamut of contemporary MR devices, virtual reality (VR) displays render completely virtual images, while augmented reality (AR) displays render virtual images on top of real world spaces.

Depth misperception, in which one's estimates of distances to and between objects is distorted, can adversely affect people's ability to determine the position of virtual objects in space. An accurate sense of depth is particularly important for training and guidance applications, since even minor disruptions in people's perception of space for these applications can have negative consequences. For example, overreliance on image guided navigation can introduce risk to patients if surgeons overshoot distances to anatomical structures in a surgical site during operation [56, 495]. And heads-up displays in automobiles can affect driving judgments and safety when the positions of pedestrians, obstacles, and other vehicles are incorrectly perceived [35, 434].

In completely virtual environments, like those rendered by VR displays, a pattern of depth compression has been established in head-mounted displays (HMDs) in which people's distance judgments to virtual targets are underestimated on average [64, 90, 94, 211, 212, 214, 224, 389, 425]. Some of the contributing factors for the distortion of depth perception in these displays are the limited field of view [64, 211, 212] and the weight of HMDs like those used to render virtual environments [501]. Yet less research has been conducted on devices that integrate both real and virtual visual information, like those rendered by AR HMDs. Although a recent trend towards distance underestimation in *action space*, which refers to intermediate distances between a viewer and approximately two and 30 meters [96], has begun to emerge [109, 132, 222, 368, 397, 455, 478, 479]—the results of distance perception studies in AR are more variable than the results of studies conducted in VR. We discuss perception and visual depth perception (Section 2.2) as well as how visual depth perception is affected by different HMDs (Section 2.4) in more detail later in this dissertation.

One possible explanation for the variability of results in AR depth perception has to do with how virtual graphics are integrated into real world scenes (Figure 1.1). A common complaint by users of AR devices is that objects appear "floaty". In other words, virtual objects seem to float above real world surfaces, rather than lie on top of them. Yet few studies in AR have investigated how to best establish surface contact between virtual objects and real world surfaces.

In the real world, the perception of surface contact seems effortless due to the presence of consistent

Figure 1.1: Augmented reality displays not only render virtual images—they also integrate computer generated images into real world scenes

lighting information [76, 144, 467]. For example, cast shadows, which are caused by obstructions between a light source and an object, are powerful cues for creating points of contact between objects and nearby surfaces. The role of cast shadows in depth perception as a preeminent cue for surface contact has been well demonstrated in traditional perception and computer graphics like those conducted by Ni and colleagues [332, 333] and Kersten et al. [228].

However, rendering realistic lighting effects in devices on a computational budget—like mobile phones and AR HMDs—can be expensive. Furthermore, AR HMDs that rely on additive light for rendering, like the Microsoft HoloLens 2 and Magic Leap 1, cannot render dark color values photorealistically. This makes the task of rendering realistic cast shadows in such displays difficult. Accordingly, much of the current research on shadows in AR HMDs has been dedicated towards rendering more realistic shadows given the limitations of additive light displays [203, 298, 299]. But are realistic cast shadows necessary to create surface contact?

In the field of AR, some researchers have theorized that visual depth cue conflicts between virtual and real objects (e.g., conflicting depth from shading information) is a contributing factor to distance misperception in these displays [1, 107, 217, 245]. However, if we look at cast shadows as an example, prior research by Kersten and colleagues [228] has demonstrated that the visual appearance of cast shadows can be manipulated to unnatural extents while still preserving the cast shadows' role in depth perception. Although Kersten's study was conducted in a completely virtual desktop environment, their work provides some evidence that non-photorealistic rendering approaches may be beneficial for enhancing depth perception in AR when realistic graphics are unfeasible.

Because cast shadows play an important role in depth perception, in Aim 1 of this dissertation, we investigated how the manner in which we render cast shadows influences people's ability to determine surface contact. For this investigation, we directly evaluated people's surface contact judgements across multiple MR HMDs—each of which utilized a different approach for rendering computer graphics. We manipulated the presence of cast shadow, the appearance of cast shadow, the shape of its associated object, and more over four experiments. The presence of cast shadows proved important for the accuracy of surface contact judgments. However, remarkably, people's surface contact judgments were often more accurate for non-photorealistic than realistic cast shadows, even when the evaluated MR display was capable of rendering dark color values. These investigations showed that while the presence of a surface contact cue, like a cast shadow, is important for determining the vertical position of an object in space, visually realistic shadows may not be necessary for creating surface contact in MR.

To better understand this behavior and to describe it in more generality, in Aim 2 we wanted to investigate the effect of surface contact cues and object height above a surface on people's distance judge-

Figure 1.2: A zoomed out view of the virtual environment (left) and the environment as viewed stereoscopically by someone with simulated central visual field loss (right)

ments from the viewer to virtual objects in space. As such, we conducted two evaluations of distance judgments in MR. In the first evaluation, we assessed depth perception to targets on and above the ground with and without cast shadows. In the second evaluation, we assessed depth perception to the same targets presented with either a realistic cast shadow or a non-photorealistic graphical element for surface contact. The results of this work proved complimentary to our surface contact perception experiments. We found evidence that people's distance judgements to objects with non-photorealistic surface contact cues were similar to those made to objects with realistic shadows. Through this work, we can more confidently assert that realistic shadows are not requisite for creating surface contact in MR displays. Rather, the presence of information to indicate surface contact is important for depth perception—regardless of the realism of the cue.

In the final aim of the dissertation (Aim 3), we extended the findings of our research to a larger audience by demonstrating how individuals with vision impairments may benefit from our empirical findings. People with low vision—a severe form of vision loss in which vision cannot be corrected with glasses—include fourteen million people in the United States alone [343]. The needs of users with vision impairments, particularly those with severe or moderate loss, are often neglected in the design process.

However, the needs of individuals with severe, and even moderate, vision impairments are often left out of design considerations, which is shown by the prevalence of inaccessible applications in more mature technology platforms, like mobile applications and web pages [398].

To better understand how realistic and nonrealistic graphics, like those studied in this dissertation, can be used to enhance accessibility in MR, we developed a vision impairment simulation (e.g., Figure 1.2) that allowed for precise control of the extent and severity of vision loss. Designers and developers of MR applications can better understand how their design decisions affect users with vision impairments by experiencing vision impairments firsthand through simulation. Vision simulations may also be employed for more rapid and accessible application development when test users with vision impairments are unavailable. In two-dimensional (2D) digital media, it has been demonstrated that people with severe vision loss benefit from non-photorealistic image enhancement techniques like edge and contrast enhancements [70]. However, the domain of accessibility in immersive, three-dimensional (3D) media is still nascent [322, 513]. As such, it is unknown if non-photorealistic effects will help or harm distance perception in MR for individuals with visual impairments. In the current work, we employ our vision simulation to better understand the impact of visual field loss and computer graphics on people's spatial perception.

1.2 Contributions

As described previously, the primary aims of this dissertation are:

- Aim 1: Compare perceptually realistic & non-photorealistic shading methods on surface contact perception for different virtual objects in MR.
- Aim 2: Determine the effects of shadow shading & object height above a surface on egocentric depth perception in MR.
- Aim 3: Develop & evaluate low vision simulation for accessibility to extend prior research to those with vision impairments.

1.3 Structure

In this dissertation we present the following:

I The Preamble
In Part I, we introduce the research questions of the dissertation and address how the dissertation hopes to answer them (Chapter 1).

II Foundations
In Part II, we recount the fundamental research from which the current work derives. In Chapter 2 we describe what constitutes a mixed reality display and how one's perception of space is influenced by the use of immersive devices like the ones studied in the current work. Then, in Chapter 3 we discuss how mixed reality has been employed to better understand how those with vision impairments perceive their surroundings. We also discuss how different types of visual impairment influence visual perception and how researchers have tried to recreate the experience of vision impairment through simulation.

III Surface Contact Perception
In Part III, we first introduce the motivation and framework for our evaluations of surface contact perception in MR (Chapter 4). We then describe the four experiments conducted to understand surface contact perception as well as the influence of non-photorealistic rendering approaches to cast shadows in Chapter 5. In these studies we ascertain the benefit of different rendering techniques for improving surface contact perception in MR (Aim 1). This research has resulted in two publications: Adams et al. [5] and Adams et al. [2]. Adams et al. [5] has been published in ACM Transactions on Visualization and Computer Graphics (ACM TVCG). Adams et al. [2] is awaiting submission to the same journal.

IV Depth Perception
Building upon our prior investigations of surface contact perception, Part IV explores the extent to which these findings generalize to the broader domain of spatial perception by examining distance perception. In Chapter 6, we motivate our evaluations of and describe our methodology for assessing egocentric depth perception. We examine the efficacy of non-photorealistic rendering of cast shadows for improving depth perception as well as the influence of an object's vertical position in space in MR in Chapter 7 (Aim 2). This investigation consists of two studies. The first experiment, Adams et al. [4] was published in IEEE Virtual Reality (IEEE VR)—a premiere conference venue for the dissemination of virtual and augmented reality research. The second experiment, Bodenheimer et al. [54], was published in the Proceedings of the ACM's Symposium on Applied Perception.

V Vision Simulation & Eye Tracking

In Part V, we begin to broach the final research aim of the dissertation. In this section, we extend our research on non-photorealistic rendering and spatial perception to the domain of accessibility by constructing an immersive visual impairment simulation (Aim 3). This research was supported in part by the Microsoft Research Dissertation Grant. The framework for the vision simulation is described in Chapter 8, and in Chapter 9 we discuss our eye tracking data quality assessment for two eye-tracked head-mounted displays. We perform this evaluation in order to better understand the technical limitations of our vision simulation system. At the time of the dissertation defense, this work has yet to be submitted for publication. However, an earlier version of the vision simulation was integrated into a research grant proposal to the National Institutes of Health (NIH).

VI Enhancing Depth Perception for Vision Loss

Finally, in Part VI, of the dissertation we deploy our vision simulation for use in one final distance perception assessment. In this work, we assess how central and peripheral vision loss impact distance perception judgements. And we look at how non-photorealistic, or stylized, graphics can be used to improve depth perception when a person experiences visual field loss. In Chapter 10, we motivate our decision to study the impact of central and peripheral vision loss on depth perception. We discuss the results of our experiment and the implications it has for the design of MR applications for both normally-sighted and visually impaired individuals.

VII Closing

In closing, for Part VII we briefly summarize the progress of the dissertation towards its research aims. In Chapter 11, we provide research and software development guidelines that can be derived from the body of research completed through this dissertation, and we provide suggestions for future research directions.

Part II

Foundations

All perceiving is also thinking, all reasoning is also intuition, all observation is also invention.

Rudolf Arnheim

2 Foundations for Perception in Mixed Reality

To effectively design and evaluate immersive technology, one must understand how it is influenced by human perception. Due to hardware and software limitations, immersive technology introduces simulated sensory information that does not perfectly match the real sensory information that a person experiences every day. Although developers can create compelling applications with contemporary mixed reality displays, conflicting sensory information may disrupt the user experience—resulting in adverse effects like distorted depth perception [213] , disorientation [401], and even nausea [344]. Fortunately, researchers are actively investigating how to leverage humans' perception of the senses to improve immersive displays.

Because different kinds of mixed reality displays take unique approaches to render images onto a person's retina, different displays can influence human perception in unique ways.

2.1 Mixed Reality (MR)

Many different kinds of displays fall within the mixed reality continuum. As such, having a common framework from which to reference different types of immersive displays is important for constructive discussions about how these devices can be evaluated and improved. Within this document we use the term *mixed reality (MR)* to reference any technology that mediates reality. Accordingly, "MR" encompasses the spectrum of immersive technologies [318, 432]. Moving forward, we will refer to displays that render completely virtual environments as *virtual reality (VR)* displays, and we will refer to devices that integrate both virtual and real environments as *augmented reality (AR)* displays [24]. Both virtual and augmented reality devices fall under the umbrella of mixed reality displays.

The current work is focused on a family of mixed reality displays called head-mounted displays. *Head-mounted displays*, or HMDs, refer to immersive devices in which graphics are rendered via a near-eye display. In this dissertation research, we exclusively investigate human perception in HMDs, as opposed to spatial mixed reality, large screen mixed reality displays, and the like. Specifically, we look at three types of mixed reality HMDs in the current work: *optical see-through augmented reality (OST AR)*, *video see-through augmented reality (VST AR)*, and VR head-mounted displays.

2.1.1 The Mixed Reality Spectrum

A number of taxonomies exist for classifying immersive displays both within industry and within academia [32, 276, 301], which has made the classification of these displays a somewhat contentious issue. Among the most popular of these taxonomies is the "Virtuality Continuum" proposed by Milgram and Kishino in 1994 [318]. In this seminal document, Milgram and Kishino describe all immer-

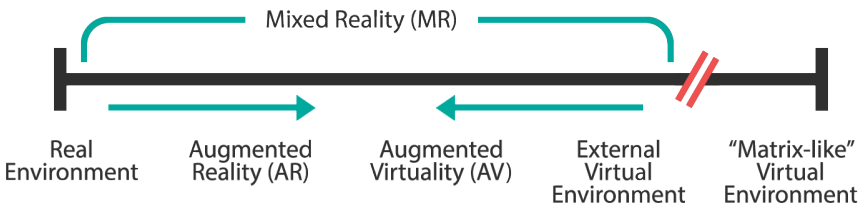

Figure 2.1: The Reality-Virtuality Continuum as described in Skarbez et al. [432]

Figure 2.2: Antique stereoscope manufactured by Underwood & Underwood. The components are made of wood, aluminum, and brown velvet. The device was patented on June 11, 1901 in New York. Stereoscope courtesy of my dear friend Joe (Yuzhou) Huang.

sive technology to lie within a "Mixed Reality" continuum, where *real environments* are shown at one end of the spectrum and completely *virtual environments* are shown at the other. Milgram and Kishino argue that any device that presents both virtual and real world objects together, within a single display, should be considered MR. However, a limitation to this classic work is that it classifies immersive displays solely based on their ability to produce visual information. This limitation is a product of the technical limitations of displays at the time, which relied predominantly on simulating visual sensory information.

More recently, the Virtuality Continuum has been updated by Skarbez, Smith, and Whitton [432]. In this document, they expand the taxonomy to include all exteroceptive senses (i.e., sight, hearing, touch, smell, and taste) as well as interoceptive senses (e.g., proprioception). A consequence of this expanded taxonomy is that the displays that we have traditionally classified as "virtual reality" displays do not represent the end of the spectrum. Because contemporary VR displays are unable to create true sensory agreement across exteroceptive and interoceptive senses, these displays are actually classified as mixed reality displays in the updated Virtuality Continuum.

Skarbez and colleagues take this argument yet another step further. They state that there is no way to avoid sensory conflicts in current immersive technology. A true virtual reality display would need to stimulate the brain directly and simulate all senses in a manner similar to the *Matrix* films. As an example, they explain that even if taste and smell were perfectly simulated, a person would ultimately become hungry since no food was ingested. Consequently, Skarbez et al. argue that true virtual reality is not achievable and is thus not aligned with the Virtuality Continuum—a point that is reflected in their updated model (See Figure 2.1). In the interest of clarity, we will continue to refer to virtual reality displays as those that produce completely virtual images, although we recognize that contemporary displays are far from creating *true* virtual realities.

2.1.2 Head-mounted displays (HMDs)

The evolution of HMDs

In 1838 Charles Wheatstone invented the first analog *stereoscope* [497], a device that allowed a person to view images with a sense of depth. The impression of depth was achieved by showing the same image, taken from different perspectives, to each eye. Shortly after their invention, stereoscopes became a popular form of entertainment and their designs became more compact like the Underwood & Underwood stereoscope in Figure 2.2. Even today, classic stereoscopes from the late 1800's and early 1900's can be found in abundance at retro arcades like the Musée Mécanique in San Francisco, California.

The first recorded, digital head-mounted display was Ivan Sutherland's *Sword of Damocles* in 1968 [453]. By combining optical see-through lenses with cathode-ray tube (CRT) monitors, the Sword of Damocles could render wireframe graphics over real world scenes. The eponymous display derived its name from a Greek parable by the same title in which the courtier Damocles sits on the throne of the tyrant Damocles only to find that a giant sword has been suspended by a single horse hair above the chair. In a similar manner, Sutherland's cumbersome augmented reality display and its position tracking system were precariously mounted upon the ceiling.

Fortunately, over the last 50 years, the weight and ergonomics of HMDs have vastly improved. The quality of computer graphics and position tracking solutions have progressed while the cost of constructing these devices has decreased. However, It was not until the cost of high density pixel displays decreased, due to the popular rise of smartphones in the late 2000's that VR began to be considered commercially viable. As a result, researchers at the University of Southern California (USC) were able to produce the technology that ultimately resulted in the production of the Oculus Rift DK1 in 2013 [14]. Since then, the field has seen a resurgence in research funding and product development.

More recently, the cost of building VR displays has plummeted from $23,900 for the nVisor SX60 in 2009 to $300 for the Oculus Quest 2 in 2021. By comparison, the price tag for current AR HMDs remains high in 2021, with Microsoft's HoloLens 2—an optical see-through AR HMD—costing $3500 and the Varjo MR3—a video see-through AR HMD—costing $5500. One of the reasons for the differences in price point across these MR displays is that there are different technical demands and limitations for each that depends on whether it is a VR HMD, a VST HMD, or an OST HMD (See Figure 2.3).

Although different MR displays may vary in modes of interaction and solutions for position tracking, there is commonality in how the displays are constructed and their fundamental limitations. For example, most contemporary mixed reality displays permit rotation and position tracking—six degrees of freedom (6DoF) tracking—within physical space. In this thesis, all of the evaluated MR displays permit 6Dof tracking.

Because we evaluate human perception in all three types of HMDs, in the next section we will discuss some of the technical differences between these displays. For more in-depth reviews of technical differences between different MR displays, see Lavalle [260], Azuma [24], Billinghurst et al. [49], and Rolland et al. [396].

Virtual reality (VR)

The concept of virtual realities captured the imagination of authors long before the technical term *virtual reality* was popularized by Virtual Programming Languages (VPL) founder Jaron Lanier in 1987 [443]. Perhaps the most compelling, forward-thinking examples of immersive technology are those

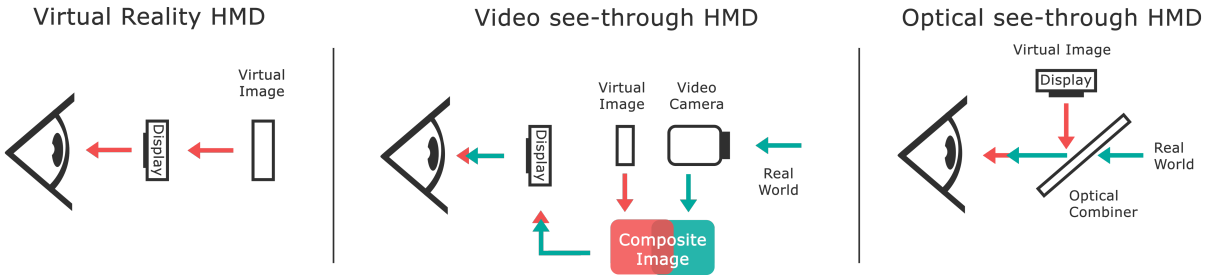

Figure 2.3: Simplified specifications for three kinds of head-mounted mixed reality displays: virtual reality, video see-through augmented reality, and optical see-through augmented reality

from the writers of the early 1900's. Short stories like Stanley G. Weinbaum's "Pygmalion's Spectacles" (1935) and Ray Bradbury's "The Veldt" (1950) introduce early conceptualizations of VR goggles and *Cave automatic virtual environment (CAVE)* environments. In a more theoretical, or arguably esoteric, manner people have argued that the conception of virtual realities existed long before any form of analog or digital technology manipulated vision. A fair argument could be made that even Plato's infamous "Allegory of the Cave" (520 CE)—in which another reality is projected as shadows upon the cave's wall—may be considered as an early inception of virtual reality.

Virtual reality HMDs are near-eye displays that rely on purely virtual rendering to create 3D images. Images are rendered on a digital panel and viewed through fixed optics positioned between the panel and the eyes of the viewer (See Figure 2.3 left). When HMDs are used to render virtual environments, the real world is completely occluded by the chassis of the head-mounted display. As a result, people's visual perception is almost entirely informed by the display. One consequence of the complete substitution of visual information is that it introduces mismatches between a person's visual and vestibular systems that can cause motion sickness [339].

Although contemporary VR HMDs are impressive feats of engineering that produce compelling virtual experiences, the devices are not without limits. In the hardware design of HMDs, there is a trade-off between field of view (FOV) and the resolution of the display such that increasing FOV often comes at the cost of decreasing display resolution and vice versa. As a result, commodity level VR HMDs typically have fields of view of approximately 110° and resolution limits that range between 24 and 50 pixel per degree (ppd). This stands in contrast to normal human vision, which provides approximately a 200° FOV [498] and a resolution of 120 ppd (or 20/10 visual acuity).

There are additional issues beyond limited resolution and field of view. For example, HMDs only accommodate for a specific range of interpupillary distances (IPDs) for users—between 60 and 75 mm. But the average adult IPD can ranges between 50 and 75 mm [105]. The problems of limited resolution, FOV, and IPD adjustments are common issues across near-eye displays, including VST and OST AR HMDs. However, the field of view of virtual reality HMDs is typically larger [49] than those provided by augmented reality displays. For a more detailed discussion of the technical challenges of state-of-the art HMDs, Zhan et al. provide an extended discussion [516].

Video see-through augmented reality (VST AR)

Video see-through HMDs create augmented reality through the use of stereoscopic video cameras, which are placed side by side on the front of the head-mounted display (See Figure 2.3 center). They

then combine virtual overlays with real world images using video composition techniques. A benefit to this approach is that it permits blending between real and virtual inputs, which is a capability that is unavailable to optical see-through displays. However, the video see-through approach also comes with a computational overhead since multiple passes through the rendering pipeline are needed to render captured videos to both eyes and to render the virtual overlays.

Video composition in most contemporary video see-through displays is performed using depth information captured from the stereoscopic cameras via pixel-by-pixel depth comparisons [24, 49]. Depth information may also be captured using small infrared or other depth sensors within the display to capture real world depth information. Depth comparisons in rendering allow both video and optical see-through augmented reality devices to provide occlusion so that real objects can cover virtual objects and vice versa, when appropriate.

A limitation to VST displays is that the resolution of the generated images is limited by the resolution of the two stereoscopic cameras. As a result, the resolution of the real world view in video see-through is degraded. Cameras can also introduce distortions, which must then be compensated for in addition to any distortions from the optics of the display. These problems are relatively solved, but they add to computational overhead of rendering. A more persistent issue for VSTs is that the video cameras that are positioned at the front of the HMD are displaced from the position of the viewer's actual eyes. This causes a disconnect between what the user expects to see and what they actually see, such that if the cameras are positioned slightly to the left of the viewer's real eyes, then the view of the user will also be shifted to the left. Even minor displacements require the viewer to recalibrate their movement to compensate for the change in viewpoint.

Like VR displays, VST HMDs can induce motion sickness after prolonged use. Even though the video cameras provide visual information of the real world environment, motion sickness is caused by the imperfect match of the video feed to what a person's visual system expects [199]. Motion sickness can be further exacerbated by other perceptual inconsistencies such as latency in the video feed [331].

Optical see-through augmented reality (OST AR)

Optical see-through HMDs are augmented reality displays that are characterized by the use of display mediums that permit direct viewing of the real world surroundings. These displays typically rely on mirrors or optical combiners to superimpose computer generated images onto real world scenes (See Figure 2.3 right).

Because optical see-through provides an unobstructed view of the real world, a user's view of the real world is not degraded in these devices. Limitations in the display's resolution only affect virtual overlays. However, the virtual overlays created by these displays are semitransparent due to the use of optical combiners and additive light to render images. Users may be able to see through virtual objects, even when they are placed in front of real world objects. The reliance on additive light for rendering, also means that OSTs cannot render black, which requires the ability to subtract light. Because the traditional computer graphics pipeline assumes the ability to subtract and add light to generate color values, traditional graphics approaches for rendering shading and shadows to not function properly in additive light displays. Another characteristic of additive light displays is that they struggle to render virtual images in bright spaces like the outdoors on sunny days.

A major limitation to optical see-through displays is that it is challenging to build OST devices with a wide field of view. Complex optics are needed to optically correct any distortions in virtual stimuli to match the real world view. In contrast, video see-through displays may use digital corrections to re-

solve visual distortions. Because optics solutions for correcting these distortions are currently expensive and because they can add weight to HMDs, even contemporary optical see-through displays can only provide a diagonal field of view between 20° and 60° [49, 396].

AR HMD Tradeoffs

There are functional tradeoffs between immersive HMDs displays, such as the ability to view the real world unobstructed in OST displays versus the comparatively easy integration of virtual and augmented stimuli in VST [49, 396]. Researchers have investigated how the tradeoffs between OST and VST displays affect a user's ability to perceive and interact with virtual objects in AR [6, 29, 311], and they have similarly evaluated how different augmented and virtual reality displays affect human perception [41, 81, 212, 213]. Although direct comparisons of different HMDs can be challenging to understand due to the variability in optics, rendering, and position tracking across experimental studies, these studies are nonetheless worthwhile as they allow us to assess specific cues and perturbations across families of devices. Only by understanding how graphical techniques impact user experience and spatial perception in different devices can researchers begin to create generalizable development guidelines for the spectrum of MR displays.

However, at present the state of this research is difficult to interpret. For example, in recent distance estimation research that compared people's perception in OST AR and VST AR displays, Medeiros et al. [311] found that OST AR displays resulted in more accurate depth perception over VST AR displays–yet Ballestin et al. found the opposite [29]. An added lay of complexity for interpretation is that both evaluations were conducted with displays that were customized in-house. In this dissertation we make comparisons between commercial displays to facilitate comparisons between our presented research and future work.

Despite the challenges presented by perception and HMD comparisons, design guidelines have begun to emerge from these studies. Ahn et al. [6] compared three AR devices: OST AR, VST AR, and mobile AR. They found that across devices, people's accuracy and speed was best in a size-matching task when they were presented with a more detailed 3D model (in this case, a 3D scanned object). In contrast, Cidota et al. [81] evaluated how 'diminishing' visual effects—in this case, fade and blur—affected depth perception when reaching to targets in OST AR and VR displays. They found that measured performance was best in VR when visual effects were applied but performance was best in AR when no visual effects were applied.

2.2 Perceiving Layout and Knowing Distances

The challenge of creating realistic representations of 3D space on flat surfaces is not a new one. Since the rediscovery of *linear perspective* by the architect Filippo Brunelleschi circa 1420 CE, the desire to create compelling illusions of depth in art has dominated visual composition in western art. Brunelleschi realized that by drawing converging parallel lines into the horizon, he could create a sense of 3D space that matched the visual perspective of the human eye. From the Renaissance onward, western artists further enhanced the illusion of depth by combining linear perspective with other techniques like *chiaroscuro*,a method in which contrasting light and shadows are used to enhance drama and depth.

The study of artistic techniques to create the illusion of depth has also influenced both the way we understand visual perception and the way we render computer graphics. In perceptual psychology, one may break down the types of visual information used to understand 3D space into two categories:

monocular and binocular depth cues. *Monocular depth cues*, which can be inferred from a single retina, are reminiscent of the techniques for generating depth in art, consisting of: linear perspective, light and shadow, relative size, texture gradient, occlusion and more. Visual examples of these pictorial cues to enhance the illusion of depth are displayed in Figure 2.4. However, because reality is not static, kinetic cues like motion parallax are also included within the family of monocular depth cues. In contrast, *binocular depth cues* rely on input from both eyes to create the perception of depth through stereopsis. Stereoscopes (Section 2.1.2) rely heavily on binocular cues to create a sense of depth.

These depth cues provide important information to the observer about where items are positioned in space. However, visual registration of depth cues in isolation provide an insufficient explanation for the complex perceptual processes and spatial reasoning abilities demonstrated by humans. As such, researchers that investigate spatial perception also emphasise the importance of ground planes, of surfaces, and of the horizon in understanding the layout of 3D space.

Figure 2.4: Examples from European art that showcase how a sense of depth can be created through monocular, or pictorial, depth cues. From left to right, the paintings gradually introduce more depth to the images through linear perspective, contrasting light and shadow, and texture gradient. *Lamentation of Christ* (c. 1305) by Giotto, *The Swing* (c. 1767) by Fragonard, and *The Calling of Saint Matthew* (c. 1600) by Caravaggio are shown.

2.2.1 Cutting and Vishton's depth cues

The human visual system relies on multiple sources of information to understand the structures of surrounding environments. Spatial information that informs the viewer about where an item is positioned in space is referred to as *depth cue* information. There are many kinds of depth cues, the most common of which are listed in Table 2.1.

The visual system integrates a large number of depth cues to create accurate representations of space. Although other lists of cues for depth information have been considered, the framework created by Cutting and Vishton [96] is the most prominent. A large body of research has been dedicated towards developing frameworks of depth cue taxonomies and towards modeling how depth cues are weighted against each other [96, 257, 451, 466]. "Taxonomy is, afterall, the beginning of science, of measurement" [96]. Although it has been demonstrated that perceptual uncertainty decreases as the number of consistent depth cues increases to the limits of the perceptual system [208, 486], how these cues are combined and utilized by one's perceptual system is still an active line of research.

In this dissertation, we rely heavily on the framework laid out by Cutting and Vishton for our discussion of depth cues. For the sake of clarity, we will briefly review common visual depth cues, their

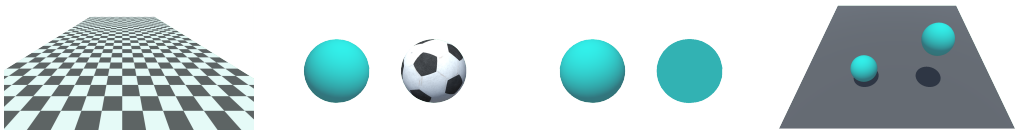

Figure 2.5: Visual examples of pictorial depth cues from left to right: texture gradient, familiar size, shading, and cast shadows

definitions, and under what circumstances these depth cues are utilized in spatial perception.

Monocular depth cues refer to the family of depth cues that are able to provide depth information from only one eye. A large subset of these cues are pictorial depth cues, which relate to depth cues that may be generated within illustrations and other still images. Simulating monocular depth information in mixed reality can be accomplished through graphical solutions and through accurate position tracking between real and virtual environments. Some of the most commonly listed monocular depth cues are as follows:

- Occlusion — otherwise known as interposition, is the obstruction of an object in the visual field when another object is positioned in front of it.
- Relative Size — is when two or more similar objects are presented at different distances from the viewer, the farther objects appear smaller. A similar cue is called *familiar size*. However, unlike relative size information, familiar size information is influenced by one's prior experiences with an object.
- Shading — is the interaction of light and shadow. As light reflects off of surfaces, different gradations of light and shadow appear onto objects and onto nearby surfaces. These light interactions are apparent even in dimly lit scenes.
- Texture Gradient — is the progressively finer appearance of textures as they recede from the viewer. It can also be seen in the appearance of groups of objects, where objects within the group appear more dense when they are positioned farther away.
- Motion Parallax — is a kinetic cue for depth in which objects that are farther away from the viewer appear to move slower than those that are near the viewer.
- Angular Declination — or the height of an object in the visual field, is the angular direction of a target relative to eye level. Angular declination is scaled to the eye height of an individual and it is typically relative to the position of the ground surface. [346].
- Brightness — a related depth cue to shading is brightness. Typically, brighter objects are seen as closer to the viewer than dim objects.
- Aerial Perspective — also called atmospheric perspective, is only present at far distances from the viewer. When one looks onto the horizon, they may notice that distant objects like mountains and building appear fainter or more blue. These modulations in contrast and hue are aerial perspective.

- Accommodation — ocular accommodation is the process of the eye adjusting its lens to focus on an object at some distance from the eye. As humans age, the muscle contractions around the eyes weaken and the lens of the eye hardens so that their eyes are no longer able to focus on targets placed at near distances. This condition is referred to as presbyopia.

Binocular depth cues create a sense of depth by employing both eyes. Simulating binocular depth within mixed reality head-mounted display is typically accomplished through optics and hardware solutions. The two binocular depth cues are:

- Vergence — is the physical rotation of eyes that occurs when the eyes focus on near and far objects. When the eyes focus on nearby objects, they simultaneously rotate inward to permit binocular fusion. Conversely, the eyes rotate outward for objects that are farther away.
- Binocular Disparity — is the difference in position between two eyes. This horizontal separation allows for the view of each eye to overlap to create the perception of depth upon fusion.

Discussions of depth perception typically refer to the perceived distances between an observer and a target positioned in space. This type of depth perception is referred to as *egocentric distance* perception. Much of the early work in psychology and the contemporary work investigating perception in mixed reality focuses on the study of egocentric distance perception since egocentric perception affects not only one's perception of space but also their ability to perform actions. This term stands in contrast to *exocentric distance* perception, which refers to the perceived distance between two external targets.

Depth cues are weighed by distance

The potency of the information provided by most depth cues is dependent on the distance of the cue from the observer. Effective distance ranges are typically divided into three areas: personal space, action space, and vista space [96]. *Personal space* refers to the space that is within 2m of the viewer, which is the space within and just beyond an arm's reach.

Action space, a space in which one may quickly perform actions such as throwing and talking, refers to distances between 2m and 30m. And *Vista space*, as the name implies, refers to distances beyond, typically farther than 30m. These divisions and the designated weights are idealized delineations, but they have been functional for guiding research in vision, perception, and computer graphics.

Table 2.1 displays each of the mentioned depth cues by cue type (monocular, binocular, & oculomotor), by effective distance range (personal, action, & vista), as well as by whether the cue provides absolute or relative information about an object's position in space.

2.2.2 Methods to evaluate distance perception

Perception is dynamic. Because perceptual systems attend to information as needed, they are influenced by the task that is being performed in a given moment. As a result, it is perhaps unsurprising that the results of distance estimation experiments, even in real world environments, can vary depending on the measurement paradigm used [13, 92, 153] and depending on the environmental context that is provided [13, 258, 508].

There are several paradigms for evaluating depth perception, both in reality and within immersive technology. Although, each paradigm requires people to make judgments about perceived distances, the spatial information that people must encode differs. Distance estimation paradigms may

require participants to register *egocentric depth information*, in which distance information between the viewer and objects in space must be encoded, or *exocentric* (also known as *allocentric*) *depth information*, in which relative distance information between different objects in space must be encoded. This distinction is important as prior research has shown judgments of distances to be anisotropic such that egocentric and exocentric distance judgments differ even when the same extents are evaluated [275, 305, 442], although, depending on the experimental task, there may be interactions between the two [402, 442].

Exocentric distance measures

For exocentric distance perception, the most commonly employed measure is *perceptual matching* (aka *visual matching*). Perceptual matching is a relative measure for depth perception, which, in AR research, involves aligning a virtual target with a real one in space. Other experimental designs that evaluate relative depth judgments include *depth ordering*, in which an array of targets must be sorted from nearest to farthest and *two alternative forced-choice (2AFC)* design, in which a viewer must decide which one of two presented targets is closer. In this dissertation we rely on a 2AFC paradigm to evaluate surface contact in Chapter 5 and we have proposed the use of visual matching to evaluate depth perception in individuals with vision impairments in Chapter 7

Egocentric distance measures

The most common techniques for evaluating egocentric depth perception in immersive technology are *verbal report* and *visually directed actions* like blind reaching and direct blind walking. Verbal report is a cognitive measure [282], whereas visually directed actions require perceptual motor coordi-

Table 2.1: The different depth cues with their effective distance ranges, modified from prior research by Cutting and Vishton [96], Nagata [325], and Kytö et al. [254].

Depth Cues and Their Properties			
Depth Cue	Cue Type	Effective Distance Range	Type of Distance
	(M)onocular, (B)inocular, (O)culomotor	*(P)ersonal, (A)ction, (V)ista*	*(A)bsolute, (R)elative*
Occlusion	M	P, A, V	R
Relative Size	M	P, A, V	R
Shading	M	P, A, V	R
Texture Gradient	M	P, A, V	R
Motion Parallax	M	P, A	A, R
Angular Declination	M	A, V	A, R
Brightness	M	A, V	R
Aerial Perspective	M	V	R
Accommodation	M, O	P	A
Vergence	B, O	P	A
Binocular Disparity	B	P, A	R

nation [370]. Responses in verbal report are recorded by asking a person to estimate the distance to a target using a familiar distance unit [282]. In contrast, visually directed actions require a participant to view a target, close their eyes, and then perform an action directed towards the target [283]. A number of different actions have been used to conduct visually directed action experiments. *Direct blind walking* requires participants to walk in a linear trajectory while blindfolded to a previously viewed target, while *blind reaching* requires participants to reach towards a previously viewed target while blindfolded. Other visually directed action paradigms, like blind throwing, exist, as well.

Alternative action-based measures that require more complex strategies, like *triangulated blind walking* [232, 233, 500] or *timed imagined walking* [153, 232, 373] may be employed when physical space is limited. Triangulated walking requires participants to indirectly walk to a previously viewed target while imagined blind walking requires participants to pretend they are walking to a previously viewed target. However, prior research has shown that these alternative action-based measures are less accurate than direct blind walking [153, 500], perhaps due to their cognitive complexity. And, although triangulated walking may be theoretically employed to evaluate farther distances, Klein et al. has shown that its accuracy deteriorates at distances greater than 10 meters [232]. In this dissertation, we rely on verbal reports of distance estimates to understand the influence of shadows on absolute distance perception in Chapter 6.

Distance perception measure trade-offs

When evaluating egocentric distance perception in immersive technology, verbal report is often more feasible to execute in practice than visually-directed actions. Verbal report may be used at any distance, while protocols like direct blind walking and blind reaching are limited to a specific range of actionable distances. In addition, technical constraints of equipment (e.g., tethered power cable and position tracking quality) can preclude the use of visually directed actions. Verbal report may also be used for measuring exocentric distance judgments between targets, although this approach is less common.

However, perhaps because verbal report is a cognitive measures, verbal distance judgments are often more variable [13, 247, 508] than visually-directed actions, and they can be susceptible to anchoring effects—in which prior judged distances influence later judgments over time [353, 464, 465]. In contrast, visually-directed actions are reputed to be a more accurate measure of people's distance perception [13, 353]. For example, direct blind-walking is generally accurate up to 20 meters in real world environments [281, 283, 392].

Yet, as pointed out in recent work by Feldstein et al. [121], even when a seemingly accurate measure like blind walking is employed, people's accuracy can vary between groups of participants. As such, even though most prior research on blind walking has shown accurate distance estimation [95, 205, 281, 392, 440], a notable amount of research has also found distance underestimation with this protocol [13, 213, 224, 502]. In contrast, people's verbal report judgments consistently exhibit some underestimation even in real world studies [224, 232, 283, 389].

Different degrees of accuracy across depth perception measures may suggest that a common, underlying perceptual representation informs two differently calibrated responses [283, 370] or it could suggest separable underlying representations for perceptions of space and action [89, 151]. The exact perceptual representations that underpin depth perception are a current area of theoretical debate [92, 283, 370]. However, contemporary researchers have begun to favor a more integrative, embodied approach that permits the interplay between representations of space and action [92]. As such, perception researchers often stress the importance of using multiple measures of distance perception to

understand how depth is perceived [121, 389, 508].

For example, verbal reports are typically influenced by the visual context of an environment even when visually-directed actions are not [247, 508]. Studies have found distance estimates in indoor environments may be more accurate or overestimated than those made in outdoor environments [13, 258, 369, 464, 465, 508], and this effect is more frequently pronounced in verbal estimates of distance than in visually directed actions [13, 464, 465, 508]. This may be due to the presence of fewer anchoring depth cues in flat, open environments. The influence of environmental context is of particular interest to the current work, since prior research by Ooi, He, and colleagues demonstrated that context effects associated with a ground plane can influence depth perception when blind walking and perceptual matching measures are employed [427, 510]. The ground plane is especially important for evaluations of depth perception and the influence of shadows on depth perception.

Distance estimation studies are typically conducted as relative pairwise comparisons to isolate factors that contribute to distance underestimation. With this approach, variability across distance estimation measures is less important but which evaluation method is selected is. For example, it may be a good idea to use verbal report when environmental context or graphical quality is manipulated since this measure is more sensitive to contextual manipulations. A case in point is that almost all depth estimation studies in VR displays have found distance judgments to be underestimated—regardless of measurement protocol [64, 90, 248, 359, 373, 389, 500, 518]. However, within this body of work, the quality of graphics has influenced distance estimates with verbal report but not visually-directed actions [248, 468].

2.3 Perception of Visual Space

2.3.1 Gibson's Ground Theory of Space Perception

In 1950, James J. Gibson introduced his ground theory of space perception in which he proposed that there is "no such thing as the perception of space without the perception of continuous background surface" [144]. In this document and many others that followed, Gibson postulated that visual space is not defined by arrays of objects in empty air—but by the layout of surfaces, sets of adjoined surfaces, and entities that are arranged in relation to surfaces [145, 146].

As the field of computer science advanced, Gibson's theories were tested with computer generated graphics. By evaluating people's distance perception, researchers were able to manipulate intermediate platforms and surface discontinuities to measure how they affected one's perceived distance to a target in highly controlled, desktop virtual environments [48, 315]. In particular, Meng and Sedgwick demonstrated that when continuous ground surface is disrupted, the visual system is unable to establish a reliable frame of reference [314, 419]. The visual system, consequently, fails to obtain correct estimates of absolute distance. As such, the accuracy of people's distance perception is disrupted when surface discontinuities along the ground—such as gaps in the floor [427] or changes in texture gradient [314]—are positioned between a viewer and a target.

2.3.2 The role of surfaces in perceiving spatial layout

When cues that link objects to nearby surfaces are absent, individuals judge distance based on *optical contact*—or the location at which the projected image of an object contacts the image of the ground beneath it—to determine position in space [314, 332, 352]. As a result, people's distance judgments to

targets positioned above the ground are perceived as farther away than those positioned on the ground [332, 397]. This phenomenon has been demonstrated in both real [383, 384] and virtual environments [48, 315]. More recently, Salas-Rosales et al. [397] have confirmed this effect in augmented reality, as well. Specifically, Salas-Rosales and colleagues demonstrated that floating targets are perceived as farther away than grounded ones in an optical see-through augmented reality display, the Microsoft HoloLens 1.

It is also possible to approximate distances in the absence of a ground plane by using *angle of declination* (i.e., the angle between the viewer's eyes and the horizontal ground plane or an object along the horizontal ground plane) alone [346, 417]. However, correct distance judgments at intermediate distances require both angle of declination and ground-surface information for correct judgments [346, 347]. The presence of a ground-surface is important for obtaining the viewer's eye height, which can be determined using the body's senses and visual information from the surrounding environment [191, 417, 447, 510]. Because angular declination is defined with reference to an individual's eye height, the presence of a reliable ground plane is requisite for accurate perception of eye height as well as angular declination [427, 510].

Gibson's formulation of the perception of space draws connections between surfaces and objects within an environment. The visual system extracts information about surface layout from patterns of light and divides the problem into a set of visual cues, with each cue describing a particular visual pattern that can be used to infer properties of surface layout or other aspects of the environment. These visual depth cues can be isolated to better understand how individual pieces of visual information contribute to our perception of depth.

The influence of height in the visual field

It is possible to approximate distances in the absence of a ground plane with *height in the visual field* alone [346, 478]. Height in the visual field is closely related to, and sometimes equated with, *angular declination* (i.e., the angle between the viewer's eyes and the horizontal ground plane) [352, 487, 510]. However—even though height in the visual field may be used to approximate distances when an implicit, invisible horizon is assumed [346, 478]—accurate distance judgments often require both height in the visual field and ground-surface information [144, 346, 347, 447]. This is because the presence of a reliable ground plane permits the accurate perception of eye height, which is an important variable for computing height in the visual field [427, 447, 510].

As such, it is important to consider how the vertical position of a virtual object can influence people's distance judgments when interpreting prior research on depth perception in AR. While some of this work has been conducted with targets positioned on the ground [132, 212, 213, 222, 372, 455], much more depth perception research has been conducted with floating targets [41, 81, 100, 102, 135, 157, 255, 311, 360, 361, 428, 429, 456].

Few AR studies have considered how the vertical position of a virtual object (e.g., whether the object is grounded or floating) can influence people's distance judgments. Dey et al. [101] found that height in the visual field influenced people's depth judgments in a mobile AR study. And Kytö et al [255] revealed that people's confidence in depth judgments was worse for floating targets that were higher in the visual field (i.e., 1.0m versus 0.5m above the ground).

Yet, in the last few years, research published by Salas-Rosales et al. [397] as well as Hertel et al. [175] has provided evidence that depth judgments in AR differ when target objects are presented as floating or along the ground. Specifically, both Salas-Rosales and Hertel provided evidence that people perceive floating objects as farther away than those presented on the ground when cast shadows were not ren-

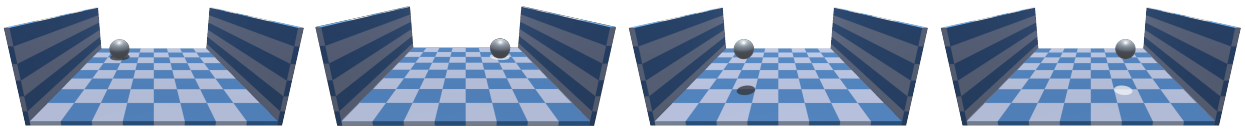

Figure 2.6: A recreation of the "ball-in-a-box" demonstration [228]. Although the sphere is located at the same position in all four images, it appears closer in the bottom images due to the placement its cast shadow. The influence of cast shadow on an object's position works regardless of whether the shadow is rendered realistic (left) or nonrealistic (right) manner. Top: the sphere seems to rest along the ground due to an anchoring shadow beneath it that connects the sphere with the back of the box. Bottom: the sphere seems to float above the ground with a detached shadow beneath it that connects it towards the front of the box.

dered in AR. These findings support prior research in both real and virtual environments that shows people rely on optical contact information to make depth judgments when surface contact cues are absent [314, 332, 352] (See Section 2.3.3). The importance of surface contact information for floating objects is further reinforced by the research reported in this dissertation in which people were less confident in surface contact judgments to target objects when cast shadows were removed [5].

2.3.3 The impact of shadows on surface contact

Cast shadows—from which humans infer size [75, 491, 514], from which humans infer shape [78, 234], and from which humans infer the distance between an object and adjacent surfaces[195, 295, 467]—play a particularly important role in our perception of space. In traditional computer graphics, evidence shows that shadows function as "visual glue" to attach virtual objects to surfaces [295, 467]. Furthermore, people become more accurate when estimating egocentric distances when objects are clearly connected to the ground via shadow [333]. Distance perception may also be enhanced by the presence of another object being wedged between the target and the ground plane beneath it [314, 315].

In fact, one's perception of the position of an object can shift dramatically depending on the placement of its associated cast shadow [297, 333]. Kersten et al.'s "ball-in-a-box" study provides perhaps one of the most poignant examples of this response [228]. Kersten and colleagues demonstrated that by changing the location of a shadow in space, even a stationary object may appear to move (See Figure 2.6). Furthermore, their research demonstrated that cast shadow shading could be manipulated to unnatural extents—in their case with light, non-photorealistic, light shading—and yet still provide a powerful tool to determine spatial location.

In Kersten et al [228], given a stationary target, lightly shaded (photometrically incorrect) shadows were less effective in producing apparent motion in depth than more traditional, dark shadows. However, in a 3D environment with motion cues, light shadows proved as effective as dark shadows for determining spatial location [228].

Related research in the domain of visual search may provide some insight into the way we processes non-photorealistic lighting information. Visual search investigations have found that light shadows are processed more slowly (a matter of milliseconds) than dark shadows in visual search tasks—providing evidence that a higher-level cognitive process may be required to process shadow shading approaches that do not conform to the darkness constraint of more naturally occurring shadows [110, 204, 390]. Should the reader desire a deeper understanding of the relationship between shadows and perception,

we suggest "The Visual World of Shadows" by Casati and Cavanagh as a notable, recent review [76].

2.4 Perception of Visual Space in MR

Although the real world is filled with rich, consistent depth information [179, 257], mixed reality displays only provide a subset of these cues. The presence of conflicting depth cue information causes perceptual uncertainty, which indicates that an individual lacks confidence about their assessment of where objects are positioned in space [107]. It has been demonstrated that perceptual uncertainty decreases as the number of consistent depth cues increases to the limits of the perceptual system [208], and it has been theorized that consistent depth cue information between real and virtual objects will improve depth perception in mixed reality, accordingly [1, 107, 245].

In computer graphics, researchers and developers have utilized information from human perception and properties of light and shadow to create increasingly realistic imagery. However, there are limitations to what can be rendered for a given display based on the pixel density, the power requirements, the optics of the display, and whether graphics need to be rendered in realtime or not. So then what are the requirements for accurate depth perception in mixed reality displays? How should virtual images be rendered to create accurate depth perception? And can we leverage this information to enhance the accessibility of this technology?

2.4.1 Depth perception in MR

Depth is compressed in VR

Virtual reality distorts distance perception. Research over the past 20 years assessing people's abilities to judge absolute distances to targets, through measures like blind walking and verbal report, have shown that people consistently underestimate distances in VR (see Creem-Regehr et al. [90] or Renner et al. [389] for recent reviews). Depending on the lab and the measure of distance perception employed, underestimation has ranged from 40-80% in virtual reality, regardless of differences in HMDs and tracking systems [90]. The problem of distance underestimation, in which virtual targets are perceived as closer than their actual position, is also referred to as *depth compression*.

Quality of graphics has also demonstrated an effect on distance judgments when a cognitive measure of distance perception, verbal report, was used [468] but not when an action based measure of distance perception, blind walking, was employed [501]. The results of these studies indicate a tentative effect of the quality of graphics on distance perception in VR. However, when interpreting distance perception research, special attention must be paid to the experimental protocol that has been used, since experimental protocol influences what spatial information is encoded to make judgments. For example, a cognitive measure like verbal report is more sensitive to the influence of environmental context. And an action based measure, like blind walking, is susceptible to the angle of declination from a viewer to a target—especially when a small target is placed on the ground. Employing multiple measures of distance perception is recommended to more confidently generalize effects.

Only with the advent of commodity-level HMDs has distance compression in virtual environments begun to diminish in severity. Researchers have now successfully isolated some factors that contribute to distance compression, including mechanical factors, such as added head weight [64, 501] and restricted field of view [233, 274, 501] These findings have been used to inform the design of contemporary head-mounted displays as well as their applications. As a result, recent studies evaluating distance

Figure 2.7: A researcher wearing the Microsoft HoloLens views a virtual cube that is placed on the floor. Image from Salas-Rosales et al. [397].

estimation in commodity VR devices have reported 10% or less underestimation [64, 94, 425]. However, depth compression still persists in contemporary virtual reality head-mounted displays [106].

Depth is misperceived in AR

Although real world depth cues are consistent and reliable, only a subset of depth cues are available for the superimposed graphics rendered by AR displays. Furthermore, in augmented reality depth information provided by virtual objects often conflicts with the depth information provided by the real world environment (e.g., inconsistent shading), which produces cue conflicts and perceptual uncertainty [107]. Figure 2.7 visualises this problem as captured by the Microsoft HoloLens, and optical see-through AR device.

Interpreting prior depth perception research in AR can be perplexing. When estimating egocentric distances, some prior studies have found that people underestimate distances in augmented reality [47, 255, 397, 455, 479]—a pattern similar to the one consistently elicited by virtual reality devices [94, 282]. For example, in 2010, Grechkin et al. found comparing blind walked estimates of the distance to objects in a virtual hallway compared to a visually matched real hallway with objects displayed through augmented reality found similar underestimation (60% of real distances) of AR judgments and VR judgments [153]. Yet other depth evaluations have found patterns of accurate estimation or even overestimation [213, 374, 454]. Furthermore, this variability in results occurs across both video see-through [255, 479] and optical see-through HMDs [94, 213, 374, 397, 454, 455].

There are several factors that may play a role in immersive AR, including the depth cues provided, the distances evaluated, and the type of display evaluated. At present, it is difficult to draw reliable connections between distance estimation results for OST and VST AR displays because of differences between devices or experimental protocols. Direct comparisons between devices may provide important insights into how the technical tradeoffs between AR HMDs influence perception. At present, it is difficult to ascertain how depth perception in one family of AR display differs from another given the paucity of direct comparisons [6, 29, 311, 361]. Even among those few comparisons, none have been conducted between two contemporary, commercially available AR HMDs.

Depth in Optical See-through AR

In OST AR studies that evaluate depth perception judgments to targets at near distances (distances $<$ 2m) have been variable. While more studies have found that people's depth judgments are overesti-

mated, especially when compared to real world objects [361, 395, 428, 429, 456], a notable amount of work has also found underestimation [307, 311, 360]. However, distance judgments in action space, which ranges between 2m and 30m, consistently trend toward underestimation [109, 132, 175, 222, 368, 397, 455, 478, 479].

In their absolute distance estimation study, Salas-Rosales et al.[397] found that distance judgement to virtual targets in the Microsoft HoloLens 1 were underestimated by 15% on average when targets were on the ground and by 7% on average when targets were floating. Keil et al. [222] similarly reported that verbal reports to distances were approximately 12% underestimated on average. Both studies evaluated distances under 7 m.

At somewhat farther distances in action space (10 m - 30 m), Gagnon et al. [132] reported perceptual matching judgements of depth in the Microsoft HoloLens 1 to virtual human avatars in space to be underestimated by 15%, whereas distance matching judgements were underestimated by 3% to real humans. However, they found notably more distance underestimation when an absolute distance measure, verbal report, was used to measure distance estimates. In this case distances were underestimated by 36% and 10% on average to virtual and real humans, respectively. The authors argue that the additional underestimation expressed in their verbal report study may be due to anchoring effects [464, 465]. Only one study thus far has been conducted in the Microsoft HoloLens 2. This work, conducted by Hertel & Steinicke [175], evaluated relative depth judgements to farther distances yet (15m - 75m) via perceptual matching, and they found distances were underestimated by approximately 14.3%, on average.

This research provides some preliminary evidence that depth perception in the Microsoft HoloLens 2 is similar to depth perception in the HoloLens 1, despite differences in specifications, like field of view. Moving forward, it will be important for future research to evaluate depth perception in action space for different, contemporary OST AR displays like the Magic Leap One to be able to generalize these research findings.

Depth in Video See-through AR

Less distance perception research has been conducted in VST head-mounted displays [109, 131, 368, 478, 479], and much of it has been conducted in recent years. Further, the majority of these studies have been conducted with either custom built or retrofitted displays, which can complicate comparisons across studies. Jamiy et al [109] and Vaziri et al. [478, 479] both created customized video see-through displays by affixing forward-facing cameras to the front of commercial VR devices. Vaziri modified the nVisor ST50 while Jamiy modified an Oculus Rift DK2. Pfeil et al. [368] and Gagnon et al. [131] both evaluated distance perception using a ZED Mini camera attached to the front of the HTC Vive and HTC Vive Pro, respectively.

Using both verbal report and blind walking, Jamiy et al. [109] found that absolute, egocentric distances in VST AR were underestimated by approximately 20% overall, for both estimation protocols. Vaziri et al. [479] then compared the effect of unaltered video input and non-photorealistic video input (via a Sobel filter) on people's distance perception with blind walking. The found that distances in both conditions were underestimated when compared with real world distance estimates. But they did not find any significant difference in responses between the two viewing conditions. In their followup work, Vaziri et al. [478], reported less distance underestimation, approximately 10% underestimation, when evaluating distance judgments in an open field.

Using the ZED Mini, Pfeil et al. [368] assessed people depth perception to real targets on the ground

via blind throwing. They found that in their VST AR condition, people underestimated targets by 8% on average—in comparison to 5% underestimation with natural viewing conditions and 7% underestimation when the real world was viewed with a restricted field of view. Although Gagnon et al. [131] evaluated distance perception in the ZED Mini, the real world environment was not visible since the viewer was presented with a virtual environment via a green screen positioned in front of them.

Interestingly, none of the aforementioned distance estimation studies in VST AR evaluated distance judgments to virtual targets. They have instead chosen to evaluate distance judgments to real world targets viewed through video-see through display [109, 368, 478, 479]. This is because, in VST AR, one's visual perception of the real world is mediated by stereoscopic video cameras. As such, researchers have assumed that distance perception—even to real world objects—will be misperceived in VST AR [49, 396]. Thus far, the results of prior research has confirmed this assumption. In contrast, OST AR displays allow for unmediated viewing of the real world. As such, depth perception research in these displays has focused on evaluating peoples depth judgments to virtually rendered targets within real world spaces. In this dissertation, we bridge this gap in literature by assessing distance judgements to virtual targets in both video and optical see-through displays (Chapter 6).

Our understanding of how people visually perceive space in video see-through augmented reality HMDs is still in a nascent stage. A wide range of technical setups have been employed to study depth perception in these devices, which makes it difficult to isolate what specific mechanical, ergonomic, or optical properties contribute to distance compression in these devices. Although, there is some evidence that field of view may be a factor [368]. It is also unclear to what extent distance judgments to virtual targets will differ from distance judgments to real targets when their appearance is mediate by stereoscopic cameras.

2.4.2 The importance of shadows in MR

As discussed in Section 2.3.3, cast shadows provide a strong and salient cue for depth perception by forming a point of contact between an object and adjacent surfaces [195, 295, 467]. This relationship has been demonstrated further by distance perception research in virtual environments in which people's egocentric distance estimates to targets were more accurate when cast shadows were present [333].

And, yet, most of the aforementioned traditional graphics research was conducted via desktop and in completely virtual environments. It is not known how the manner in which shadows are rendered across head-mounted virtual and augmented reality devices affects a viewer's sense of ground contact to improve spatial perception in AR. This may be especially problematic for augmented reality devices, which combine both real and virtual depth cue information. This process often results in conflicting depth information and increased perceptual uncertainty [1, 107, 208].

For both video and optical see-through displays—and, indeed, any graphical device on a computational budget—rendering lighting effects can be expensive. Due to this and due to the dearth of commercially available AR HMDs prior to 2016, few studies in immersive AR have examined the effect of shadows on depth perception [5, 102, 135, 175, 449].

However, researchers have begun investigating how graphically provided depth cues must be rendered to enhance spatial perception in AR devices. While the results of some research evaluating depth cues like shading and texture have been mixed, findings suggest that shadows successfully improve the accuracy of depth perception in both immersive OST AR [102, 135, 372] and VST AR [47, 449] displays. This literature also suggests that the manner in which shadows are rendered makes a difference for accuracy.

Shadows in Optical See-through AR

Rendering cast shadows can be challenging, especially in OST displays that rely on additive light for rendering. In these displays, the darker the color value, the more transparent the appearance of the color becomes until the color becomes completely transparent when black is rendered. Because of this, much of the prior research investigating cast shadows in AR has been directed towards how to best render them in OST AR displays [193, 203, 235, 298, 299] or how new display technology may be constructed to allow for subtractive rendering in optical see-through displays [206, 219].

The inability to produce dark color values with fidelity in OST AR is especially problematic for rendering cast shadows. Manabe et al. [298, 299] developed methods to produce more perceptually valid shadow rendering techniques for OST devices. Ikeda et al. [203] used simultaneous contrast to create the illusion of dark color values within the umbra of a shadow by applying brighter color values outside of the shadow's penumbra. In their evaluation, they found that their illusory shadow method caused the umbra of the shadow to be perceived as dark–despite the fact that nothing was rendered within the shadow's umbra itself.

This finding provided evidence that simultaneous contrast illusion may be an effective approach for rendering cast shadows in additive light displays [203]. Our first experiment in this dissertation uses a similar technique, which we refer to as the gradient shadow, to render more realistic shadows in optical see-through displays. However, our technique uses a simple linear falloff to produce a gradient, whereas the algorithms proposed by Manabe, Ikeda, and colleagues [203, 298, 299] produce more complex lighting interactions. For example, Ikeda et al. [203] use a photograph as a texture to estimate the radiance of the surface along with an empirically determined constant to account for viewpoint changes in their falloff algorithm.

Research on shadow generation algorithms gives rise to another question: how much fidelity is needed for accurate spatial perception of shadows in AR? In this dissertation we evaluate people's surface contact judgments for both perceptually motivated shadow rendering methods, including a simple, dark gray shadow as well as a variant of the method developed by Manabe, Ikeda, and colleagues [299], and a non-photorealistic rendering (NPR) method for shadows. Over a series of studies, which are reported in Chapter 5, we found that more realistic rendering techniques for shadows did not inherently benefit surface contact judgements. Further, people's judgements for the NPR method, which produced a white shadow, were more accurate in both OST AR and VST AR head-mounted displays. In our proposed research, we extend this research to the problem space of egocentric distance perception in order to better generalize our results and understand how our research may be used to interpret prior research on the influence of shadows on depth perception in OST AR.

Previous research on depth perception in OST AR has provided evidence that lighting misalignment—in which the position of real world and virtual lights do not coincide—may adversely affect distance perception [135], unless the misalignment is due to the use of drop shadows [102]. Drop shadows are dark silhouettes ("shadows") that are displayed ("dropped") immediately below an object–regardless of the position of light sources in a scene. In a study of depth perception in immersive OST AR, Diaz et al. [102] found that participants' depth perception was significantly better in a drop shadow condition over a coherent lighting condition. This same work also suggests that the salience of a shadow may affect spatial perception such that more transparent shadows are less effective as depth cues.

Diaz et al. [102], Gao et al. [135], and Hertel et al. [175] have all found that people's accuracy in relative depth judgments improves when floating targets are rendered with cast shadows.Hertel et al. [175] advanced this research a step further by comparing relative depth judgments for floating targets with

cast shadows to grounded targets without shadows. However, because Hertel and colleagues did not include a condition where objects were rendered on the ground with a shadow, as well, it is impossible to interpret the relationship between cast shadow and height in the visual field with this study alone. We hope to address this gap in literature with our distance estimation studies in Chapter 6.

Shadows in Video See-through AR

Unlike OST AR devices, VST head-mounted displays are immersive augmented reality devices that benefit from the ability to use traditional shadow shading techniques for rendering since they are able to render dark color values [49]. Accordingly, most of the research investigating shadow rendering has focused on estimating lighting information from real world scenes so produce more realistic alignment between real and virtual lights. This line of research is especially prominent in mobile AR [37, 156]. Perhaps not surprisingly, much of the research on rendering in VST AR has been conducted in mobile AR displays as opposed to VST HMDs, which may be attributed to the ubiquity of mobile devices.

Research conducted in mobile VST displays has evaluated how the presence of shadows and other monocular cues affect a viewer's depth perception [47, 100, 104]. The findings of this research encourage the pursuit of using monocular cues to improve depth perception in AR [47, 100, 104]. However, it may be that not all of these findings transfer to immersive VST HMDs since the afforded depth cues are different. For example, HMDs provide stereoscopic viewing. Given the dearth of research on rendering in VST AR that evaluate how lighting information affects people's perception of space, we believe our research will provide valuable groundwork for future research in this area.

There are two ways of spreading light: to be the candle or the mirror that receives it.

Edith Wharton

3 Foundations for Vision Impairment

Understanding how visually impaired individuals interact with their surroundings is essential for designing accessible applications in mixed reality. In general, accessible products benefit all users—even those without disabilities. For example, closed captions, which were initially designed to allow those with hearing impairments to understand videos on television, are also beneficial viewers who want watch media in noisy environments and for viewers who require translated subtitles to enjoy media in a non-native language. Likewise, in the current work, understanding how graphics in MR affects perception and interaction for visually impaired individuals, may also help developers design more effective solutions for people with healthy vision.

Mixed reality, if designed properly, is well poised to help individuals with accessibility needs. Augmented reality may visually enhance surroundings or supplement visual information with 3D audio cues for those with visual impairments [317]. And virtual reality may provide closed captions for those who are deaf or even simply distracted. However, at present most MR applications are not designed with accessibility in mind [322]. Even though accessibility criteria have developed in recent years for 2D technology [388], it can be difficult for developers to comprehend how their design decisions affect people with accessibility needs in immersive media. How to best render visual information for those with visual impairments, for example, is an open area of research [517].

Inaccessibility precludes a more diverse and interesting audience for MR applications. If immersive technology is to commercially succeed, this accessibility problem must be resolved. Accurate spatial perception may be especially important for assistive technologies that are designed to help individuals navigate space in mixed reality. Virtual reality can allow those with mobility issues to explore. For those with vision impairments, mixed reality applications may enhance the visual images of a person's surroundings or supplement visual information with spatialized audio. For those with hearing impairments, augmented reality applications may provide closed captions so they are note excluded from verbal conversation.

The domain of assistive technologies, in which devices help individuals perform everyday tasks, and accessible technologies, in which devices are inherently designed to include the widest audience possible, can seem daunting as there are many forms of sensory impairments that can affect how people interact with technology. Therefore, in the current thesis, we focus exclusively on visual impairments. And in our coverage of foundations for designing accessible technology for mixed reality, we will discuss vision impairments, how vision impairments can affect interactions, how vision impairments may be simulated to allow normally-sighted individuals to better design for these populations, and even the limitations of disability simulations for vision impairments. In covering these topics, we hope to not only inform the reader about essential background information for the current thesis but to also provide the reader with insights that will allow them to promote accessibility in their own work.

3.1 Vision Impairment

Visual impairments are often the result of *scotomas*—or a blind spots in one's visual field. Scotomas vary in size, intensity, number, and placement within a person's visual field. As such, low vision conditions are heterogeneous and vary widely between conditions and even between patients with the same condition. Two of the most common types of visual field loss that can affect people's vision are: central and peripheral field loss.

The most common eye diseases that result in peripheral vision loss are glaucoma and retinal detachment. Macular degeneration is a medical condition that primarily affects older populations and results in vision loss to the center of the visual field. This loss is due to the deterioration of the macula, which lies near the center of the eye's retina. The resulting degradation of visual acuity, or scotoma, may present itself as a partial vision loss or complete occlusion. Clear central vision is essential for observing fine detail, and central vision deficits affect performance in daily activities such as reading and walking [113, 168]. At present macular degeneration accounts for 8.7% of all blindness worldwide and is the most common cause of blindness in developed countries [506]. Alarmingly, the number of people worldwide with age-related macular degeneration alone is projected to rise to 196 million by 2020, advancing to 288 million by 2040.

3.1.1 Vision impairment affects perception

An estimated 3.2 million Americans are diagnosed with low vision, in which vision is impaired to the degree that it cannot be corrected with glasses alone [475]. Visual impairments may interfere with one's ability to interpret visual information, making many everyday tasks challenging. For example, central field loss, which is caused by diseases like age-related macular generation, can affect one's ability to read. While peripheral vision loss, which may be caused by glaucoma or retinal detachment, can affect one's ability to drive. Tasks like driving may even become dangerous.

Geruschat et al. [142] found that macular degeneration patients appear to have different gaze behavior in comparison to fully-sighted people during high-risk activities. In a traffic crossing study, Geruschat and colleagues [141] then compared traffic gap detection among pedestrians with normal vision, central vision loss, and peripheral vision loss. While their results suggested that all groups could identify crossable and uncrossable gaps accurately, there was a significant effect of low vision in measures of latency and safety. The study also found that decisions at the exit lane of a roundabout are more difficult than those at then entry lane. These experiments, done in the real world at an uncontrolled intersection with a handheld trigger as the indicator of deciding to cross are best viewed as complementary to the current work, which evaluates human behavior in virtual and augmented environments.

3.1.2 Studying vision impairment in MR

Virtual Reality provides an effective medium for the study of human behavior. The appeal of this technology lies in its capacity to control environmental factors. VR also finds applications in research where conducting an experiment may be too dangerous or infeasible for real-world execution.

For example, VR has been used to investigate pedestrian behavior at intersections [46, 312, 321, 421]. Real traffic scenarios present unnecessary risk to participants and are difficult to control with accuracy. For the visually impaired and other vulnerable populations, the ability of pedestrians to make street-crossing decisions may be compromised, resulting in unsafe decision making at these special

crossroads. This concern is motivated by real-world studies [19, 162, 163], which have established that blind individuals make poor gap judgments at traffic roundabouts. Accordingly, these populations have been of interest for behavioral analysis within the domain of pedestrian safety.

Hassan et al. [167] compared normally sighted, visually impaired, and blind pedestrians' street crossing decisions, and found that visually impaired participants' performance was as accurate and reliable as normally sighted participants. Unsurprisingly, blind pedestrians were the least accurate in making street crossing decisions. Hassan et al. [169] continued this investigation among elderly people with macular degeneration, elderly people with normal vision, and young normally sighted pedestrians. Again, no significant differences were found between macular degeneration and age-matched, normally-sighted pedestrians in street crossing decisions. However, the study found a risky tendency for macular degeneration pedestrians to make unsafe street-crossing decisions

The use of virtual reality in this work was critical. Virtual reality provides an environment in which a dangerous scenario — traffic crossing with visual impairment — can be investigated in a controlled and rigorous manner. Testing subjects with true visual impairment in true traffic situations is difficult, and some form of proxy is often used. For example, Geruschat et al. [141] used subjects with actual macular degeneration or peripheral vision loss at a live intersection with real traffic, but had them press triggers to indicate when they would cross rather than actually cross. In contrast, a VR study like the ones conducted by Wu et al. [511, 512] has considerably easier recruitment, ease of execution, no real danger, and can allow actual locomotion. This suggests that VR may be an beneficial tool to design and conduct effective research in perception and action for those with vision impairments.

To illustrate how mixed reality can be used to improve the lives of those with vision impairments in this section we focus a specific case: traffic crossing evaluations. However, traffic crossings provide only one example of mixed reality may be used to understand and design better, safer experiences for those with vision disabilities.

3.2 Vision Impairment Simulation

If developers are to design accessible products that people with disabilities can use, then they must understand how affected populations interact with their surroundings. Disability simulations allow individuals to experience what it is like to have a specific impairment. These simulations are not only informative for product and application design, but they can also help caregivers or medical professionals better understand the needs of those whom they care for.

Mixed reality provides a particularly unique opportunity for disability simulation. Within MR people may experience firsthand how specific disabilities, like visual impairments, affect the strategies that one employs to interact with their surroundings. With virtual reality, we may better understand how people with low vision or simulated low vision navigate hazardous situations, such as road crossings, to ensure that these spaces are designed with safety in mind [512]. And with augmented reality low vision simulations, we may evaluate how those with visual impairments interact with physical products to ensure that they are accessible to the widest audience possible. Figures 3.1 - 3.3 display an example of simulated vision impairment applied to a street-crossing scenario in VR.

Some immersive low vision simulations already exist. However, most are based on simplified symptoms of eye diseases [496] and are unable to produce the irregular scotomas that individuals experience in reality. Even fewer of these simulations have been implemented with eye tracking [244, 512]. Yet the biggest obstacle for the adoption of immersive low vision simulations for design may be that there are

Figure 3.1: View of street-crossing without vision loss [512]

Figure 3.2: Virtual roundabout [512]

Figure 3.3: View of street-crossing with vision loss [512]

no rigorous evaluations of their efficacy. In short, there are no complete, empirical studies that evaluate the ability of low vision simulations to replicate the impairments of those with low vision. This dearth of evaluations is likely due in part to the low latency demands of eye-trackers for effectively simulating low vision and to the difficulty of recruiting a sufficient number of people with visual impairments for system evaluations. As a result, most researchers have either only been able to conduct preliminary analyses [244, 496] or they have focused on applying their simulation to applications [512].

Virtual reality has been applied to simulate visual impairments for medical training and education purposes, as well. These simulations provide first-person experiences for medical professionals to better understand the daily difficulties encountered by patients. Ai et al. [7] and Jin et al. [210] simulated various forms of eye diseases in the context of a virtual apartment and received positive user response. They simulated macular degeneration through the application of an opacity mask and a wavy mask. Banks et al. [30] created a similar, specialized visual eye disease simulator for architects to view their designs through the perspective of a visually impaired onlooker. This work provides engineers with a better understanding of how to design public spaces for better accessibility and easier navigation. In the study presented by Lewis et al. [270], a Gaussian blur and distortion shader were applied to simulate macular degeneration. They also conducted an effectiveness test, which showed that using their visual impairment simulator improved users' understanding of visual impairments in general. Expanding the virtual microcosm, Vayrynen et al. [477] designed a navigation task amidst a city environment that allowed for participants to experience various visual impairments such as macular degeneration, cataracts, glaucoma, and myopia in a dynamic setting.

In augmented reality, various types and levels of visual impairment have also been simulated. Through the coupling of head-mounted displays and stereoscopic cameras, Ates et al. [21] and Werfel et al. [496] produced simulation tools to generate experiences using real-time video feedback. Both developments focus primarily on user experience, invoked empathy, and understanding as metrics for evaluation. While most studies recreate computational estimations of low vision experiences, an assessment tool was designed by Pamplona et al. [354] to capture retinal information from a high-contrast light-field display. Although this information is not displayed in real-time, it is able to create an accurate depiction of the visual occlusion experienced by a participating subject. Still, at present there are no rigorous, empirical evaluations of the efficacy of these immersive low vision simulations as compared to real low vision patients.

3.2.1 Analog simulation

Simulating vision impairments does not always require complex solutions. In fact, many experiments that evaluate the influence of visual acuity and restricted field of view on people's perception of space to date have relied on analog solutions. Often these simulations rely on degraded-vision goggles, such as googles with theatrical lighting gel applied [383, 384, 460, 519] or welding goggles in which field of view is restricted by cutting out a hole in black cardstock [33, 34].

A common approach for simulating loss in visual acuity is to use Bangerter occlusion foils (Ryser Optik, St. Gallen, Switzerland) Although Bangerter foils were originally designed to treat conditions like amblyopia (aka lazy eye), they have proven useful for simulating degraded visual acuity as well as contrast sensitivity [342, 366]. As such, they have been employed to simulate low vision in a number of studies [111, 266–268, 461]. Tarampi et al. [460] further demonstrated that certain types of theatrical filters can successfully decouple visual acuity loss from contrast sensitivity loss when visual acuity alone is of interest for simulation. Although several physical materials have been evaluated for recreating degraded visual acuity, there is a finite supply of possible materials to employ and the evaluation of these materials can be difficult.

Another challenge of analog simulations is appropriately restricting FOV—an issue which is candidly discussed in Legge et al. [266]—due to the problem of binocular overlap. As such, much of the prior research investigating the influence of visual field restrictions using analog approaches has been conducted with monocular rather than binocular viewing [33, 34, 266]. Monocular viewing is sometimes desired, like in the studies conducted by Rand et al. [383, 384]. In these studies, Rand and colleagues studied how depth judgments to targets were influenced by the visible horizon and by ground contact when visual acuity was degraded. The authors argued that binocular viewing was undesirable since it would introduce additional depth information that was not of interest for their evaluation.

The analog solutions for visual field loss discussed thus far have two limitations. They are unable to produce binocular visual field restrictions, a case which is more representative of real low vision patients' vision, and they are unable to connect FOV restrictions to eye movements. Instead, they rely on head-based FOV restrictions, which are a less accurate representation of visual field loss.

3.2.2 Digital simulation

The development of digital simulations for vision impairments has progressed in conjunction with technological improvements in computer graphics, displays, and eye tracking. Due in part to the rapid advancement of technology, most research publications on digital vision simulations describe system details and propose algorithms but few have sought to evaluate the efficacy of their systems to reproduce the perceptions and actions of those with real vision impairments.

Algorithms for accurate simulations of human vision began to emerge in the field of image processing in the 1990s. Initial investigations began with investigations of whether properties of human vision could be recreated in digital images. For example, Peli and colleagues used simulation to evaluate computational models for foveal vision [362] as well as peripheral vision [363]. Accordingly, much of this initial work focused on the development of image processing techniques to represent aspects of human vision and vision impairment.

In 2002 Perry and Geisler [367] proposed a system in which Goldmann perimeter data could be used to display scotomas over video images. Banks and McCrindle [30] then described how image processing techniques could be used to emulate the visual characteristics of vision impairments. They list a

number of eye effects that required simulation for effective modeling of impairments, including blur, opacity, color manipulations and more. Hogervorst and van Damme [182] then constructed their own simulation method for low vision by applying non-patient specific scotomas and glares to images.

A limitation to early proposed digital simulations is that they did not work in real time with eye tracking due to technical limitations of the time. Even so, researchers developed other ways to approximate people's eye position. Mankoff et al. [300] rendered opaque mask overlays relative to a mouse's position to simulate scotomas on desktop monitors. Vinnikov et al. [484] then developed a gaze contingent system to estimate real-time gaze direction on desktop monitors. However, gaze tracking in their system was limited to large head movements.

As external eye trackers improved in accuracy, researchers began to evaluate the effect of scotomas on people's perception in virtual desktop environments. Kwon and colleagues evaluated the influence of a simulated central scotoma (a circular opacity that subtended about 10° visual angle) in normally sighted individual's ability to perform visual search tasks in static images using an eye tracker mounted to a desktop monitor [252, 253]. They found that image enhancements improved search time and accuracy in older but not younger adults with simulated scotomas [253] and they revealed that people rapidly adopted a peripheral locus for fixation when their central vision was occluded by a simulated central scotoma [252].

Regrettably, even though desktop simulations are able to render vision impairments in real time, few empirical evaluations of these systems have been conducted [57, 252, 269, 270]. Kwon et al. [252], which compared visual search and chase task performance between simulated central scotomas and normally-sighted controls, is a notable exception. In their experiment, they found that oculomotor control with peripheral vision could be as precise and accurate as that with central vision when explicit training was conducted to compensate for simulated central vision loss. Although Lewis and colleagues [269, 270]. did not perform an assessment of people's behavior in their desktop vision impairment, they conducted user experience evaluations with expert reviewers (e.g., opticians and a visually impaired consultant) using pre-post questionnaires. Further, people's behavior with simulated vision impairments were not compared against real patients with real vision impairments in any of the mentioned evaluations.

3.2.3 MR HMD simulation

A major appeal to simulating vision impairments in MR head-mounted displays is that they can provide an immersive experience of what someone with vision impairment experiences. Another potential benefit of HMDs is that they can integrate eye-tracking into their simulations to create more accurate representations of visual field restrictions. However, performing accurate eye tracking within the dark chassis of an HMD is a non-trivial task. Unlike the eye trackers employed with desktop virtual environments, eye tracking cameras integrated within the chassis of an HMD are surrounded in darkness and they can be jostled when head movements are executed [280]. These environmental factors can interfere with a camera's ability to quickly and accurately detect a user's eye position. However, both speed and accuracy are important for tracking eye movements.

Saccades—which are rapid eye movements between fixation points—can encapsulate a wide range of eye movement distances and durations. For example, their extent can measure less than a degree in visual angle and as large as 90 degrees visual angle [170, 200, 265]. Saccade durations similarly encapsulate a wide range of values, and they can be greater than 100 ms for large saccades [43, 143]. More subtle eye movements, like the saccades one executes when reading, can be as fast as 20 to 40 millisec-

onds [387]. The speed of saccades and other eye movements has ramifications for how much delay is tolerable in gaze-contingent displays. For example, prior research on the influence of latencies in gaze-contingent displays has shown that delays greater than 45 ms may disrupt fixation behavior [286]. However, the degree of eye tracking latency that is detectable to a viewer is also related to where it occurs within a person's visual field. To seamlessly mask central vision eye tracking experiments have provided evidence that a system's latency should be less than 25 ms to take advantage of vision suppression during eye movements [308, 399, 423]. In contrast, latencies up to 60 ms are often not detected outside of central vision [287].

In recent assessments of eye tracking in the HTC Vive Pro Eye—a commercial MR HMD released in 2019—Sipatchin et al. [430] evaluated the temporal precision of its eye tracking to be approximately 58.1 ms. Stein et al. [439] then conducted an evaluation of eye tracking latencies across multiple commercial VR HMDS, including the HTC Vive Pro Eye, the Foveo, and the Varjo XR-1. In their evaluation, they found that the eye tracking latency of the HTC Vive Pro Eye was significantly worse than the other two systems with an eye tracking delay of 50 ms. Both research groups argued that the eye tracking latencies shown in this display were problematic for gaze-contingent MR—an argument which is supported by prior research on eye tracking latency [286, 406]. Stein et al.'s eye tracking assessment determined approximate eye tracking delays of 35 ms in the Varjo VR-1 and 15 ms in the Fove-0. The authors argue that the latencies exhibited by the Fove-0 and Varjo VR-1 may be more suitable for gaze-contingent experiments [287].

In regard to vision impairment simulation, Sipatchin et al. [430] explain that a system that exhibits notable eye tracking delay (e.g., a 58.1 ms delay) may be able to simulate peripheral field loss but simulating central visual field loss would require improvements. A smaller latency is required for manipulations of central vision like the ones employed to simulate central field loss since latencies that are longer than saccade durations can dissociate the scotoma from a person's true eye position [406]. In behavioral assessments of vision simulations, this can result in participants "peeking" around scotomas [430], which can interfere with experimental control.

Using the HTC Vive Pro Eye, Krösl et al. published a series of investigations into how specific vision impairments may be simulated with eye tracking in virtual environments. In their initial system, Krösland colleagues [241] modeled reduced visual acuity and they evaluated its impact on people's ability to assess maximum recognition distances (MRD) of escape-route signs. They then extended their system to create eye-tracked simulation of cataracts—called ICthroughVR [243]—and they again evaluated people's ability to assess MRDs of escape-route signs. In both studies they found that the presence of simulated vision impairments, in this case for reduced visual acuity and cataracts, negatively affected people's ability to correctly assess MRDs. Although eye tracking latency was not measured, the authors comment that their proof-of-concept eye-tracking integration exhibited noticeable lag. They also reported more errors when participants wore glasses.

Direct evaluations between those with vision impairments and those with equivalent simulated impairments can be difficult and is often not possible. However, when vision impairment is correctable via surgery, as is the case with cataracts, direct comparisons can be done. Accordingly, Krösl et al. [244]. conducted a pilot assessment in which five cataract patients were able to provide feedback about how accurately the simulation represented their cataract condition. With one eye corrected and one eye still affected by cataracts, participants were able to view either the simulated cataract or their corrected vision in one eye. The authors recruited five participants with cataracts to participate in the evaluation of their vision impairment simulation system (the CatARact system), which was a notably improved version of their original cataract simulation system (ICthroughVR) [243]. The updated system also

permitted video see-through AR experiences. The qualitative feedback acquired from Krösland colleague's assessment stressed the importance of representing the heterogeneous and diverse symptoms of vision impairments for an accurate simulation of cataracts.

Using the Fove-0, Jones et al. [215, 216] developed the Open VisSim to simulate vision impairments with eye tracking in VR. They also extended their system to AR by deploying their software in a modified HTC Vive that was integrated with Tobii eye-tracking [216]. A Zed Mini stereoscopic camera was affixed to the front of the display to provide video feed input for AR. The Open VisSim produces visual blur, spatial distortions, and color vision deficits with eye tracking. It also provided support for a data-driven approach for producing visual field restrictions via perimetry data like the system proposed by Perry and Geisler [367]. However, they describe their system as one that runs in "near real time", indicating that the current version of their system may suffer from inaccuracies or temporal delays. The actual accuracy of vision tracking in their system is not reported. Although eye tracking latency was not explicitly measured, both Krösl et al. [243, 244] and Jones et al. [215, 216] noted latency issues in their systems.

3.2.4 Limitations

There are limitations to what these simulations can elicit. For example, they cannot emulate the adaptive strategies that those with disabilities have learned to employ over time nor can they recreate sociohistorical context that those with disabilities experience. Disability simulations are invaluable design tools, but they are no substitute for input from real individuals with disabilities [469]. As such, disability simulations can provide invaluable tools for expediting the design process for accessible applications, but before a product is released developers should consider consulting real individuals with vision impairments for feedback.

Further, due to the difficulties of integrating eye tracking into head-mounted displays, many of the HMD-based systems for vision impairment simulations to date do not use eye tracking in their solutions [21, 103, 241, 293, 446, 477, 496]. Even when vision impairment simulations have been adopted for use in head-mounted MR displays, many of these publications focused on describing technical systems while forgoing evaluations [21, 215, 242] or while conducting only informal evaluations [103, 430, 446, 496] Although these system descriptions are important for the development and improvement of technical solutions for simulating vision impairments, evaluations of their efficacy are needed for confident adoption in MR HMDs. Some evaluations have been conducted with normally sighted individuals alone in which participants experience normal vision and simulated vision impairments [216, 241, 243, 477]. However, to the author's knowledge, Krösl et al. [244] is the only HMD-based simulation that has evaluated vision impairment with real patients. Krösl et al.'s work specifically evaluated the how a single vision impairment, cataracts, could be simulated.

3.3 Perception of Visual Space with Vision Impairment

A notable amount of research has been directed towards understanding the influence of vision impairment of view on people's ability to perceive space. An inability to correctly perceive space can also pose risks to populations with low vision. For example, vision impairment is associated with an increased risk of falls in elderly patients with low vision conditions [285, 404]. In a study of 148 older individuals (aged 63 to 90), Lord et al. [285] found that impaired depth perception, contrast sensitivity, and low-

contrast visual acuity were the strongest risk factors for reported falls over a 12 month period. They conclude that adequate depth perception is an important factor for detecting hazards and maintaining balance in an environment.

In the following section, we discuss how prior research has evaluated the influence of visual acuity, peripheral visual field loss, and central visual field loss on people's depth perception using both real and simulated vision impairments.

3.3.1 Visual acuity loss

Loss of visual acuity can interfere with people's ability to correctly perceive the position of objects in space. For example, the appearance of everyday surfaces like steps and ramps along a sidewalk can be obscured. In a study evaluating the ability of people with simulated visual blur to detect steps and ramps, Legge et al. [268] found that visible edge contours in an environment were influenced by visual perspective. The authors found that participants were more easily able to identify steps up as opposed to steps down due to differences in luminance contrast between the riser of a step and its contiguous surface planes [268]. In addition, the height of the element in the visual field influenced people's ability to correctly identify these surfaces in space. The bounding contour between the far edge of the target (the top of a step or ramp) and the wall behind it played an important role in determining whether the ramps and steps were positioned up or down. Bochsler et al. [53] later extended this work to real low vision patients. They found that people with simulated visual acuity and people with low vision made similar responses.

Because the distal horizon as well as the height of an object relative to the ground influences depth perception judgements [314, 315], an inability to correctly register ground surface information can interfere with an individual's ability to correctly judge distances to targets.This problem is related to the horizon-distance relation suggested by Sedgwick [417], in which the visual system utilizes the angle of declination from a line-of-sight parallel to a target located relative to the ground plane in order to determine absolute distances. Support for the horizon-distance relation has been found through empirical studies in which the appearance of the horizon was manipulated. For example, Meng et al. [314, 315] demonstrated that when the horizon is artificially raised, angle of declination increases, and distances are overestimated. Similarly, Rand et al. [384] demonstrated that, even when visual acuity is artificially degraded with vision goggles, people are able to make accurate distance judgments to targets along the ground in action space (2m -30m) using architectural features within a space. However, participants overestimated distances to targets when the position of the horizon was artificially manipulated [384].

In follow-up work, Rand and colleagues [383] tested people's distance judgements to targets positioned both on and above the ground. Although distance judgements remained accurate under reduced visual acuity when an object was positioned above the ground on top of a visually detectable stand (a stand with high color contrast against the background), distance judgements to targets on an undetectable stand (a stand with high color contrast against the background) were judged as farther away. This response pattern is in line with prior research which stresses the importance of surfaces and surface contact for determining distances to targets in space [332, 352]. The overestimation of distances in this work is explained by the absence of visual information about surface contact to link objects in space to nearby surfaces. In the absence of surface contact cues, people rely on optical contact to determine the position of an object in space [333, 352, 397], where optical contact refers to the location at which the projected image of an object contacts the image of the ground beneath it. As a result, targets positioned above the ground appear to be on the ground but farther away rather than floating. This

concept is intimately related to Gibson's ground theory of spatial perception, which we discussed in Section 2.3.1.

Rand et al. then conducted a second study in which floating targets were positioned at various heights in the visual field, relative to the horizon [383]. When objects were positioned above the horizon, they were perceived as larger in size than either targets placed on the ground or targets placed above the ground below the horizon. Because all targets were of the same size and all targets were positioned at the same distance from the viewer, these results suggest that participants perceived floating targets to be farther away when they were located above the visible horizon. The authors argue that size is an indirect measure of distance, following the size-distance invariance hypothesis (SDIH) [147, 230].

Because the position of objects relative to the horizon can perturb distance judgements to targets in space when vision is artificially degraded [383, 384], special attention to how virtual objects are connected to nearby surfaces when they are floating may be necessary for ensuring accurate spatial perception for people with low vision in MR displays. This is further reinforced by prior research, which has shown that normally-sighted individuals perceive floating targets as farther away in AR [397]. Although the presence of shadows, which provide a powerful cue for connecting objects to nearby surfaces, [2, 5, 295, 467] can help improve distance judgements to floating targets in AR (Section 6). It may be necessary to use graphics with higher contrast to ensure shadows are visible, since people with reduced visual acuity are unable to leverage visible cues that connect targets to the ground when they are not visible [384]

3.3.2 Visual field loss

Artificially restricting an individuals' visual field is a strategy that has been applied to understand human perception in contexts outside the study of vision loss. Field of view restrictions have been used to provide experimental control for the study of visual space perception [95, 509, 510] and to better understand how visual field restrictions caused by virtual and augmented reality HMDs influences depth perception [64, 95, 212, 233, 501]. The majority of this research has investigated peripheral visual field loss.

Peripheral Field Loss

To determine how people use ground surface information to accurately perceive distances, Wu et al. [510] manipulated people's peripheral field of view and head motion. In their first two experiments, they found that distances were underestimated when peripheral field of view was reduced to 21.28° x 21.28° and 13.98.28° x 13.98° and head motion was prevented. In their third experiment, they evaluated people's distance judgements when either horizontal or vertical field of view was variably restricted and head movement was prevented. They found that when vertical field extents were reduced to values less than approximately 21°, distances were underestimated. However, horizontal restrictions did not result in significantly different judgements from their control condition, which permitted full view.

The authors argued that the results of the experiments underscore the importance of near ground information for accurate distance judgements [144, 418]. Finally, to better understand the role of near-ground (<2m) and far-ground information, Wu and colleagues [510] evaluated blind walking distance judgements when near-to-far and far-to-near head movements were permitted. While near-to-far scanning remained accurate, far-to-near scanning resulted in distance underestimation. The results of the experiments conducted by Wu and colleagues [510] underscore the importance of near ground infor-

mation for accurate distance judgements. Wu and colleagues argue that the visual near-ground provides a valuable anchor for the construction of accurate visual representations of space, an argument that is consistent with related research that has shown the ground is used as a frame of reference when judging absolute distances [95, 427].

In related work, Creem-Regher et al. [95] found that restricting a person's view of their own body and the floor beneath them (out to approximately 1.5m) did not result in different distance judgments as measured through blind walking. However, when they evaluated the influence of head motion when field of view was restricted, they found that distance estimation performance with a virtual field that was restricted to approximately 38° x 32° did not differ from their control (a full visual field condition) when head rotation was permitted. Restricting head motion and field of view, however, led to distance underestimation. This finding is particularly interesting when compared to similar research in which occlusions of portions of the ground surface between a viewer and a target in space resulted in egocentric distance underestimation [171, 314, 509]. These conflicting results may be evidence for more complex processing of visual information for distance judgments.

In assessments or distance perception where participants are permitted to locomote and extensively view targets without restricted head motion, differences in distance estimation responses between those with normal vision and those with restricted peripheral vision do not always occur [126]. In such evaluations, researchers have argued that people may learn to compensate for any distance compression effects incurred by loss of peripheral vision [510].

Central Field Loss

The majority of the research discussed thus far evaluates the influence of peripheral field restrictions on people's ability to accurately perceive space. Further, unlike in the case of peripheral field loss research, much of the prior research conducted to understand the influence of central field loss on people's spatial perception has been conducted with indirect measures of distance judgments, such as grasping [111, 355, 482], traffic crossing[141, 169], and fall risk [285, 489].

Prior research has argued that low vision conditions, especially conditions that cause central vision loss, adversely affect depth perception by interfering with stereopsis [71, 482]. Stereopsis is a particularly important depth cue for judgements of distances in near space (< 2 meters) [451, 452]. In a study evaluating low vision patients' grasping judgments to targets, Melmoth et al [313] found that without binocular stereo vision, participants made less accurate grasp selection and they were more likely to rely on nonvisual information. Verghese et al. [482] then conducted a similar study in which patients with central field loss were less accurate at grasping targets when stereopsis was impaired. These assessments are informative for understanding how people with central field loss perceive depth at near distances but open questions remain regarding how people with central field loss perceive distances at farther distances.

Vision simulation can be helpful for studying the influence of central scotomas on depth perception. However, due to the difficulties of simulating central scotomas (See Section 3.2, less research has been conducted to understand how central field loss affects people's perception of space when central field loss is simulated.

Some work in VR HMDs has been conducted with simulated macular degeneration conditions in road crossing scenarios. Wu et al. [512] simulated a central scotoma represented by opacity and blur to study how central field loss affected people's ability to detect gaps in traffic when crossing the street. In their study they found the people with large, central visual scotomas simulated scotomas (approxi-

mately 20 – 40° in this study) selected longer gaps between vehicles in traffic and they waited longer to initiate crossings. This response pattern is similar to what has been observed in real patients with macular degeneration in traffic crossing scenarios [141, 169].

Some work studying the influence of simulated central scotomas has been conducted using desktop virtual environments. However, these experiments focused on how simulated vision impairments influenced visual search performance. Kwon et al [252, 253] used circular disks that covered about 10° visual angle to understand how individuals with macular degeneration performed visual search. Interestingly, they found that when individuals were exposed to central vision impairments, they quickly learned to compensate for this vision loss by adopting a peripheral locus for fixation when performing visual search. Further, participants with central field loss in Kwon et al. [252] were able to perform the task as accurately as those with unobstructed central vision with training.

Part III

Surface Contact Perception

Shadows appear to me to be of supreme importance in perspective, because, without them opaque and solid bodies will be ill-defined

Leonardo da Vinci

4 Surface Contact Perception in MR

In this chapter we briefly review the theoretical motivation behind our decision to investigate shadows' influence on surface contact perception across virtual and augmented reality displays (Aim 1). Because visual space is not defined by arrays of objects in empty air—but by the layout of surfaces, sets of adjoined surfaces, and entities that are arranged in relation to surfaces [145, 146]—some visual element must exist that connects objects to surfaces in space. Shadows naturally, yet often subtly, accomplish this important task. However, in immersive technology we do not yet know how much the shadows rendered by virtual and augmented reality displays must match those observed in reality for accurate spatial perception.

4.1 Motivation and Goals

A common complaint from users of augmented reality technology is that virtual objects appear to "float" within real world scenes. A detached—or "floaty"—appearance is a strong indicator that the depth cues provided by virtual stimuli are insufficient to accurately locate the stimulus' position within a real world environment. Given that the human visual system integrates information from a variety of cues to interpret depth, the presence of unreliable depth cues can cause unstable depth perception when these cues are not combined in a consistent manner [114, 257]. Drascic and Milgram [107] as well as Adams [1] have pointed to cue conflicts as a potential factor contributing to inaccurate depth perception in augmented reality.

All visual information is produced by structured patterns of light and shadow. As such, one of the most salient visual cues for the layout of objects within a scene is the cast shadow [297]. Cast shadows may be defined as holes in light that occur when an opaque or semiopaque object blocks the light that falls onto a surface [76]. By providing relative position information between an object and the surface upon which an object's shadow rests, cast shadows provide important depth information in both the real world [297, 514] and in augmented reality [255, 449]. As such, a significant body of work has been conducted on how to best render shadows in graphical displays [279].

According to Gibson's ground theory of space perception, one's perception of space is defined by the layout of surfaces, and the position of an object in space is defined by its relationship to surfaces [145, 146]. Within this framework, cast shadows function as a "visual glue" to attach virtual objects to surfaces [295, 467]. Furthermore, it has been demonstrated that people become more accurate when estimating egocentric distances to objects placed above the ground when they are clearly connected to the ground via shadow [333]. However, it is unclear how to best create this visual glue for augmented reality devices, especially for those devices that rely on additive light displays such as optical see-through devices. Given that this type of AR display cannot remove light—and thereby darken—virtual or real

Figure 4.1: A VR HMD with VST capabilities: the HTC Vive Pro with mounted Zed Mini camera

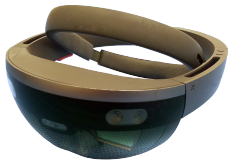

Figure 4.2: An OST HMD: the Microsoft HoloLens 1.

objects, rendering shadows in these devices is a challenge. We are therefore also interested in assessing how non-photorealistic shadows (cf. [62, 485]) affect perceived visual glue and depth perception.

Recent work on optical see-through AR displays leverages the human visual system to create shadows [203, 299]. These devices create the illusion of dark color values using simultaneous contrast illusion, which approximates a shadow. Other methods have been used for video see-through AR displays [337], and shadow generation in virtual reality (VR) is considered mature [108, 433]. Nonetheless, for generally deployable MR applications, it would be desirable to have a clear and unified understanding of which techniques work across widely available technologies. Therefore, in the present work we investigate how surface contact is affected by cast shadows across a set of display types.

4.2 General Methods

In order to evaluate how shadow shading methods affect a viewer's certainty in estimating ground contact in head-mounted extended reality displays, we conducted four experiments across three unique display conditions: an optical see-through augmented reality display, a video see-through augmented reality display, and a virtual reality display. The subsequent sections discuss the technical setup of our experiments (Section 4.2.1) as well as the specific solutions used for rendering shadows (Section 4.2.2) and positioning virtual objects based on viewing angle (Section 4.2.3).

4.2.1 Materials

We employed three immersive HMDs for our investigations. We used the Microsoft HoloLens 1 for our optical see-through display condition, and a wireless HTC Vive Pro was used to render the virtual reality scene. The same Vive Pro was used in conjunction with a Zed Mini stereoscopic camera for the video see-through display condition. Head tracking was used across all conditions to allow natural viewing of experimental stimuli. Applications for each device were developed in Unity version 2017.4.4f1 with the C# programming language.

The Microsoft HoloLens 1 has an approximate per eye resolution of 1268×720 and field of view of $30° \times 17°$. Although the augmented field of view (FOV) of the HoloLens is narrow, outside of this viewing area users' vision is not occluded by the device. This OST display relies on additive light to render images and is therefore unable to render black color values. For our experiments, position tracking was performed using the HoloLens' native inside-out tracking solution.

The virtual reality environment was rendered using a wireless HTC Vive Pro, which has a maximum per eye resolution of 1440×1600 and an approximate field of view of $110° \times 113°$. Position tracking

for this condition was performed using the Vive's lighthouse tracking system. In addition, because this condition relied on completely virtual imagery, a virtual model of the real world environment was created. This model included photographed images of the real room along the walls and custom 3D models designed to match the table and foam floor tiles present within the real world environment. An image of the virtual environment can be seen in Figure 4.4. An image of the real world environment can be seen in Figure 4.3.

The video see-through device also relied on the HTC Vive Pro for rendering. However, it created an augmented reality environment by combining virtual overlays with real time video footage, which was captured using the Zed Mini stereoscopic camera system. The Zed Mini was affixed to the front of the head-mounted display. The use of the Zed Mini's camera feed constricted the Vive's resolution and field of view to 1280×720 and $90° \times 60°$, respectively. Position tracking was performed using the Zed Mini's inside-out tracking solution, which integrated with the HTC Vive's tracking system.

An HTC Vive tracking puck was used to position virtual target objects within the real world environment for both the VST AR and VR conditions. At the start of an experiment and between each experimental block, the tracking puck was placed near the viewer at predetermined positions in the physical room, and virtual objects were rendered at the puck's position virtually. Similarly, for the OST AR system, this position calibration involved placing a virtual HoloLens spatial anchor at the same predetermined positions.

For the the VST AR and VR conditions, selections were performed using a wireless mouse. With the Microsoft HoloLens, users selected inputs using the HoloLens' clicker in Study 1. However, because a two-alternative forced choice paradigm was used for Studies 2-4, inputs were performed with a mouse for all three HMDs after Study 1. For all experiments, a gaze-directed paradigm was used to guide selections. A small gaze cursor appeared at the center of the user's vision whenever their forward head orientation pointed towards a user interface element within the HMDs. However, this cursor disappeared when viewing target objects so as not to disrupt their vision when evaluating surface contact.

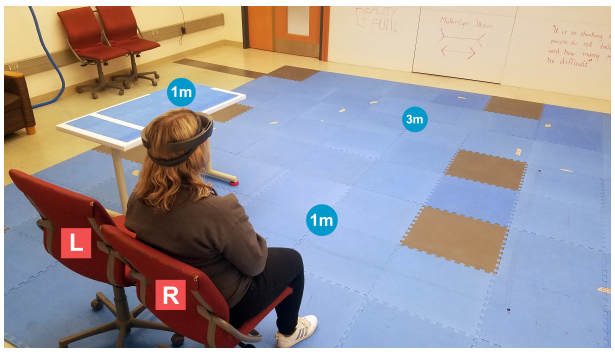

Figure 4.3: A participant views experimental stimuli in the Microsoft HoloLens. The image marks the left (L) and right (R) chairs as well as the three distance conditions: 1m table, 1m floor, 3m floor.

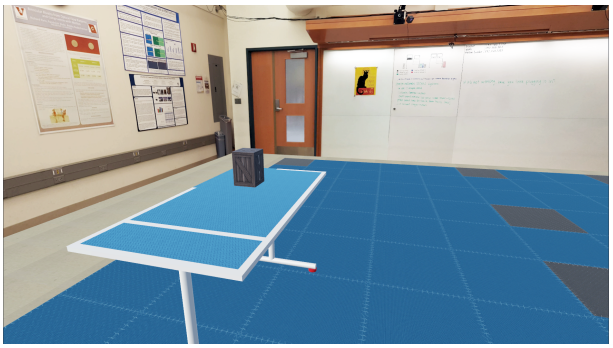

Figure 4.4: A screenshot of the virtual environment from the user's perspective is displayed. A target object is placed 1m away on a nearby table. The target is rendered without a cast shadow.

4.2.2 Shadows

Shaders to render three distinct hard shadows were programmed using a variant of the HLSL language that is compatible with the Unity game engine. A directional light was positioned so that a target ob-

ject's shadow would lie behind and to the right of the object. To accomplish this, the orientation of a virtual, directional light was set to 141° along the x axis and −141° along the y axis. A depiction of the three shadow conditions for each device in Study 1 can be seen in Figure 5.1. It should be noted that the images displayed in the figure do not perfectly match those presented by the immersive HMDs since there are display and capture differences. For example, the HoloLens screen capture relies on a monocular video feed for image capture but the actual user only experiences AR through the stereoscopic optical see-through display.

The dark gray shadow condition represented the most traditional method. It rendered a dark color value within the umbra of the shadow and therefore created a perceptually valid impression of a shadow for most devices.

The white shading condition, which added white light to create a shadow instead of subtracted, represented a photometrically incorrect shadow. Accordingly, it was also the most perceptually incorrect, or non-photorealistic, shading method included in this study.

The gradient shadow condition was designed as another perceptually correct method–especially for OST devices, which are unable to render black. Our gradient shadow method used simultaneous contrast to change the visual appearance of two adjacent colors and give the illusion of a dark shadow by rendering light outside of the shadow's umbra. In our method, the intensity of the light near the edge of the umbra also gradually decreased as the distance from the shadow increases, which created a gradient of light along the ground surface. Within the shadow's umbra nothing was rendered.

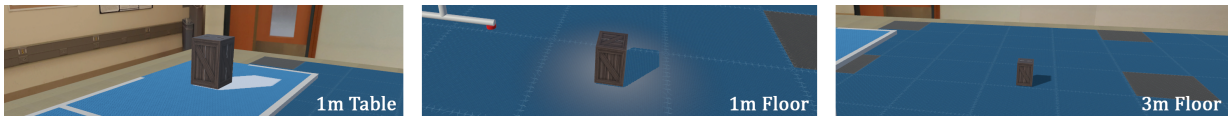

Figure 4.5: Target objects in the above ground condition for Experiment 1 are positioned at three distances from the user's perspective. The white shadow, gradient shadow, and gray shadow conditions are displayed from left to right.

4.2.3 Vertical displacement

For both studies, in order for participants to judge if targets were in contact with a surface, stimuli had to be presented both on and slightly above the ground for discrimination. Because we evaluated surface contact judgments across multiple distances, we displaced each target vertically based on viewing angle to ensure fair comparisons across distances as much as possible. Participants also viewed stimuli while seated throughout both experiments for consistency. The average eye height of the viewer was calculated by summing the average eye height of a person while seated (i.e., the distance from their bottom to their eyes while seated) from Harrison et al. [166] and the height of the seat of the chairs used in our setup, which resulted in a value of 1.171 m for h_e. In both experiments, some or all stimuli were placed on a table in front of the user (See Figure 4.3). For these conditions, h_e was adjusted to account for the table by subtracting the table height (0.7461 m) from this value.

Using the average eye height of a viewer, denoted as h_e, and the distance to a given target, d_t, we were able to solve for a series of three triangles from which we could extract the degree of vertical displacement, d_v, for target objects placed above the ground. Eye height was calculated by adding the average sitting eye height [166] to the height of the chair used in our study, which resulted in $h_e = 1.171$. In

both experiments, when target objects were placed on a nearby table, instead of on the ground, we subtracted our table height from this value to obtain h_e. The trigonometric formulas used for this calculation are shown in the equations below:

$$\sigma = \tan^{-1}\left(\frac{d_t}{h_e}\right) + \omega \tag{4.1}$$

$$d_v = \left(\frac{\tan(\sigma)\ h_e - d_t}{\tan(\sigma)}\right) \tag{4.2}$$

For equations 4.1 and 4.2, ω represents the degree to which viewing angle was modified and σ represents the updated viewing angle to the vertically displaced target object. Figure 4.6 shows each variable in context for clarity. In Study 1, a viewing angle of 0.3° (ω) was selected since it was the height at which people could discern that an object was off the ground more than half the time during preliminary testing. Screenshots of experimental stimuli in the above ground condition can be seen in Figure 4.5. In Studies 2, 3, & 4 multiple vertical displacements were used.

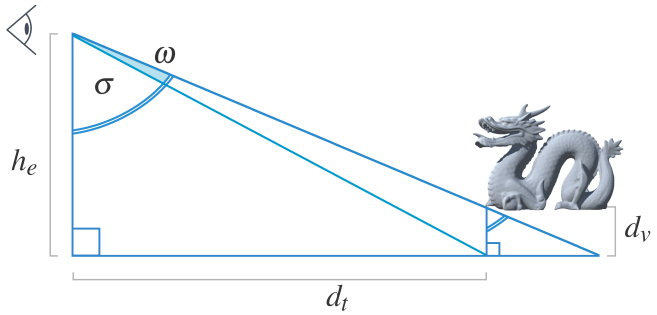

Figure 4.6: Visual depiction of trigonometric solution for vertical displacement.

Shadow is a colour as light is, but less brilliant;
light and shadow are only the relation of two tones.
Paul Cezanne

5 Investigating Surface Contact Perception in MR

In this chapter we begin our investigation into how the manner in which graphics are rendered can influence one's perception of space in mixed reality displays. Of particular interest in this chapter is how realistic and non-photorealistic rendering methods depth perception (as evaluated through surface contact perception). Although we began this work motivated by theories of cue integration [208, 256], which suggest that more consistent, realistic visual depth information between virtual and real objects will result in more accurate depth perception [1, 107, 245], our initial evaluation (Study 1) on shadow rendering and surface contact perception provided evidence to the contrary.

Specifically, in our first investigation we found that people's confidence ratings for surface contact with a non-photorealistic rendering approach, in this case a white shadow, resulted in higher ratings than any other condition in an optical see-through (OST) and a video see-through (VST) AR display. Spurred by this curious experimental outcome, we conducted three followup investigations into the influence of realistic and non-photorealistic rendering approaches on surface contact perception (Studies 2,3, & 4). The development and execution of this work motivated Aim 1 of this dissertation. Further, it provided the groundwork for our future investigations into the influence of realistic and non-photorealistic shadows on depth perception in persons with normal vision (Aim 2) and persons with image impairments (Aim 3).

In our research we focus on evaluating a specific depth cue: cast shadows. Shadows function as "visual glue" to create a point of contact between an object and nearby surfaces[295, 467]. Because creating connections between objects in space and surfaces within an environment is important for an accurate perceptions of spatial layout and depth [144, 145], we focus our investigations on this depth cue. For more discussion on the role of surfaces and the role of cast shadows in spatial perception, see Section 2.3. Another reason we focus on cast shadows for these evaluations is that evaluating a specific depth cue allows for better interpretations and comparisons of empirical results across the studies reported within this dissertation.

Over four studies, my collaborators and I evaluate people's sense of surface contact given both realistic and non-photorealistic shadow shading methods in OST AR, VST AR, and VR head-mounted displays. We manipulate cast shadow shading, object shading, object geometry, and object orientation to better understand what characteristics of virtual objects and their cast shadows influence the accuracy of surface contact perception. We also conduct our research across multiple HMDs to determine what rendering guidelines would generalize across HMDs that use different approaches for rendering. At the onset of this investigation, we anticipated that rendering approaches for cast shadows would benefit surface contact judgments differently between OST AR, VST AR, and VR displays due to the unique optical and graphical properties of each display type. Contrary to our expectations, surface contact judgments in the two AR displays often benefited from the same, non-photorealistic rendering

approach. This outcome encouraged the investigation of different rendering techniques across OST AR and VST AR displays in our later investigations of egocentric depth perception in Aim 2 of the dissertation (Chapter 6).

The results of our research encourage the use of non-photorealistic rendering approaches for cast shadows to enhance surface contact judgments, which is an outcome that may prove beneficial for people with vision impairments: a population that often benefits from high contrast imagery like that provided by the non-photorealistic cast shadow condition evaluated in these studies. This outcome also contributed to the development of the final research aim of this dissertation (Aim 3) in which we evaluate depth perception with real and simulated vision impairments.

The results of our investigations into surface contact perception have resulted in two publications. The first, Adams et al. [5], covers material discussed in Studies 1 and 2. These experiments cover our initial investigation into surface contact perception as well as our investigations into the influence of how contrast between an object's shading and the shading of its cast shadow may have influenced our initial research findings. The second paper, Adams et al. [2], covers material discussed in Studies 3 and 4. In these two evaluations we evaluate the effects of object geometry and orientation on people's surface contact judgements. Throughout these investigations of different object properties influence on surface contact perception, we evaluated both realistic and non-photorealistic cast shadows.

5.1 Study 1 - The way shadows are rendered influences perception of surface contact

In our first study, we evaluated how the manner in which we render shadows affected ground contact perception across multiple distances. Furthermore, we evaluated stimuli in our first experiment across multiple distances to ensure that any effects we found were due to our stimuli and not due to the use of a specific viewing angle, since depth cues can vary in effectiveness depending on the distance of information from the viewer [6, 96].

We anticipated that rendering shadows that were more consistent with the real world environment–and therefore more perceptually valid–would improve ground contact perception in AR devices. Accordingly, in Study 1 (Section 5.1), we evaluated a variety of shadow shading methods, which included both perceptually motivated methods and a photometrically incorrect shading method. Curiously, our predictions were proven wrong and the photometrically incorrect shading method had the most pronounced effect on people's certainty when estimating ground contact in AR. Based on our previous discussion of the importance of cast shadows as a cue for surface contact to inform depth perception, we designed our first study to test several hypotheses:

H1: Since prior research has shown that cast shadows are an effective cue for establishing ground contact, we anticipated that the presence of a shadow would significantly affect ground contact perception.

H2: In addition, given the unique display properties of our two augmented reality devices, we anticipated that the ability of a shadow to create a sense of ground contact would vary depending on the shading technique employed to render it in AR. A priori, we did not anticipate significant differences in ground contact perception for our virtual reality HMD condition as the completely virtual scenes generated by this device benefited from rich and consistent depth cues.

H3: Finally, we anticipated that perceptually valid shading methods would be more beneficial in discerning ground contact over a photometrically incorrect shading method since this method would

better match the real world cues for depth given by shadows.

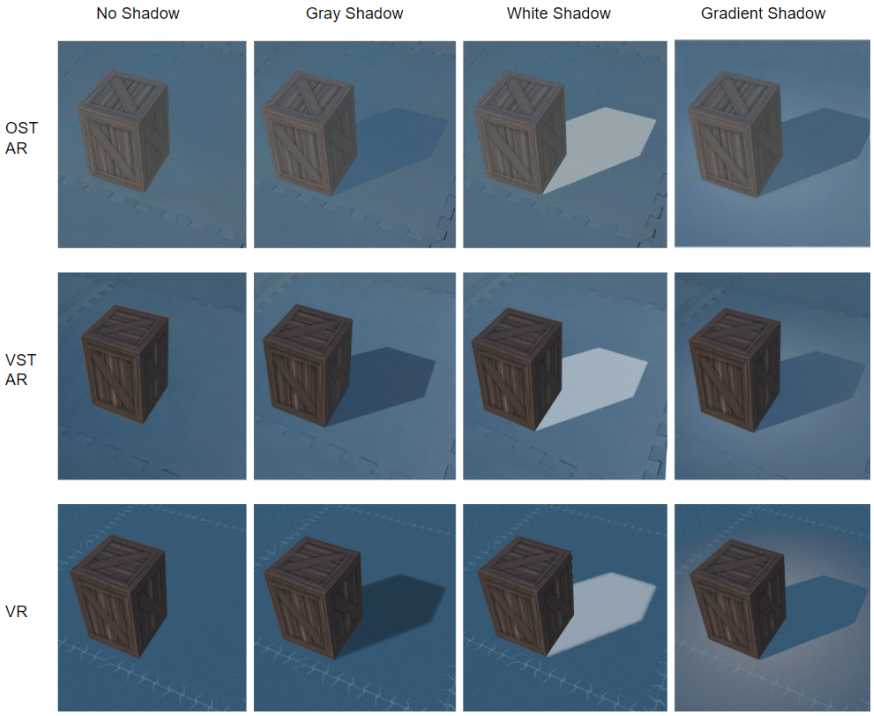

Figure 5.1: Close up images of the experimental stimuli used in Experiment 1 are displayed for each MR device. All stimuli are presented on the ground in this image. The same shaders for cast shadows were used across devices for the (1) no shadow, (2) gray shadow, (3) white shadow, and (4) gradient shadow shading conditions. OST AR images were captured with the HoloLens' native mixed reality capture feature, which relies on video input. VST AR images relied on the Zed Mini's video feed. Color correction has been applied to both augmented reality images to better match what participants saw during the experiment.

5.1.1 Participants

Thirty six individuals (27M, 9F) aged 18–32 from Vanderbilt University were recruited to participate. All participants had normal or corrected to normal vision, and each person was offered donuts for 40 minutes of their time. Our experimental methods were approved by the local institutional review board, and written consent was obtained from all subjects prior to participation.

5.1.2 Design

Our protocol was modeled after the one used by Madison et al. [295] in which participants were asked to rate their confidence in perceived contact between a target object and the surface beneath it when virtual stimuli were presented either above or in contact with a surface. For both this prior study and our current research, participants rated their certainty in perceived ground contact for each stimulus using a 5-point response format with verbal anchors where values mapped to: *(1) definitely touching, (2) maybe touching, (3) unsure, (4) maybe above,* and *(5) definitely above.* The input prompt used in the current work is shown in Figure 5.2.

We used a mixed factorial design for this experiment. Specifically, a 3 (display type) x 4 (shadow type) x 3 (distance) design, with head-mounted display type (OST AR, VST AR, or VR) as a between-subjects variable and shadow type (none, gray, white, or gradient) and distance as within-subjects variables. Section 4.2.2 discusses the shadow types in further detail. Targets were placed at distances of 1m away from the viewer on a table, 1m away from the viewer on the ground, and 3m away from the viewer on the ground.

To mitigate viewing order effects, our experiment was blocked with respect to distance condition, and the order in which each distance condition was presented was counterbalanced across subjects. Within each block the order of presented stimuli was randomized without repeating within a series of 16 trials. This coincided with an experiment interruption in which participants were prompted to indicate if they required a break.

Within each display condition, participants were exposed to four shadow sub-conditions across three distances. Because we evaluated four shadow shading conditions–that were presented on or above the ground across three distances–there were 24 unique stimuli in total. Additionally, each unique stimulus was viewed 10 times, making the experiment consist of 240 trials total. We used a repeated measures design because it is an effective method for reducing the effect of variance between participants by permitting an individual to act as their own control.

5.1.3 Procedure

Before beginning the experiment, the participant recorded their basic demographic information and gave written consent. Then, the participant was introduced to one of the three immersive head-mounted displays, and they were instructed on how to wear and interact with the system. During this tutorial, they were also shown how to use a gaze directed interface.

Next, the researcher explained the experimental task to be performed and guided the participant to the experimental setup, which can be seen in Figure 4.3. In this room, stimuli were presented either on the floor, which was covered in blue foam squares, or on a nearby table. A disposable cloth of a similar blue color was draped along the top of the table with white tape wrapped around its edges. The tape was included to improve the table's salience for the two AR systems, which relied on inside-out tracking. Throughout the experiment participants sat in one of two adjacent chairs in the room. Both the approximate locations and chairs used in the experiment are marked in Figure 4.3 for clarity.

Virtual target objects were placed in front and slightly to the left of the participant. To view stimuli that were placed on the floor, a participant sat in the right chair, and to view stimuli that were positioned on the table, a participant sat in the left chair. The left chair was included to view targets atop the table since the table was physically offset towards the left of the room. Before each experimental block, a participant was guided to the appropriate chair for viewing. The researcher also gave participants a short break while the system was calibrated for the next distance condition. Calibration was performed to ensure that target objects appeared at the correct position in space. For the OST AR system, this process involved positioning a HoloLens' world anchor at a predetermined position in the room. For the VR and VST AR systems, calibration entailed placing a Vive Tracking puck.

After calibration was complete, the participant was given their head-mounted display. Upon donning the display, the participant saw a single prompt, which asked if they were ready to begin the next portion of the study. The next block of the experiment began once the user selected the 'ready' button below this prompt. Each participant was asked to respond to experimental trials as quickly and as comfortably as possible. The participants viewed one stimulus at a time. Once they determined their

answer, they clicked once to reveal an input prompt with the 5 potential confidence in ground contact ratings. This prompt appeared immediately in front of the user to prevent strenuous head movement, and the target object was removed from sight. The next trial began once the user selected a value between 1 to 5 on the input prompt. Clicking anywhere else allowed the user to toggle between viewing the current target and its corresponding input prompt. After every 16 trials, the participant was asked if they needed a break via a virtual prompt within the simulation. After 80 trials, the system was recalibrated for the next distance condition and the next experimental block began.

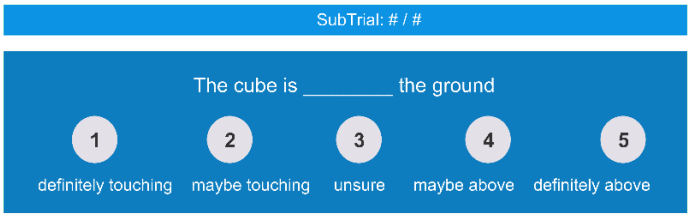

Figure 5.2: Input prompt for measuring one's certainty in estimating ground contact

5.1.4 Results

Our study used ordinal subjective assessments to investigate differences in perceived ground contact across shadow shading conditions. Participants gave ratings of their confidence in perceived surface contact using a 5-point response format like the one used in Madison et al. [295]. Due to the use of an ordinal response format, the resulting data were not normally distributed and nonparametric statistical analyses were employed for correct interpretation. First, we employed Friedman tests to determine if there were differences in confidence ratings between experimental conditions. For each participant we evaluated the average confidence rating for perceived contact across 10 repeated trials. We then used Wilcoxon signed-rank tests in post-hoc analyses with Bonferroni correction to understand specific effects of shadow and distance conditions. Bonferroni correction is recommended for evaluations with multiple comparisons to compensate for an increased chance of Type I error. As a result, for our evaluations of the three shadow methods and the three distance conditions, the significance level was set at $p < 0.0167$.

5.1.5 Shadows improve surface contact

Based on an abundance of prior research in psychology and computer graphics, which demonstrates that shadows provide an effective cue for establishing ground contact [135, 295, 333, 449], we anticipated that the presence of a shadow would result in a significant difference in people's confidence in perceived contact between the no shadow and the other shadow conditions–regardless of display system. To evaluate this hypothesis, we ran a Friedman test on average confidence of contact ratings between the no shadow condition and all other shadow conditions when collapsed. Table 5.1 summarizes the average confidence ratings across participants as well as the results of this analysis.

For the OST AR display, when target objects were placed on a surface, Friedman tests showed a significant difference in people's confidence of contact ratings between the shadow conditions when collapsed together and the no shadow condition ($\chi^2(2) = 4.45, p = 0.035$). However, when objects were placed above the ground, there was no significant difference in confidence ratings between the no

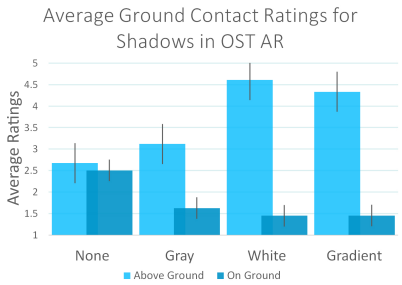

Figure 5.3: Average confidence in ground contact rating with 95% CI of ground contact with shadows in OST AR

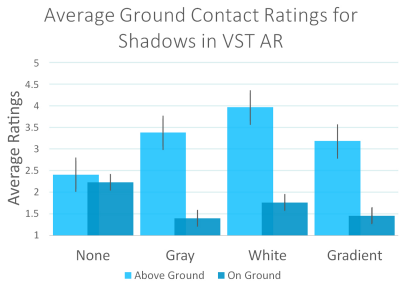

Figure 5.4: Average confidence in ground contact ratings with 95% CI of ground contact with shadows in VST AR

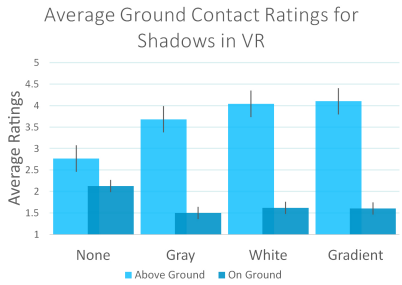

Figure 5.5: Average confidence in ground contact ratings with 95% CI of ground contact with shadows in VR

shadow and collapsed shadow conditions ($\chi^2(2) = 1.60, p = 0.206$). This indicates that people's confidence in perceiving when an object was placed above the ground was higher when a cast shadow was present—but only when the target object was truly placed on the ground.

For both the VST AR and VR display conditions, people were more confident in assessing surface contact in the presence of a shadow, regardless of whether the target object was placed on the ground or above it. For the VST AR device, people's confidence differed between the no shadow and shadow conditions for targets with the same degree of significance for both on and above ground targets ($\chi^2(2) = 8.33, p = 0.004$). There was a difference in people's confidence of contact given the presence of a shadow for the VR display, as well, for both the on ground target objects ($\chi^2(2) = 8.33, p = 0.004$) and the above ground target objects ($\chi^2(2) = 12.00, p = 0.001$).

As expected, we found significant differences in confidence of contact between the no shadow and collapsed shadow conditions for all devices–a result which confirms our hypothesis (**H1**). However, it was curious that confidence ratings for virtual objects in OST AR were significant only when placed on the ground. To better understand this finding, we conducted a post-hoc analysis on the average confidence ratings for above ground objects for each shadow condition in the OST AR device. Wilcoxon signed rank tests indicated that people's ratings for the no shadow condition were significantly different from the white ($Z = -3.1, p = 0.002$) and gradient ($Z = -2.8, p = 0.005$) shadow conditions, but not the gray shadow condition. These findings were reasonable, given that dark color values are known to appear more transparent in additive light displays like those used in OST AR. If a shadow is too transparent for a viewer to discern when an object is placed above the ground, then it may adversely affect the viewer's ability to determine surface contact such that their confidence ratings more closely resemble the ratings expressed in the no shadow condition.

5.1.6 Realistic shadows are not necessary to improve surface contact

Next, we evaluated differences in confidence of contact for on and above ground targets for the different shadow shading methods via Friedman tests for each device. In the OST AR device, Friedman tests across the three rendered shadow conditions revealed no significant differences in people's confidence ratings between the methods when an object was placed on the ground. However, the same analysis did find a difference when objects were placed above the ground ($\chi^2(2) = 18.500, p < 0.001$). Post-hoc tests revealed that confidence of contact ratings for all three shadow conditions were significantly dif-

Table 5.1: Results of Friedman analysis–Friedman's Q–between the no shadow and shadow conditions for each device

	On Ground				Above Ground			
	Avg Rating				Avg Rating			
	No Shadow	Shadow	$\chi^2(2)$	Sig	No Shadow	Shadow	$\chi^2(2)$	Sig
OST	2.5	1.6	4.455	$= 0.035^*$	2.7	3.1	1.600	$= 0.206$
VST	2.2	1.4	8.333	$= 0.004^*$	2.4	3.4	8.333	$= 0.004^*$
VR	2.1	1.5	8.333	$= 0.004^*$	2.8	3.7	12.000	$= 0.001^*$

ferent from each other. Namely, ratings for the white shadow ratings differed from the gray shadow ($Z = -3.1, p = 0.002$) and the gradient shadow ($Z = -2.8, p = 0.005$); the ratings for the gradient shadow differed from the gray shadow ($Z = -2.9, p = 0.004$). For each shadow condition the average confidence of contact ratings were: 3.1, 3.8 and 4.6 for the gray, gradient, and white shadows, respectively. Figure 5.3 further illustrates the differences in these ratings, where the white shadow shading condition is given the highest confidence rating and the gray shadow shading condition is given the lowest confidence rating among the shaded shadow conditions. For OST AR, brighter–and therefore more salient–shadows appeared to greatly influence people's confidence of surface contact when objects were placed above the ground.

For the VST AR device, we found a significant difference in confidence of contact ratings between shadow conditions for both on ground ($\chi^2(2) = 7.787, p = 0.020$) and above ground ($\chi^2(2) = 12.667, p = 0.002$) target objects. The average ratings for the no shadow conditions and the three, shaded shadow conditions are visualized in Figure 5.4. Post-hoc analyses revealed that when target objects were on the ground, the white shadow was significantly different from the gray shadow condition ($Z = -2.432, p = 0.015$). Furthermore, for above ground objects, the white shadow was significantly different than both the gray ($Z = -2.472, p = 0.013$) and the gradient shadow condition ($Z = -2.590, p = 0.010$). Interestingly, for both on and above ground targets, the white shading method for shadows generally resulted in higher average confidence of contact ratings for target objects when compared to the other shading methods (See Table 5.2).

In VR there was no significant difference in confidence of contact between the three shadow shading methods (See Figure5.5 and Table 5.2). Although confidence was unaffected by shadow shading method in VR, confidence ratings in AR displays proved quite sensitive, which confirmed our second hypothesis (**H2**) that shadow shading method would influence people's confidence in surface contact perception in augmented reality displays. However, we were unable to confirm (**H3**), which predicted that people would make more confident surface contact ratings given realistic cast shadows over non-photorealistic shadows in AR. This outcome becomes particularly curious when we examine the non-photorealistic (the white shadow) shading condition's performance across the OST AR and VST AR devices. In the OST AR display, people's confidence was significantly higher when target objects were placed above the ground. Their confidence matched the ground truth of the target's position in space. However, in VST AR, people's confidence ratings were higher when presented with the white shadow, regardless of whether the object was placed on the ground or above it, which may be an undesirable outcome for establishing ground contact.

Table 5.2: Results of Friedman analysis between the shadow conditions (dark gray, white, and gradient) for each device

	On Ground					Above Ground				
	Avg Rating					Avg Rating				
	Gray	White	Grad	$\chi^2(2)$	Sig	Gray	White	Grad	$\chi^2(2)$	Sig
OST	1.6	1.4	1.5	1.317	$= 0.518$	3.1	4.6	3.8	18.500	$< 0.001^*$
VST	1.4	1.8	1.5	7.787	$= 0.020^*$	3.4	4.0	3.4	12.667	$= 0.002^*$
VR	1.5	1.6	1.6	1.227	$= 0.541$	3.7	4.0	3.9	4.667	$= 0.097$

5.1.7 The distance to a target affects people's ability to discern surface contact

We displayed targets at multiple positions within personal and actions space to better understand how our results were influenced by viewing conditions. We believed this was especially important for our first investigation since depth cues can vary in effectiveness across distances [96]. We first conducted Friedman tests across the 1m table, 1m floor, and 3m floor conditions to determine if there were differences in confidence of surface contact ratings across the viewing distances. Table 5.3 shows the average ratings and significance values across conditions for this analysis. We then ran post-hoc tests to compare confidence ratings between specific distance conditions.

For the OST AR display, we found no significant difference in confidence of contact ratings across viewing distance conditions when the object was in contact with the ground. However, there was a significant effect of distance when objects were placed above the ground ($\chi^2(2) = 16.468, p < 0.001$). Post-hoc tests showed that people's confidence in contact ratings for the 3m floor distance were significantly higher than the ratings for the 1m floor ($Z = -2.8, p = 0.006$) and 1m table conditions ($Z = -3.1, p = 0.002$).

We also found no difference in confidence of contact ratings across viewing distances when objects were placed on the ground for VST AR. However, there was a significant difference between viewing distance conditions for above ground objects ($\chi^2(2) = 12.667, p = 0.002$). In a similar pattern to we found in the OST AR condition, in VST AR we found that people were more confident that target objects were placed above the ground for the 3m floor condition than either the 1m floor condition ($Z = -2.7, p = 0.008$) or the 1m table condition ($Z = -3.1, p = 0.002$).

Similarly, in the VR display there was no effect of viewing distance for the on ground objects; however, there was a significant difference between viewing distance conditions when objects were placed above the ground ($\chi^2(2) = 18.766, p < 0.001$). In VR, people's confidence of surface contact was significantly higher for the 3m floor and 1m floor conditions when compared to the 1m table condition, with the same degree of significance for both comparisons ($Z = -3.1, p = 0.002$). The average confidence ratings for each viewing distance condition are displayed in Table 5.3. The average ratings for above ground objects were higher in the 1m floor and 3m floor conditions with scores of 4.2 and 4.1, respectively. In contrast, the average rating for the 1m table distance condition was only 2.5.

For all devices, we found significant differences in confidence of surface contact as a function of viewing distance for target objects positioned above the ground, but not for those positioned on the ground. In addition, confidence ratings for above ground objects were significantly higher for the 3m floor condition than either one or both of the 1m floor and 1m table conditions for all devices. A pos-

sible explanation for this finding is that the 3m floor condition permitted the viewer to see underneath the target object, thereby making it easier for the viewer to determine when objects were positioned above the ground since their view of the cast shadow underneath the object was more clear. However, people's ratings for the 3m floor and 1m floor conditions did not significantly differ in the VR condition, so this finding may be unique to AR devices.

Table 5.3: Results of Friedman analysis between the distance conditions (1m on table, 1m on floor, 3m on floor) for each device

	On Ground					**Above Ground**				
	Avg Rating					Avg Rating				
	1m Tab	1m Fl	3m Fl	$\chi^2(2)$	Sig	1m Tab	1m Fl	3m Fl	$\chi^2(2)$	Sig
OST	1.5	1.4	1.7	3.268	$= 0.195$	3.4	3.7	4.5	16.468	$< 0.001^*$
VST	1.4	1.8	1.4	2.227	$= 0.328$	3.0	3.3	4.4	12.667	$= 0.002^*$
VR	1.7	1.7	1.7	0.304	$= 0.859$	2.5	4.2	4.1	18.766	$< 0.001^*$

5.1.8 Discussion

Overall, the results of our comparisons between the no shadow condition and all other shadow conditions when collapsed re-affirm that shadows provide a powerful cue for ground contact (**H1**). However, our investigation cautions against the use of dark color values for shadows in OST AR devices as they may be less effective than other, more visible, techniques for establishing ground contact. Although we did not find a significant effect of shadow condition for the OST AR display when objects were placed above the ground, post-hoc analyses revealed that people's confidence in the gray shadow condition was similar to the no shadow condition. People's low confidence when presented with the gray shadow in OST AR likely influenced the outcome of our Friedman test between the no shadow and shadow conditions.

A priori, we also anticipated that shadow shading techniques would vary in effectiveness for both AR devices (**H2**). We confirmed this hypothesis, but we were surprised to find that the photometrically incorrect shading method–the white shadow–generally resulted in higher confidence of surface contact when objects were placed above the ground for both AR devices. This outcome was particularly unexpected since we had predicted that more perceptually valid shadow methods, like the gray and gradient shadows, would be more beneficial for establishing ground contact. As a result, we could not confirm our third hypothesis (**H3**). This counterintuitive result encouraged further evaluation of the effect of high contrast, and therefore more visible, object and shadow shading conditions in our second study.

It should be noted that we saw different effects of shading conditions between the OST AR and VST AR devices. This can be attributed to differences in the displays as both AR devices combine real and virtual images in very different ways. Whereas people's confidence of surface contact in the OST AR device proved highly sensitive to all shading methods when objects were placed above the ground, people's confidence of contact in the VST AR condition was sensitive to a single shading condition–the white shadow–regardless of whether an object was placed on or above the ground. In addition, for the OST AR device we found that the white shadow resulted in the highest confidence and the gray

shadow resulted in the lowest confidence. The gradient shadow's confidence ratings fell between the two. One interpretation of these results is that more salient shading techniques are more effective for determining ground contact in OST AR displays. For example, the white shadow may have been more prominent against the darkly textured target object used in our first evaluation.

Our first study on ground contact perception suggested that non-photorealistic shadow shading techniques in AR may be beneficial for attaching virtual objects to the surfaces beneath them. This finding was surprising given that researchers previously theorized that consistent depth cue information between real and virtual objects is necessary to improve depth perception in augmented reality [1, 107, 245]. These findings motivated a second investigation in which we hypothesized that the color contrast between the target object and its shadow was a contributing factor to our initial results.

5.2 Study 2 - Color contrast between object and shadow influences surface contact

The results of our first experiment indicated that shadow shading plays an important role in determining ground contact in augmented reality. Interestingly, we found that people performed better with white shadows—or non-photorealistic shadows—than with the other shading methods when discerning ground contact for both OST AR and VST AR devices, especially when objects were placed above the ground. Given these results, we suspected that the high contrast of the white shadow's color value against the darkly textured target object used in our study allowed participants to more confidently assess ground contact. To better understand how object shading may have influenced the findings of Experiment 1, we evaluated different object and shadow shading conditions in Experiment 2. We also used the same testing environment with a medium blue background so that we could draw comparisons between the two studies.

Therefore, for Experiment 2, we designed our study to evaluate the effect of color contrast between a target object and its shadow. We employed a 2 x 2 design in which the shading methods used for both the target object and its shadow were manipulated to use either light or dark color values. The 2 x 2 design resulted in four unique shading conditions that allowed us to parse out what aspects of perceived surface contact were affected by shadow shading method alone versus the contrast of an object and its shadow shading method. For evaluation we elected to use a psychophysical approach. Psychophysics is a class of psychological methods that quantitatively measures perceptual responses to changes in physical stimuli [118]. This change in protocol allowed for greater sensitivity in measuring perception of ground contact and it enabled us to cast our evaluation as a within-subjects evaluation across devices. The same three, immersive head-mounted displays from Experiment 1 were used to test the perception of shadows in this experiment.

Based on the results of our first experiment, we developed three hypotheses. First, we anticipated that high color contrast between objects and their shadows would improve the likelihood of correct assessment of ground contact in augmented reality displays (**H1**). We also predicted that white shadow shading methods would improve participants' ability to discern surface contact in augmented reality (**H2**). However, given the results of our first study and our prior discussion, we did not anticipate any significant effects of shadow shading method on task performance for the virtual reality head-mounted display (**H3**).

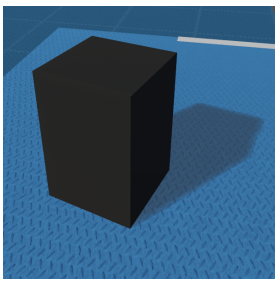

Figure 5.6: Dark object with dark shadow condition (DODS)

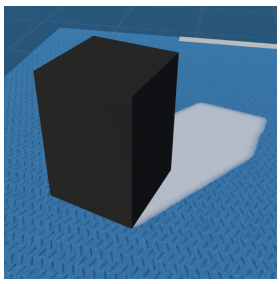

Figure 5.7: Dark object with light shadow (DOLS)

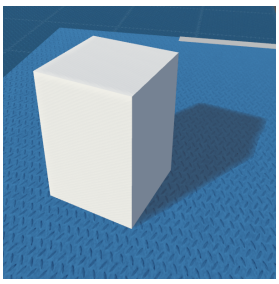

Figure 5.8: Light object with dark shadow (LODS)

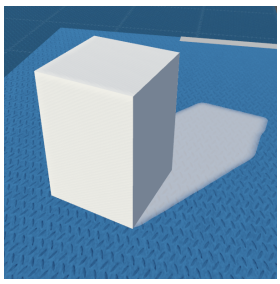

Figure 5.9: Light object with light shadow (LOLS)

5.2.1 Participants

Six individuals in total (3M, 3F) aged 20–45 from Vanderbilt University volunteered to participate in the second experiment, which used a psychophysical paradigm. These methods rely on a small number of participants to make a large number of simple, behavioral responses that reveal underlying perceptual processes. Psychophysical paradigms have proven highly replicable since they employ judgments or adjustments with low individual variance [112, 435]. Although this family of paradigms typically rely on a smaller number of participants, six participants were required to counterbalance the presentation order of the three immersive displays. All participants had normal or corrected to normal vision. Our experimental methods were approved by the local institutional review board, and written consent was obtained from all subjects prior to participation.

5.2.2 Design

For Experiment 2, we evaluated the relationship between object and shadow shading across the same OST AR, VST AR, and VR devices that were used in Experiment 1. However, in this experiment we employed a psychophysical approach to evaluate how light and dark color values affect the relationship between a target object and its shadow when determining ground contact. By using this approach we were able to restructure our experiment as a within-subjects evaluation for further experimental control, and we were able to efficiently evaluate ground contact perception for target objects at multiple heights for a more sensitive measure of performance. Specifically, we used a two-alternative forced forced choice (2AFC) design with method of constant stimuli—a classic method [154]. We used a within-subjects 2 x 2 factorial design in which two levels of shading for a target object and its shadow were presented.

The shading levels contained light and dark color values such that two high contrast and two low contrast conditions were created. In this context, we refer to contrast as a difference in color—rather than luminance—since we are unable to directly compare luminance values using traditional methods over the three unique display types. Grayscale color values were used to inform shaders for both the target object and its shadow. The low contrast conditions were: [light object x light shadow] and [dark object x dark shadow]. The high contrast conditions were: [light object x dark shadow] and [dark object x light shadow]. High contrast conditions had a difference of 200 RGB color values and low contrast conditions had a difference of 30 RGB color values. Specifically, we use grayscale RGB color values of 250 and 20 to inform the target object shader and we use grayscale RGB color values of 220

and 50 for the shadow shader in this experiment. The same shadow shaders that were developed for Experiment 1 were also used for Experiment 2 (See Subsection 4.2.2). This design resulted in 4 unique combinations of experimental stimuli.

For each stimulus pair in our temporal 2AFC protocol, the participant was asked "Which stimulus is closer to the ground?" Each stimulus was presented for 600 msec. In between each stimulus pair, there was an interval of 800 msec in which a random pattern was presented to avoid visual aftereffects. Participants responded using the left mouse button to indicate the first stimulus and the right mouse button to indicate the second stimulus. One stimulus was presented on a surface and the other was presented at some height above the surface. In this second experiment, all objects were positioned 1m away from the viewer on the same table that was used in Experiment 1.

Objects were presented at one of six heights between 0 and 3mm at regular intervals of 0.03 degree changes in viewing angle (See Section 4.2.3). This change in visual angle corresponded to approximately a 0.06mm change in height per step. Each height comparison was presented 20 times each, except for when both the first and second object heights were both presented on the ground with 0 vertical displacement. In this case, the height conditions were only presented 10 times each.

The experiment was blocked by device and counterbalanced across subjects. Each device block consisted of 440 trials, resulting in each participant completing 1320 trials throughout the experiment. The experiment took approximately two hours to complete, with participants taking around 30 minutes to complete the experimental task in each device. Within each experimental block, stimulus pairs were presented pseudo-randomly such that there were no repetitions of any unique stimulus combination before all other unique stimuli were presented once. The experiment was self-paced. Both the user's response and their response time were recorded for each trial. The next trial began 1000 ms after the participant responded to the previous trial–unless the experiment was paused. In total, across all subjects we collected 7,920 datapoints.

5.2.3 Procedure

This experiment was conducted after the Covid-19 outbreak. Therefore, special considerations (e.g., social distancing) had to be implemented to protect both the experimenter and the participants. First, the researcher explained the experimental protocol to the participant and gave them an informed consent form. Throughout the experiment, participants were informed that they could take a break at any time, and the programs used to run the study asked participants after every 44 trials if the subject required a break from the task.

Participants were asked to calibrate the equipment themselves–instead of the experimenter–to avoid the spread of germs through shared head-mounted displays. Otherwise, the calibration phase for the current experiment was the same as in Experiment 1. After completing the trials in one head-mounted display, the participant filled out a short post survey before continuing onto the next head-mounted display used for the experiment and thus the next block of trials. At the end of the experiment, participants filled out a short final survey.

5.2.4 Results

We analyzed our data using a binomial mixed model to understand the influence of shading condition on participants' judgments. Mixed models are a form of generalized regression that is appropriate for non-normally distributed outcomes and repeated-measures designs. The four shading conditions were

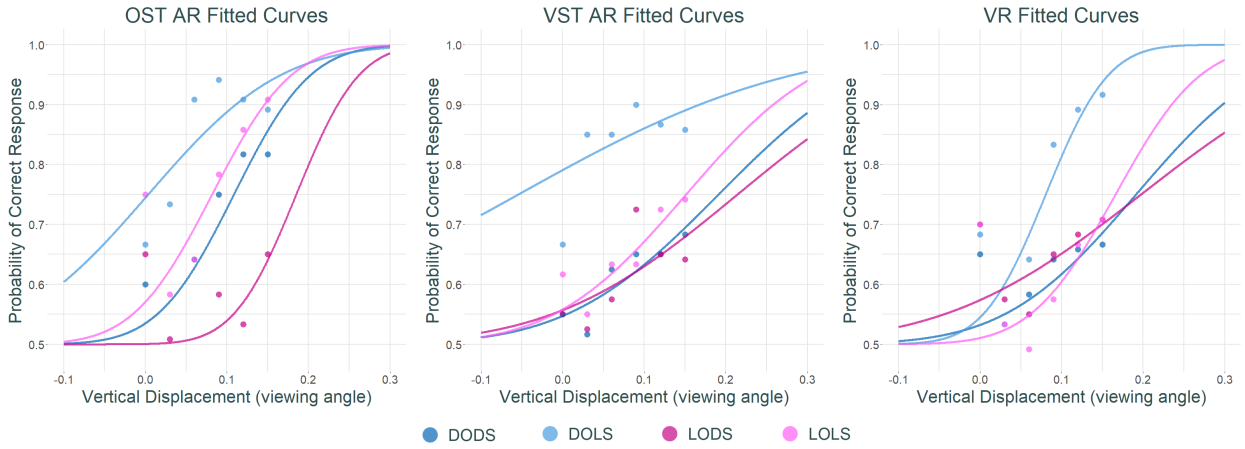

Figure 5.10: Response data from the 2AFC task in Experiment 2 has been fitted with a cumulative normal curve using a generalized linear model. The figure shows the psychometric curves for response data for the OST AR display (left), the VST AR display (middle), and the VR display (right) conditions.

included as within-subject predictors in our model: 1) dark object with dark shadow (DODS), 2) dark object with light shadow (DOLS), 3) light object with dark shadow (LODS), and 4) light object with light shadow (LOLS).

In this analysis, positive effects indicate an increased likelihood of correct ground contact judgement in our 2AFC task. Across all three devices, we found a main effect of shading condition such that participants viewing the dark object with light shadow (DOLS) condition had an increased likelihood of correctly judging a target's ground contact in comparison to the other shading methods. In Figure 5.10 participants' forced choice response data has been plotted with a psychometric function for each display condition for clarity.

In the video see-through augmented reality device, the DOLS condition resulted in an increased likelihood of correct response in comparison against the DODS ($\beta = 1.397, SE = 0.149, p < 0.0001$), the LODS ($\beta = 1.41, SE = 0.149, p < 0.0001$), and the LOLS ($\beta = 1.25, SE = 0.150, p < 0.0001$) shading conditions. For the virtual reality device, the DOLS condition performed similarly against the other conditions with an increased likelihood of correct response in comparison to the DODS ($\beta = 0.807, SE = 0.13, p < 0.0001$), the LODS ($\beta = 0.658, SE = 0.132, p < 0.0001$), and the LOLS ($\beta = 0.830, SE = 0.131, p < 0.0001$) conditions.

From the results of our first experiment, we expected more nuanced relationships between object and shadow shading methods to be revealed for the OST AR device, and this expectation was met. Each shading condition performed significantly different from each other such that the order of highest to lowest likelihood of correct response was: 1) dark object with light shadow (DOLS), 2) light object with light shadow (LOLS), 3) dark object with dark shadow (DODS), and 4) light object with dark shadow (LODS).

In a similar fashion to the results found in the other displays, the highest likelihood of correct response was found in the DOLS shading condition. Statistically speaking, people's responses given the dark object with light shadow condition (DOLS) were more accurate than when they were presented with any of the other shading conditions: LOLS ($\beta = 0.894, SE = 0.1613, p < 0.0001$), DODS ($\beta = 1.215, SE = 0.158, p < 0.0001$), LODS ($\beta = 1.919, SE = 0.155, p < 0.0001$). People also performed better in the light object with light shadow condition (LOLS) than both of the dark shadow

conditions: DODS ($\beta = 0.321, SE = 0.136, p = 0.0181$) and LODS ($\beta = 1.0253, SE = 0.132, p < 0.0001$). And they performed better in the dark object with dark shadow condition than the LODS ($\beta = 0.704, SE = 0.127, p < 0.0001$) condition.

5.2.5 Discussion

Across all devices, the dark object with light shadow (DOLS) condition resulted in an increased likelihood of correct judgement of distance to the ground. This result echoes the findings of our first experiment in which the white shadow condition resulted in significantly different responses for both augmented reality devices. Prior research in computer graphics has shown us that light, photometrically incorrect cast shadows may be as effective as dark, perceptually correct cast shadows in spatial location tasks [228]. Our current research finds a similar trend in that people often perform better with the white shadow condition than with other perceptually valid shading approaches—especially in our employed additive light display.

However, it is interesting to note that while the DOLS condition resulted in a significant increase in performance in the VR condition for Experiment 2, we did not see this pattern of results in the first experiment. We believe this is due in part to the difference in paradigm. Whereas our first evaluation was useful for uncovering multiple cues that may be interacting with a viewer's ability to perceive ground contact, in general, our second evaluation used a more sensitive paradigm to allow us to isolate to what degree cast shadow shading cues were affecting a viewer's sense of ground contact. Thus, we may have been better able to isolate subtle differences in performance in Experiment 2. Across all devices, participants generally expressed that the dark object with light shadow (DOLS) shading condition was the easiest condition to see the shadow, which explains why it performed pointedly well.

The DOLS method was the only one to result in a significant difference in performance from the other conditions in both the video see-through and virtual reality device (DOLS > LOLS , DODS , LODS) . However, the optical see-through device proved highly sensitive to all shading conditions. In order of highest likelihood of correct response to lowest likelihood, the shading conditions performed as follows: DOLS > LOLS > DODS > LODS. It is important to note that the two high contrast conditions (dark object with light shadow and light object with dark shadow) resulted in highly polarizing performance. This is likely due to the HoloLens' reliance on additive light to render objects. In addition, both light shadow conditions resulted in more accurate ground contract perception over the dark shadow shading methods. While participants expressed that the DOLS condition was easy to interpret since the shadow was very visible, participants complained that the opposite was true for the LODS condition. Participants reported that the bright appearance of the object made it difficult to see the dark shadow underneath it.

In summary, we were unable to confirm our first hypothesis, in which we predicted that people would perform better in the high contrast color conditions than the low contrast conditions when determining ground contact in augmented reality (**H1**). However, our results did indicate that a viewer's perception of ground contact is highly sensitive to high color contrast between objects in their shadows in optical see-through AR. We were also only able to partially confirm our second hypothesis in which we predicted that the white shadow condition would improve surface contact perception in augmented reality displays (**H2**). We found this relationship to be true in the OST AR device; however, in VST AR the white shadow only resulted in an increased likelihood of correct response when the target object was dark (DOLS). Finally, unlike in our first study, we found a significant effect of shading condition in the virtual reality condition where the DOLS shading condition performed significantly better

than all other conditions, which means that were were unable to confirm our third hypothesis (**H3**).

5.3 Study 3 - Surface contact judgements are more accurate for a rectilinear object than for a sphere

Preliminary evidence from Study 1 (Section 5.1) and Study 2 (Section 5.2) suggested that non-photorealistic shadow shading techniques in AR may be beneficial for attaching objects to the surfaces beneath. This finding is surprising given what we know about depth perception. However, in our initial evaluations we only assessed surface contact judgements for simple, rectilinear shapes. It is possible that people's judgements may have been influenced by the straight lines of our target object. Therefore, we decided to conduct additional research to investigate the influence of object shape and orientation on people's ability to use photorealistic (dark) and non-photorealistic (light) shadows for surface contact judgments. We conducted two psychophysical studies.

In both studies, shadows were manipulated to have either dark or light color values. In our third study, which is discussed in Section 5.3, we manipulated both shadow shading technique and object geometry to understand how the interplay between these two factors affected perceived surface contact in MR. In our fourth study (Section 5.4), we compared a simple geometric shape to a complex one, and we displayed them at different orientations. From the findings of our two studies, we were able to extrapolate MR developer and design guidelines, which are discussed in Section 5.5.

For our followup work, we used the same three, immersive head-mounted displays (HMDs): a VR display, a VST AR display, and an OST AR display. Our hardware setup (Subsection 4.2.1), the software environment used for rendering to each display as well as the shaders used for rendering shadows (Subsection 4.2.2), the equation used to displace test stimuli vertically based on visual angle (Subsection 4.2.3) were all the same as those used in the prior experiments.

We evaluated how realistic and non-photorealistic shading techniques for cast shadows affect one's ability to perceive surface contact for a cube, an icosahedron, or a sphere (Figure 5.11). These three objects were selected for their distinctive geometric properties. In particular, the shape, shading, and shadows of the cube and sphere are recognizable and different. The icosahedron represents an in-between case, with its indistinct shape, but shading closer to that of the sphere and shadow closer to that of the cube. Our hypotheses for this experiment are as follows:

H1: We predicted that people's ability to correctly perceive ground contact would be affected by an object's geometry, especially along the bottom edge that would be in contact with a surface. Specifically, we anticipated that people's likelihood of correct response would be lower for the sphere across all devices. A perfect sphere has only a single point of contact with a ground surface and that singular point is occluded when viewed from above. In contrast, both the icosahedron and the cube were flat along the edge and benefited from many points of contact with the surface beneath them, although the surface area of the bottom of the icosahedron was smaller than that of the cube.

H2: We anticipated that people would be more likely to correctly perceive surface contact when presented with the light shadow shading method, in comparison to the traditional dark shading method in both AR devices. Although, our prediction may seem counterintuitive, some prior research has shown that non-photorealistic shadows may function as well as photorealistic shadows as a depth cue for spatial perception [5, 228].

H3: We also anticipated that there would be interactions between shape and shadow shading methods. However, *a priori* we were uncertain how these interactions would be expressed.

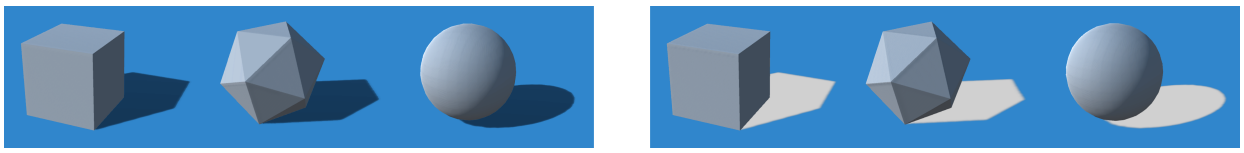

Figure 5.11: Three geometric shapes were evaluated in Experiment 1. Each object was rendered with either dark (realistic) or light (non-photorealistic) shadow shading methods and was displaced vertically based on the viewer's visual angle to the target. Changes in vertical displacement were subtle, since a two-alternative forced choice (2AFC) psychophysical approach was used to evaluate surface contact perception.

5.3.1 Participants

Six individuals in total (3M, 3F) aged 23–30 volunteered to participate in our experiment, which used a well-founded psychophysical paradigm. Psychophysics is a class of psychological methods that quantitatively measures perceptual responses to changes in physical stimuli [118]. These methods can use a small number of participants to make a large number of simple, behavioral responses that reveal underlying perceptual processes, gaining experimental power from a large number of observations. Psychophysical paradigms have proven highly replicable and robust since they employ judgments or adjustments with low individual variance [112, 435]. Psychophysics methods have previously been employed successfully by other MR research groups [65, 129, 414].

All participants had normal or corrected to normal vision. Our experimental methods were approved by the local institutional review board, written consent was obtained from all subjects prior to participation, and each participant was paid 20 USD for 2-3 hours of their time.

5.3.2 Design

To address our hypotheses, we utilized a 3 (MR display) $\times$ 3 (target shape) $\times$ 2 (shadow shading) $\times$ 6 (target height) within-subjects design to evaluate the effects of object shape and shadow shading on surface contact judgments for objects at different vertical displacements. Although we conducted our experiment with three different MR displays, we did not make any direct statistical comparisons across displays due to the high variance in optical and graphical properties across devices. However, we evaluated the same experimental conditions for each MR display. As such, this section describes the experimental design that was used for each MR device.

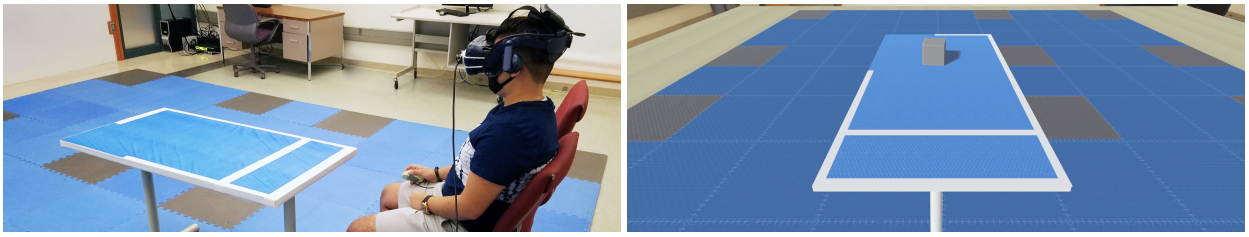

Figure 5.12: A participant views the experiment in the VST AR condition (left). An image of the virtual environment used for the VR condition (right).

Each participant evaluated 3 object shapes with 2 different shadow shading conditions,which resulted in 6 unique combinations of experimental stimuli. The target shapes evaluated were a cube, an

icosahedron, and a sphere. The shadows were shaded with either dark or light color values. Figure 5.11 displays all of the target shape and shadow shading conditions. All targets were rendered on a table 1 meter in front of the viewer (Figure 5.12). The size of the targets was adjusted to be 11.2 cm wide. Accordingly, the cube had a length, width, and height of 11.2 cm, the sphere's diameter was 11.2 cm, and he icosahedron was scaled so that its width was approximately 11.2 cm.

We employed a temporal two-alternative forced choice paradigm with method of constant stimuli—a classic psychophysical method [55, 118]—to evaluate people's perception of surface contact. Two-alternative forced choice, or 2AFC, is a method for assessing someone's sensitivity to a change in stimuli. It involves a *forced choice*, because participants must chose between one of two options. A temporal 2AFC paradigm may also be referred to as a two-interval forced choice (2IFC) paradigm. After viewing both stimuli in a sequence, participants in our experiment are then asked "Which object is closer to the ground?". Compared targets were always of the same shape and shadow condition for each trial.

How the six levels of vertical displacement were presented was dictated by *method of constant stimuli*. As such, at least one of the target stimuli was always placed in contact with the surface beneath it. The other object was then displaced vertically by one of six heights between 0 and 6mm at regular intervals of .06° changes in viewing angle. When displacement was 0, both the first and second targets were positioned in contact with the surface.

Changes in vertical displacement were subtle. All six height displacements are displayed in Figure 5.11, and the calculations for vertical displacement are described in detail in Section 4.2.3. Each height comparison was presented 20 times each, except for when both the first and second object heights were equal (at 0). In this case, the height conditions were only presented 10 times each. Stimulus pairs were presented pseudo-randomly such that there were no repetitions of shape $\times$ shadow $\times$ height combinations before all combinations were presented once. In addition, the order of presented stimuli was balanced so that the number of trials in which the target was presented in contact with the ground first and the number of times it was presented second were equivalent for each condition.

Trials for the task were blocked by device and order of device was counterbalanced across subjects. For each MR device, a participant completed 660 2AFC trials. Each participant thus completed 1,980 trials total, and 11,880 data points were collected across all subjects. We collected data over a large number of trials, which is common practice in psychophysical paradigms, to increase accuracy and ensure low variance in our study [377]. By evaluating a large number of simple behavioral responses on a small number of participants, this family of paradigms is better able to evaluate perceptual behaviors that have little variance between individuals [112, 435].

We further confirmed that variance in our collected data was not due to between participant variance by calculating the intraclass correlation coefficient (ICC). The ICC measures the amount of variance accounted for by a grouping variable. For our analysis, individual participants was selected as the grouping variable, where an ICC value of 1 indicates that any variance in the data is between participants while a value of 0 indicates that no variance in the data is due to between participant factors. We found that $\tau_{ost} = .009$ for data collected with the OST AR display, which indicated negligible variance was caused by between subjects factors. Similarly low ICC values were found for the VST AR and VR display data with ICCs of $\tau_{vst} = .012$ and $\tau_{vr} = .042$, respectively.

5.3.3 Procedure

Before the experiment began, each individual was informed about the study and filled out a consent form. They were told that they may stop the experiment at any time and that they were allowed to

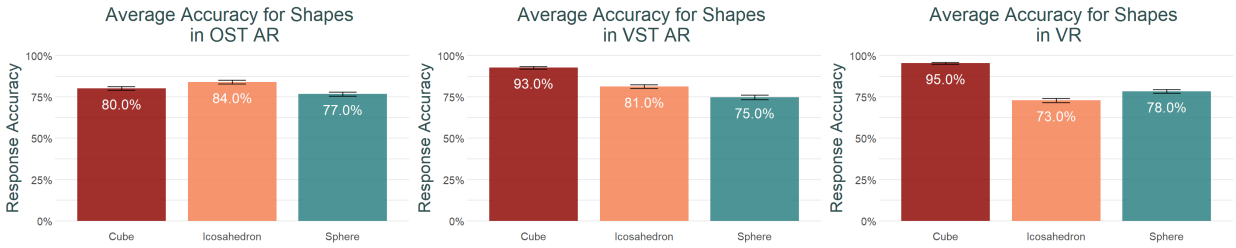

Figure 5.13: Experiment 3 — Average percentage of correct responses for each target object shape by display condition: optical see-through AR (left), video see-through AR (center), and virtual reality (right). The effects of shape on surface contact judgement were complex, with significant but different effects of target shape for each display. However, people's judgements with the cube were significantly better than judgements with the sphere in all three devices.

take breaks during the experiment if needed. Both the experimenter and the volunteer wore face masks while maintaining 2 m of separation between each other during the experiment.

Because the experiment was blocked by device, participants picked up one of the three designated MR HMDs and started up the system. They were then verbally instructed on how to interact with the system and they performed 10-20 practice trials that were randomly selected trials from the experiment. After the volunteer expressed that they were comfortable with the system and experimental paradigm, the experiment began.

For the temporal 2AFC paradigm with method of constant stimuli, each stimulus was presented for 600 ms. In between each stimulus pair, there was an interval of 800 ms in which a random pattern was presented at the position of the object to avoid visual aftereffects. After the presentation of each stimulus pair, the participant was asked "Which stimulus is closer to the ground?". They were told that at least one of the objects was positioned on the ground. They then responded using the left mouse button to indicate the first stimulus and the right button for the second stimulus. The experiment was self-paced, and both the user's response as well as their response time were recorded for each trial. The next trial began 1000 ms after the participant responded to the previous trial—unless the experiment was paused. After every 66 trials, participants were presented with a visual prompt that asked if they needed to take a break.

Between devices, participants were also required to take a break to prevent fatigue. During the break, they filled out a brief survey that asked them to describe any strategies that they used to determine ground contact. After finishing the experiment in its entirety, they were asked to fill out a final survey. Although we did not directly compare peoples' performance across MR devices in this study, due to the high variance in display properties, in the final survey we asked participants about their experiences in each device. We believed this information would be informative for interpreting our results. Therefore, in the post experiment survey, they were asked to rank the difficulty of the experimental task for each device and to rank the displays' quality of graphics. They were also asked if their strategy for determining ground contact differed across devices.

5.3.4 Results

Participants were asked to make binary decisions about which object was positioned closer to the ground (first or second object) in a 2AFC task. Therefore, for our statistical analyses we used binary logistic re-

gression models, which are appropriate for dichotomous outcome variables, to analyze participants' judgments. We used the glm function from the stats package in R [379] to conduct logistic regressions by specifying binomial errors and a logit link function. Because we wanted to analyze people's perception in each device, we ran separate models for each of the MR displays. For each display, we modeled binary outcomes (accuracy: correct (1) or incorrect (0)) for our predictors: object shape, shadow shading, and height. Height was recorded in millimeters then centered at zero and treated as continuous. Object shape (3 levels: cube, icosahedron, sphere) and shadow shading (2 levels: dark, light) were treated as categorical factors. Interactions between shape and shadow were included in the models.

We used these models to test three planned comparisons for each device to understand how object shape and shadow influenced people's surface contact judgments. These comparisons were: 1) whether surface contact judgments differed for different object geometries (i.e., **H1** the main effect of object shape); 2) the difference in surface contact judgments between dark and light shadow shading across all other conditions (i.e., **H2** the main effect of shadow shading); and 3) interactions between object shape and shading (i.e., **H3**).

In order to examine whether there were main effects, we coded shadow and shape factors using deviation coding (also known as effect or sum coding). For the shape factor, the sphere was set as the reference group (i.e., coded as -1). For the shadow factor, the light shadow condition was coded as .5 and dark shadow was coded as -.5. Using this deviation coding also allowed us to observe whether there was a main effect of height. The general logistic regression equation is depicted in Equation 5.1 below.

$$\begin{aligned}\log\left(\frac{p}{1-p}\right) = {} & B_0 + B_1(\text{shadow}) + B_2(\text{cube}) \\ & + B_3(\text{icosahedron}) + B_4(\text{height}) \\ & + B_5(\text{shadow} \times \text{cube}) \\ & + B_6(\text{shadow} \times \text{icosahedron})\end{aligned} \tag{5.1}$$

For the shape variable (**H1**), we were interested in whether the sphere lead to lower accuracy than the other two shapes, collapsed across shadow and height. In order to answer this question, we conducted planned contrasts on the aforementioned model comparing each shape to one another, using the Bonferroni correction to account for multiple comparisons.

In order to examine whether there was a main effect of shadow shading (**H2**), the shadow regression coefficient represents the difference between dark and light shadows averaging over shape and height. We were also interested in a potential interaction between shape and shadow (**H3**). Specifically, was each shape affected by the shadow shading? In order to answer this question, we calculated the shadow simple slopes for each shape.

For the sake of simplicity, we discuss the implications of our findings as well as which factors are significant in text, and we report the full details of the effects of factors in our analyses in Table 5.5

Optical See-through AR

We show the average percent of correct response for each of the evaluated main effects in Table 5.4. The results of our statistical analyses are reported in Table 5.5. We expected participants to be more likely to correctly indicate which object was on the ground as the height of target objects increased. The improvement of participants' performance as height increased is demonstrated statistically by a main

Table 5.4: Experiment 3 — The accuracy for each shape and shadow condition tested reported with the standard errors.

Accuracy for Experiment 3			
Condition	OST AR	VST AR	VR
Cube	80% [1.2%]	93% [.7%]	95% [.6%]
Icosahedron	84% [1.1%]	81% [1.1%]	73% [1.3%]
Sphere	77% [1.2%]	75% [1.3%]	78% [1.0%]
Dark	77% [.1%]	78% [1.0%]	81% [.9%]
Light	84% [.1%]	88% [.8%]	83% [.9%]

effect of height in our logistic regression model ($OR = 1.09, p < .001$). Collapsed across all shape and shadow conditions, the odds ratio (OR) of 1.09 indicates that for every 1mm increase in height, the odds of correctly stating which object was closer to the ground increased by a factor of 1.09.

H1: Does object shape matter? People were least accurate when assessing surface contact given the sphere, with 77% accuracy on average. In comparison, people were 80% accurate when presented with the cube and 84% accurate when presented with the icosahedron (See Figure 5.13). In our paired comparisons of people's accuracy for each target shape, when collapsed across shadow and height, we found that people's accuracy when they were shown the sphere was significantly worse than when their accuracy when shown the cube ($OR = 1.42, p < .01$) or the icosahedron ($OR = 1.67, p < .001$). In other words, participants were 1.42 times or 1.67 times more likely to choose the correct object with the cube or icosahedron, respectively, than with the sphere. There was not a significant difference in response accuracy between the cube and icosahedron ($OR = .85, p = .50$).

H2: Does shadow shading method matter? People were 77% accurate when assessing surface contact when presented with a dark shadow and 84% accurate when presented with a light shadow (See Figure 5.14). The main effect for shadow ($OR = 1.71, p < .001$) indicates that a correct response was 1.71 times more likely when the object was presented with a light shadow compared to a dark shadow.

H3: Is there an interaction between shape and shadow? Analysis of the shadow by shape simple slopes indicated that the effect of shadow was only significant for the cube ($OR = 4.21, p < .001$), a finding that indicates that people were more 4.21 times more likely to make a correct response when the cube was rendered with a light shadow. The predicted probabilities of correct response for each shape and shadow are displayed in Figure 5.15 (left). On average, when presented with the cube, people were 70% ($SE = 1.9\%$) accurate with the dark shadow when presented and they were 90% ($SE = 1.2\%$) accurate with the light shadow.

Video See-through AR

The the average probabilities of correct response for each main condition are shown in Table 5.4, and the results of our statistical analyses are reported in Table 5.5. For VST AR we again found a main ef-

5.3 Study 3 - Surface contact judgements are more accurate for a rectilinear object than for a sphere

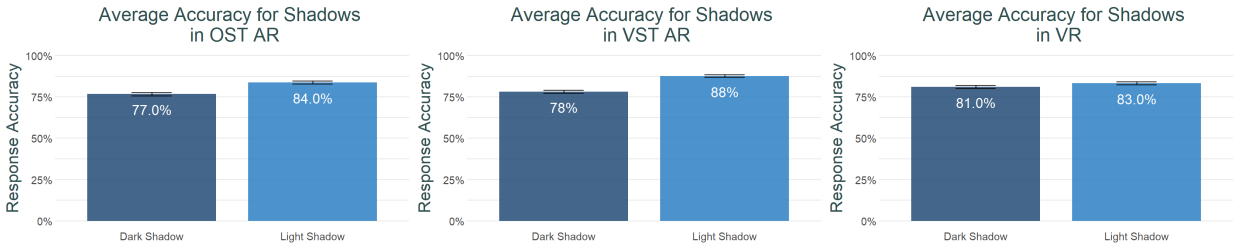

Figure 5.14: Experiment 3 — Average percentage of correct responses for each target object shadow by display condition: optical see-through AR (left), video see-through AR (center), and virtual reality (right). People were significantly more accurate when shadow shading was light, rather than dark in both AR device conditions.

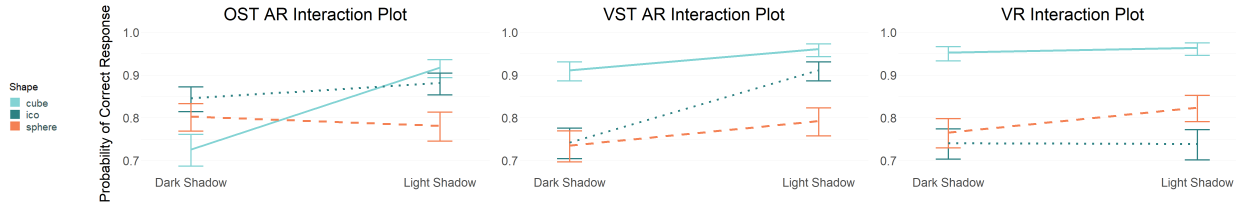

Figure 5.15: Experiment 3 — Predicted probability of correct response for shape (cube, icosahedron, sphere) and shadow (dark, light) interactions with 95% confidence. In VST AR, all shape x shadow interactions were significant. In contrast, for OST AR the effect of shadow was only significant for the cube and for VR it was only significant for the sphere.

fect of height ($OR = 1.07, p < .001$). Accordingly, as the height of the vertically displaced object increased, participants were more likely to correctly indicate which object was on the ground.

H1: Does object shape matter? People were most likely to produce a correct response when presented with the cube with 93% accuracy on average, and they were least likely to make a correct response when presented with a sphere with 75% accuracy. People's accuracy for the icosahedron was 81% on average (Figure 5.13). Planned contrasts for the shape variable indicated that the odds of providing a correct response significantly differed between all shape comparisons. The cube was more likely to yield a correct response than either the sphere ($OR = 4.90, p < .001$) or the icosahedron ($OR = 2.93, p < .001$). And the odds of providing a correct response for the icosahedron was higher than the odds for the sphere ($OR = 1.67, p < .001$).

H2: Does shadow shading method matter? The main effect for shadow ($OR = 2.28, p < .001$) indicates that a correct response was 2.28 times more likely when the object was presented with a light shadow compared to a dark shadow. Overall, people were 88% accurate on average when presented with the light shadow and 78% accurate when presented with the dark shadow (See Figure 5.14).

H3: Is there an interaction between shape and shadow? Analysis of the shadow by shape simple slopes indicated that the effect of shadow was significant for all three shapes. That is, the probability of providing a correct response was higher when a light shadow was presented regardless of whether the object presented was a sphere ($OR = 1.38, p < .05$), a cube ($OR = 2.40, p < .001$), or

Table 5.5: Experiment 3 — Results of planned comparisons using binary logistic regression models for each display condition are displayed. *B* is the regression coefficient, SE_B is the standard error of the regression coefficient, *OR* is the odds ratio, and CI_{OR} is the confidence interval associated with the odds ratio. Negative values for *B* indicate that the first factor in the comparison was more accurate, whereas positive values indicate that the second factor was more accurate.

Results for Experiment 3

		OST AR				VST AR				VR			
Predictor		*B*	SE_B	*OR*	$[CI]_{OR}$	*B*	SE_B	*OR*	$[CI]_{OR}$	*B*	SE_B	*OR*	$[CI]_{OR}$
Shape	(cube vs ico)	0.16	.12	0.85	[.64, 1.13]	−1.08 ***	.15	2.93	[2.06, 4.17]	−2.10 ***	.16	8.20	[5.66, 11.9]
	(cube vs sphere)	−0.35 **	.11	1.42	[1.09, 1.85]	−1.59 ***	.14	4.90	[3.51, 6.84]	−1.78 ***	.16	8.20	[4.08, 8.67]
	(ico vs sphere)	−0.52 ***	.11	1.67	[1.29, 2.17]	−0.51 ***	.11	1.67	[1.29, 2.17]	0.32 **	.20	1.06	[.57, .92]
Shadow	(dark vs light)	0.54 ***	.09	1.71	[1.43, 2.05]	0.82 ***	.11	2.28	[1.85, 2.83]	0.21	.11	1.23	[.99, 1.54]
Shape×Shadow	(cube: dark vs light)	1.44 ***	.17	4.21	[3.02, 5.85]	0.88 ***	.24	2.40	[1.49, 3.85]	0.27	.28	1.31	[.76, 2.27]
	(ico: dark vs light)	0.31	.16	1.36	[.99, 1.88]	1.28 ***	.17	3.58	[2.58, 4.98]	−0.01	.13	0.99	[.76, 1.29]
	(sphere: dark vs light)	−0.13	.14	0.88	[.66, 1.16]	0.32 *	.14	1.38	[1.05, 1.81]	0.36 *	.14	1.43	[1.08, 1.90]
Height		0.09 ***	.01	1.09	[1.08, 1.10]	0.07 ***	.01	1.07	[1.06, 1.08]	0.06 ***	.01	1.06	[1.05, 1.07]
Intercept		1.63 ***	.05	5.10	[4.62, 5.63]	1.88 ***	.06	6.55	[5.87, 7.35]	1.85 ***	.06	6.36	[5.69, 7.16]

$*p < .5$ $**p < .01$ $***p < .001$

an icosahedron ($OR = 3.58, p < .001$). The predicted probabilities of correct response for each shape and shadow are displayed in Figure 5.15 (center). The average accuracy for people's judgements with the cube was $90\%(SE = 1.2)$ on average for the dark shadow condition and $95\%(SE = 0.8)$ for the light shadow condition. For the icosahedron people were $73\%(SE = 1.8)$ accurate with the dark shadow and 90% $(SE = 1.2)$ accurate for the light shadow. For the sphere people were $72\%(SE = 1.8)$ accurate with the dark shadow and $78\%(SE = 1.8)$ accurate for the light shadow.

Virtual Reality

The average rate of correct response for each main condition is reported in Table 5.4, and the results of our statistical analyses are reported in Table 5.5. For VR we found a main effect of height ($OR = 1.06$, $p < .001$) such that for every 1mm increase in height, the odds of correctly stating which object was closer to the ground increased by a factor of 1.06.

H1: Does object shape matter? People were most likely to make a correct response when presented with the cube (95% accuracy on average), and they were least likely to make a correct response when presented with the icosahedron (73%). People's accuracy for the sphere was 78% (Figure 5.13). The probability of providing a correct response significantly differed for all shape comparisons. Specifically, planned contrasts, when collapsed across shadow and height, showed that participants were significantly more likely to make a correct response when presented with the cube than the icosahedron ($OR = 8.20, p < .001$) or sphere ($OR = 5.95, p < .001$). They were then more likely to make a correct response given the sphere over the icosahedron ($OR = .73, p < .01$).

H2: Does shadow shading method matter? The average accuracy for surface contact judgements was 81% for dark shadows and 83%for light shadows (See Figure 5.14). The main effect for

shadow was not significant ($OR = 1.23, p = .07$), suggesting that, averaged across shape and height, shadow shading did not have a significant effect on judgments of ground contact in virtual reality.

H3: Is there an interaction between shape and shadow? Analysis of the shadow by shape simple slopes indicated that the effect of shadow was only significant for the sphere ($OR = 1.43$, $p < .05$). The predicted probabilities of correct response for each shape and shadow are displayed in Figure 5.15 (right). On average, when presented with the sphere, people were 76% ($SE = 1.8\%$) accurate with the dark shadow when presented and they were 81% ($SE = 1.6\%$) accurate with the light shadow.

How difficult was each condition?

We examined post-experiment survey responses to better understand how participants interpreted our experimental stimuli. Participants were asked to rank the devices from easiest to hardest for determining ground contact, and they were asked to rank devices on graphics quality from lowest to highest. For difficulty rankings, they assigned each MR display to one of three categories: easiest, middle, or hardest. For quality of graphics rankings, they assigned each display to best, middle, or worst.

There was no clear easiest display, given that all three MR devices were rated as the easiest display by two out of six (2/6) participants. However, the AR devices were seen as more difficult than the VR condition, with both the OST AR display (4/6) and the VST AR display (2/6) receiving ratings for being the most difficult display. Participants also stated that the video see-through display had the lowest perceived quality of graphics with three out of six (3/6) worst votes and 3/6 median votes, and that the optical see-through display had the highest quality graphics (5/6).

Peoples' reported strategies in the post-experiment survey support the idea that the curved lines of the sphere and its shadow increased the difficulty of discerning surface contact. For discussion, we will refer to each participant with an acronym (e.g., P1 for the first participant). P1, P3, and P4 reported looking at shadows "*beneath*" the objects or "*beneath the front side*" of objects. P3 also commented that they tried to remember the point of optical contact for a given object. P2 and P5 explicitly stated that they looked for the "*edges*" of shadows to make judgments. A reliance on shadow edges would also explain why people's responses for the sphere were generally less accurate, regardless of MR display. Participants reported that they employed the same strategies for discerning ground contact across devices.

5.3.5 Discussion

In our third experiment, we again found that light shadows improved ground contact judgments compared to dark shadows for both AR devices and that they performed comparably to dark shadows in VR. These results encourage the use of non-photorealistic rendering solutions for improving surface contact judgments across MR displays. Perceived surface contact plays an important role in depth perception and improving it may help alleviate the disconnected appearance between virtual objects and real surfaces in AR.

In addition, we found that object shape influences surface contact judgments, where judgments were more accurate for the cube than the sphere in all three MR devices. Judgments for the icosahedron were more accurate than the sphere in VST AR and OST AR. Given that there was only case where people's judgements were more accurate for the sphere than another target shape, our current study

provides some evidence that spheres may be less effective than other shapes in establishing surface contact. This finding may have implications for depth perception in MR studies, as spheres are a common target object used in distance perception research for these devices [15, 135, 360, 361]. In addition, while our results for the main effect suggest an overall advantage for light shadows for AR, there were some differences in these effects for specific shapes and devices. The benefit of light shadows was largely driven by the cube in OST AR. But in VST AR all shapes were significantly influenced by the presence of a shadow.

There are limitations to the present experiment. All objects were viewed from a single vantage point, which may affect how generalizable our results are to other geometric shapes and perspectives in MR. This concern is supported by prior research, which has demonstrated that spatial perception is influenced by viewing angle [6, 261]. Some prior behavioral studies in AR have made conscious efforts to display objects at multiple orientations to subvert any possible effects of orientation [135]. Others have displayed objects at varying heights and distances [5]. Thus, to understand better how the placement of an object may affect people's ability to determine surface contact, we will evaluate target objects at different orientations in the next experiment.

A second limitation is that we do not evaluate any 3D objects that have a comparable complexity to those that might be used in typical AR and VR applications. All three objects that we evaluated were simple geometric shapes. As highlighted by Powell et al. [376], an object's geometry has the ability to influence one's perception of space. In Powell and colleagues' study, researchers evaluated people's ability to reach and grasp virtual spheres, icosahedrons, and apples in VR. They found that the richer geometry cues provided by the icosahedron and apple positively influenced reaching and grasping behaviors. In a similar manner, Do et al. [104] evaluated the effect of object color and luminance on objects of different shapes for depth judgments in mobile AR. They found interactions with color and luminance on depth perception. The findings of both of these studies motivate us to evaluate a more complex geometric shape in our next experiment.

5.4 Study 4 - Object complexity does not inherently improve surface contact judgements, but an object's orientation can

In this experiment we investigated if shape complexity and orientation influenced the perception of surface contact when dark and light shadows were present. Figures 5.16 and 5.16, which display the experimental stimuli used in Experiment 2, present a strong visual argument for investigating orientation given we can observe changes in shape and shading between an object and its cast shadows in these images with rotation. To address these concerns, we evaluate two target objects, a cube and the Stanford dragon, at two orientations, rotated by 0° and rotated by 45°, in our second study. Our hypotheses for these experiments were as follows:

H4: We anticipated that people would be worse at discerning the Stanford dragon's surface contact because it has fewer points in contact with the ground and the 3D model does not have a straight edge along the bottom—unlike the cube primitive. In essence, we argued that the likelihood of correct response would be more affected by the bottom edge of an object's geometry than its inherent complexity. This finding would provide evidence against the idea that more complex geometries are inherently better for depth perception [28, 104, 376].

H5: Based on the results from our third study, we predicted that people's surface contact judgments would be more accurate for the light shadow condition than the dark shadow condition across both

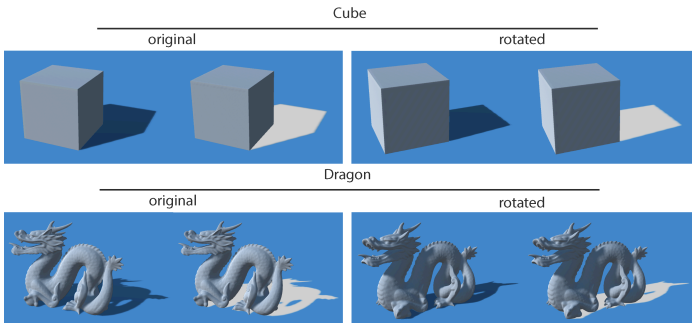

Figure 5.16: In Experiment 4, a cube and the Stanford dragon are rendered in each display and they are shown at two orientations. All other shading and displacement factors were the same as in Experiment 3.

AR devices.

H6: We predicted that, overall, a change in object orientation would affect participants' surface contact judgments since changes in orientation can alter both the perceived shape of the surface contact and its associated cast shadow for the viewer. Because the starting position (0°) was different for the cube and the dragon, it's possible that the 45° rotation might have deferentially affected the perception of the shapes. Therefore, we also include an interaction between shape and orientation in our analysis.

5.4.1 Participants

The same six individuals (3M, 3F) aged 23–30 volunteered to participate in our second experiment. Our experimental methods were approved by the local institutional review board, written consent was obtained from all subjects prior to participation, and each participant was paid 20 USD for 2 — 3 hours of their time.

5.4.2 Design

Our experimental design for the second evaluation mirrored the ones employed by our previous studies, in which a temporal two-alternative forced choice (2AFC) paradigm with a method of constant stimuli was employed. All experimental details were the same, except for the independent variables. We used the same testing environment, height displacement values (between 0 and $6mm$), the same number of comparisons, and the same number of trials. We employed two models, the cube from Experiment 1, and the Stanford Dragon,* tesselated to have 113, 000 polygons.

Each participant evaluated two cast shadow conditions (dark and light) over three different object + orientation conditions: the cube rotated by 45°, the Stanford dragon rotated by 0°, and the Stanford dragon rotated by 45°. This resulted in 6 unique combinations of experimental stimuli. However, our analysis evaluated 8 combinations of stimuli, as we included the non-rotated cube condition (at the two shadow conditions) from Experiment 1. This was done to prevent participant fatigue from an increased number of trials. In the second experiment, 1,980 points of data were collected from each volunteer for a total of 11, 880 datapoints. When combined with the cube data from the third experiment, each participant completed 2, 640 trials for a total of 15,840 trials across all subjects.

*http://graphics.stanford.edu/data/3Dscanrep/

5.4 Study 4 - Object complexity does not inherently improve surface contact judgements, but an object's orientation can

Using the same approach that we employed for Experiment 1, we evaluated the extent to which participant variance accounted for variance in our collected data by calculating the intraclass correlation coefficient (ICC) for Experiment 2. For the OST AR display condition we found $\tau_{ost} = 0.006$ and for the VST AR display we found $\tau_{vst} = 0.028$. For the VR display condition $\tau_{vr} = 0.023$. Because our ICC values were near zero, we determined that our experimental findings were not significantly influenced by between participant variation.

5.4.3 Procedure

The procedure for the final experiment followed the same protocol used in Experiments 2 and 3. However, this time volunteers did not undergo training for the devices and experimental paradigm since they had already performed the experimental task before.

5.4.4 Results

For our analyses, we used the binary logistic regression model approach that we employed in the previous study (See Section 5.3.4: Experiment 3) to analyze participants' 2AFC judgements for each of the three MR devices. Separate GLMs with logit link functions were run for each device to understand how an object's cast shadow shading, geometric complexity, and orientation affected people's surface contact judgments in each device. Specifically, we modeled binary outcomes (correct or incorrect) for our predictors: object shape, object orientation, shadow shading, and height. Height was recorded in millimeters then centered at zero and treated as continuous. Object shape (2 levels: cube and dragon), orientation (2 levels: rotated by 0° and by 45°), and shadow shading (2 levels: dark and light) were treated as categorical factors. These factors were deviation coded. For shadow shading, dark was coded as −.5 and light was .5; for shape, cube was −.5 and dragon was .5; for orientation, 0° was −.5 and 45° was .5.

We tested three planned comparisons for each device. These comparisons were: 1) whether surface contact judgments differed for different object geometries (i.e., **H4** the main effect of object shape); 2) the difference in surface contact judgments between dark and light shadow shading across all other conditions (i.e., **H5** the main effect of shadow shading); and 3) whether an object's orientation influenced people's surface contact judgments (i.e., **H6** the main effect of orientation). We also included an interaction between object shape and shading and an interaction between object shape and orientation to better understand the relationship between these variables.

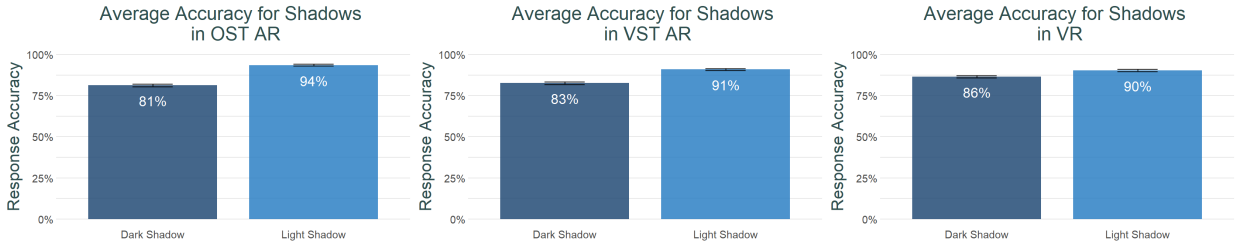

Figure 5.17: Experiment 4 — Average percentage of correct responses for each target object shadow by display condition: optical see-through AR (left), video see-through AR (center), and virtual reality (right). People were significantly more accurate when shadow shading was light, rather than dark in all three display conditions.

Table 5.6: Experiment 4 — The accuracy for each shape and shadow condition tested reported with the standard errors.

Accuracy for Experiment 4			
Condition	OST AR	VST AR	VR
Cube	85% [.7%]	92% [.5%]	94% [.5%]
Dragon	90% [.6%]	81% [.8%]	83% [.8%]
Dark	81% [.8%]	83% [.8%]	86% [.7%]
Light	94% [.5%]	91% [.6%]	90% [.6%]
Origin (0°)	84% [.7%]	87% [.7%]	86% [.7%]
Rotated (45°)	91% [.6%]	87% [.7%]	91% [.6%]

Optical See-through AR

We show the average percent of correct response for each of the evaluated main effects in Table 5.6, and the results of the logistic regression for the OST AR display data are presented in Table 5.7. We expected participants to be more likely to correctly indicate which object was on the ground as the height of target objects increased. The improvement of participants' performance as height increased is demonstrated statistically by a main effect of height in our logistic regression model ($OR = 1.13$, $p < .001$). An odds ratio of 1.13 indicates that for every 1mm increase in height, the odds of correctly stating which object was closer to the ground increased by a factor of 1.13.

H4: Does an object's complexity matter? People were more accurate when assessing surface contact with the dragon than the cube, with 90% and 85% accuracy on average overall (Figure 5.18). Our statistical analysis revealed that this relationship was significant since—when averaged across shadow shading, orientation, and height—the dragon shape in our model was more likely to elicit correct responses than the cube with an odds ratio of 1.38 ($p < .01$). As such, people were 1.38 times more likely to correctly assess surface contact when presented with the dragon. This outcome differs from what we observed in both the VST AR and VR conditions as reported in Sections 5.4.4 and 5.4.4. In these devices surface contact judgements to the cube were more accurate than judgements to the dragon .

H5: Does shadow shading method matter? The main effect for shadow ($OR = 3.70, p < .001$) indicates that a correct response was 3.70 times more likely when the object was presented with a light shadow compared to a dark shadow, when collapsed across shape, orientation, and height. On average, people were 94% accurate when presented with the light shadow and 81% accurate when presented with the dark shadow (See Figure 5.17). There was also a significant interaction between shape and shadow ($OR = .41, p < .001$). Analysis of the shadow simple slopes by shape indicated that, for both shapes, a light shadow was more likely to yield a correct judgment of ground contact than a dark shadow (supporting the main effect of shadow). The simple slopes are depicted in Figure 5.19. However, this effect was stronger for the cube ($OR = 5.79, p < .001$) than the dragon ($OR = 2.37$, $p < .001$).

5.4 Study 4 - Object complexity does not inherently improve surface contact judgements, but an object's orientation can

H6: Does object orientation matter? There was a significant main effect of orientation ($OR = 1.96, p < .001$) indicating that, overall, participants were more likely to make a correct judgment when the object was rotated 45° from its starting position. On average, people's judgements to targets at their original orientation were 84% accurate and their judgements to rotated targets were 91% accurate in the HoloLens. This main effect was qualified by an interaction between object shape and orientation ($OR = .60, p < .01$). In order to interpret the interaction, we calculated the orientation simple slopes for each shape. For both shapes, the likelihood of correctly judging ground contact was higher for the 45° orientation than the 0° orientation (supporting the main effect of orientation). However, this effect was stronger for the cube ($OR = 2.53, p < .001$) than for the dragon ($OR = 1.52$, $p < .01$).

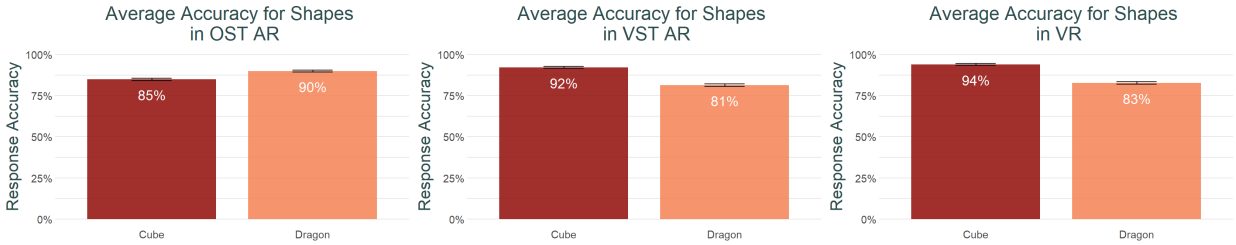

Figure 5.18: Experiment 4 — Average percentage of correct responses for each target object shape by display condition: optical see-through AR (left), video see-through AR (center), and virtual reality (right). Object complexity did not inherently benefit surface contact judgements. People's judgements were significantly more accurate when presented with the cube than the dragon in both the VST AR and VR conditions. The opposite was true in OST AR.

Video See-through AR

We show the average percent of correct response for each of the evaluated main effects in Table 5.6. The results of our statistical analyses are reported in Table 5.7. For VST AR we again found a main effect of height ($OR = 1.09, p < .001$) such that for every 1mm increase in height, the odds of correctly stating which object was closer to the ground increased by a factor of 1.09. Accordingly, as the height of the vertically displaced object increased, participants were more likely to correctly indicate which object was on the ground.

H4: Does an object's complexity matter? The cube shape was more likely to elicit correct responses than the dragon ($OR = .32, p < .001$), suggesting that object complexity is not inherently beneficial to surface contact perception. People's judgements were 92% accurate on average when assessing surface contact with the cube and 81% accurate on average when assessing contact with the dragon (Figure 5.18).

H5: Does shadow shading method matter? The main effect for shadow ($OR = 2.47$, $p < .001$) indicates that a correct response was 2.47 times more likely when the object was presented with a light shadow compared to a dark shadow. On average, people were 91% accurate with the light shadow and they were 83% accurate with the dark shadow when evaluating surface contact (See Figure 5.17). In addition, there was a significant interaction between shape and shadow ($OR = .61$,

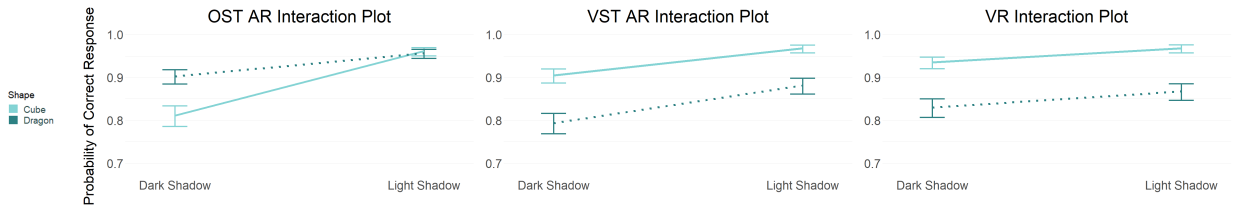

Figure 5.19: Experiment 4 — Predicted probability of correct response for shape (cube, dragon) and shadow (dark, light) interactions with 95% confidence. For all three devices a light shadow was more likely to elicit a correct judgment of surface contact than a dark shadow, regardless of shape. In addition, this effect was stronger for the cube than for the dragon.

$p < .05$). Analysis of the shadow by shape simple slopes indicated that the effect of shadow was significant for both shapes (See Figure 5.19 (center)). For both shapes, the probability of correct response was higher when a light shadow was presented compared to a dark shadow. However, this effect was stronger for the cube ($OR = 3.17, p < .001$) than the dragon ($OR = 1.93, p < .001$).

H6: Does object orientation matter? There was no main effect of orientation in VST AR. People's surface contact judgements to targets when positioned at their original orientations and when rotated were 87% accurate on average. The interaction between orientation and shape was also insignificant (See Figure 5.19). In other words, the probability of providing a correct response was not significantly different between the two orientations in VST AR. This outcome differs from the results of our analyses in the OST AR and VR conditions, both of which found that people's judgements were more accurate on average to the rotated targets.

Virtual Reality

The average rate of correct response for each main condition is reported in Table 5.6, and the results of our statistical analyses are reported in Table 5.7. For VR we found a main effect of height ($OR = 1.06$, $p < .001$) such that for every 1mm increase in height, the odds of correctly stating which object was closer to the ground increased by a factor of 1.06.

H5: Does an object's complexity matter? The cube was more likely to elicit correct responses than the dragon ($OR = .27, p < .001$), suggesting that object complexity is not inherently beneficial to surface contact perception. People's judgements were 94% accurate on average when assessing surface contact with the cube and 83% accurate on average when assessing contact with the dragon (Figure 5.18).

H4: Does shadow shading method matter? The main effect for shadow ($OR = 1.68, p < .001$) indicates that a correct response was 1.68 times more likely when the object was presented with a light shadow compared to a dark shadow. Judgements of surface contact were 90% accurate on average in VR with the light shadow and they were 86% accurate on average with the dark shadow (See Figure 5.17). The interaction between shape and shadow was also significant ($OR = .64, p < .05$). Unlike our results for Experiment 1, the analysis of the shadow by shape simple slopes indicated that the effect of shadow was significant for both shapes in VR (See Figure 5.19 (right)). The probability of correct

Table 5.7: Experiment 4 — Results of planned comparisons using binary logistic regression models for each display condition are displayed. *B* is the regression coefficient, SE_B is the standard error of the regression coefficient, *OR* is the odds ratio, and CI_{OR} is the confidence interval associated with the odds ratio. Negative values for *B* indicate that the first factor in the comparison was more accurate, whereas positive values indicate that the second factor was more accurate.

Results for Experiment 4

		OST AR				VST AR				VR			
Predictor		*B*	SE_B	*OR*	$[CI]_{OR}$	*B*	SE_B	*OR*	$[CI]_{OR}$	*B*	SE_B	*OR*	$[CI]_{OR}$
Shape	(cube vs dragon)	0.32 **	.11	1.38	[1.12, 1.70]	−1.15 ***	.10	0.32	[.26, .39]	−1.31 ***	.11	0.27	[.22, .33]
Shadow	(dark vs light)	1.31 ***	.10	3.70	[3.02, 4.55]	0.91 ***	.10	2.47	[2.03, 3.04]	0.52 ***	.11	1.68	[1.36, 2.08]
Orien	(0° vs 45°)	0.67 ***	.10	1.96	[1.62, 2.38]	−0.06	.10	0.94	[.78, 1.14]	−0.13	.11	0.88	[.71, 1.08]
Shape×Shadow		−0.90 ***	.21	0.41	[.27, .61]	−0.50 *	.21	0.61	[.40, .91]	−0.45 *	.21	0.64	[.42, .97]
	(cube: dark vs light)	1.76 ***	.15	5.79	[4.35, 7.71]	1.15 ***	.17	3.17	[2.25, 4.46]	0.74 ***	.18	2.11	[1.47, 3.02]
	(dragon: dark vs light)	0.87 ***	.15	2.37	[1.77, 3.17]	0.66 ***	.11	1.93	[1.55, 2.40]	0.29 **	.11	1.34	[1.08, 1.67]
Shape×Orien		−0.51 **	.20	0.61	[.41, .88]	0.20	.19	1.22	[.84, 1.79]	0.77 ***	.21	2.16	[1.43, 3.29]
	(cube: 0° vs 45°)	0.93 ***	.13	2.53	[1.95, 3.27]	−0.16	.16	0.85	[.63, 1.16]	−0.52 **	.18	0.60	[.42, .85]
	(dragon: 0° vs 45°)	0.42 **	.14	1.52	[1.15, 2.02]	0.04	.11	1.04	[.84, 1.29]	0.26 *	.11	1.29	[1.04, 1.61]
Height		0.12 ***	.01	1.13	[1.11, 1.14]	0.09 ***	.01	1.09	[1.08, 1.11]	0.08 ***	.01	1.08	[1.07, 1.10]
Intercept		2.50 ***	.07	12.13	[10.71, 13.83]	2.25 ***	.06	9.50	[8.51, 10.67]	2.39 ***	.06	10.88	[9.69, 12.28]

$*p < .5$ $**p < .01$ $***p < .001$

response was higher when a light shadow was presented regardless of the object presented, although the shadow effect was stronger for the cube ($OR = 2.11, p < .001$) than the dragon ($OR = 1.34, p < .001$).

H6: Does object orientation matter? The main effect of orientation was not significant. However, there was a significant interaction between shape and orientation ($OR = 2.16, p < .001$). On average, people's judgements to targets at their original orientation were 86% accurate and their judgements to rotated targets were 91% accurate. Analysis of the orientation by shape simple slopes indicated that a correct response was more likely when the cube was presented at 0° than when the cube was rotated 45° ($OR = .60, p < .01$). Figure 5.19 displays the simple slopes. In contrast, the probability of a correct response was higher when the dragon was rotated 45° than when it was presented at 0° ($OR = 1.29, p < .05$).

How difficult was each condition?

After completing the experiment, volunteers were asked to rank each device by difficulty and by quality of graphics as in Experiment 1. All participants ranked the video see-through display as the most difficult device (6/6) and the device with the lowest quality of graphics (6/6). In contrast, the optical see-through display was most consistently rated as the easiest display (4/6) and the display with the highest quality of graphics (5/6). Ratings were also corroborated by comments at the end of the survey. Both P2 and P4 stated that the HoloLens "*felt more clear*," and P1 commented that "*the resolution of the Zed was really painful for my eyes.*"

Participants were also asked about the strategies they used to discern surface contact. P3 commented that the dragon was "*easier but more mentally exhausting... since the dragon moved.*" Their comment implied that they had to visually search for a shadow before making a judgment. Their complaint is

reasonable since, as the shape of the dragon was asymmetrical, changes in orientation were more pronounced (See Figure 5.16). P3 further bemoaned of the dragon that "*when too detailed it was overwhelming, [but] when not detailed enough it became equally difficult to use*." P2 commented that the dragon provided more "*pockets of shadow,"* which altered their strategy such that " *It wasn't just watching an edge, more like watching for shape and quantity of shadow.*" This comment hints that people may have been able to use additional depth information provided by the dragon (e.g., shape from shading) to inform their judgments of surface contact.

5.4.5 Discussion

In Experiment 4, for both AR display conditions, we again found that people's judgments of surface contact benefited from the presence of non-photorealistic, light cast shadows. However, in contrast to our third study, we also found that light shadows improved surface contact judgments in VR. Looking at shape $\times$ shadow interactions for the cube in Experiment 3 and Experiment 4 in VR may provide some insight into this difference in shadow effectiveness for VR between experiments. Both experiments evaluated the same cube object, but the object was presented at a new orientation in Experiment 4. Because we found a significant difference in orientation for the cube and because we found an effect of shadow shading condition in Experiment 4, we infer that the change in orientation for the cube made the light shadows more important in VR.

Our hypothesis that surface contact judgments for the dragon would be worse than the cube was partially supported. Although we found better performance with the cube compared to the dragon in both VST AR and VR, we found better performance with the dragon in OST AR. With the current experiment alone, we are unable to explain why people's judgments for the dragon in the OST AR device alone were significantly more accurate than those for the cube. However, the current study provides evidence that more complex geometries do not inherently benefit surface contact perception. Although complex 3D shapes can provide more depth information via self-shading cues, it appears that other factors such as the bottom edge of the object may be more important. However, additional future research is needed to verify this claim.

Given the effects of orientation that we found in both the OST AR and VR displays, it will be important to consider the orientation of objects in MR in future application development. This finding is in line with prior research which shows that the angle from which we view a 3D object may affect spatial perception judgments related to that object[6, 261]. The lack of difference in orientation for the video see-through conditions requires further investigation.

5.5 Summary

The results of our research provide evidence that nontraditional shading techniques for rendering shadows in AR displays may enhance the accuracy of one's perception of surface contact. This finding implies a possible tradeoff between photorealism and accuracy of depth perception, especially in OST AR displays. However, it also supports the use of more stylized graphics like non-traditional cast shadows to improve perception and interaction in AR applications. We also find that geometric shape can affect one's perception of ground contact in subtle but tangible ways.

In the current work, we make several contributions by evaluating how people's judgments are affected by virtual object shape, orientation, and shadow shading in three different MR displays. We

demonstrate that non-photorealistic, light shadow shading can enhance surface contact judgments. We show that these judgments are complex and affected by an object's shape as well as orientation. And we conduct fundamental groundwork that encourages the further study of non-photorealistic rendering techniques to improve spatial perception in MR.

Perception of surface contact is important for determining the scale of space and for governing our interactions within space. Cast shadows provide important information about surface contact, and thus provide important cues to our visual system for perceiving and acting with our surroundings [5, 333, 467]. In certain types of MR displays, rendering cast shadows is difficult. Therefore, this paper manipulated cast shadow rendering across different object shapes in MR displays. Our goal was to see if a unified framework would emerge to help designers better design for accurate spatial perception in MR, given the difficulties with rendering realistic, dark shadows in current technology.

We hypothesized that people would be more likely to correctly perceive surface contact when shadows were illuminated with light rather than dark color values. This hypothesis was based on prior work [5, 228]. We found support for these hypotheses (**H1** and **H3**) for two types of AR devices across two experiments. We also hypothesized that surface contact judgments would be complex and affected by the shape of an object associated with a given surface (**H2**). Loosely speaking, we formed this hypothesis around the idea that if a shadow forms visible glue with a surface, then the object to which that glue is applied would also affect how well the shadow worked [295, 467]. We found support for this hypothesis as well, although we evaluated relatively simple object shapes in our first study. In our second study, we then investigated the effect of an object's complexity and orientation on surface contact judgments. Consistent with the reasoning above, we hypothesized that orientation would affect the perception of ground contact (**H5**). We found that orientation mattered for surface contact judgments in OST AR and VR, but we were unable to find a difference in VST AR, only partially confirming this hypothesis. We also hypothesized that object complexity would matter, and that it would be more difficult to perceive surface contact with more complex shapes (**H4**). This hypothesis was confirmed in VST AR and VR, but not for OST AR, where the opposite was true.

One finding that we can extract from this set of experiments is that non-photorealistic (light) shadows will likely work well across MR devices if the application can support such shadows. This finding was clear in both AR displays, and we saw this effect in VR in our second experiment. There are many AR applications in which perception may be important, but photorealism is not as critical (e.g., AR games [164, 227] or training applications where perception is more critical than realism [180, 284]). Our findings give designers some latitude for general design.

Our findings on shape, complexity, and orientation are more complicated. It may be that they do not generalize well across MR devices and that general guidelines are not appropriate here, but more work is needed. For object shape, we found only one case across all three MR devices where people's judgments were more accurate for the sphere than another target shape (where judgments were more accurate for the sphere than the icosahedron in VR). Do et al. [104] found that people's depth judgments to spheres were more sensitive to changes in luminance than other shapes in mobile AR. Both our results and those of Do et al. [104] may caution against the use of spheres for AR applications that require accurate spatial perception. Finally, our findings on orientation may have implications for the development of MR applications. The angle from which we view a 3D object can affect spatial perception judgments related to that object [6, 261]. For the development of MR applications, however, understanding how specific shapes and orientations of objects influence where people perceive them to be positioned in space may be important.

A limitation of our method is that, although we can assert with confidence the conditions under

which surface contact judgments for light shadows are better than for dark shadows, our methods do not tell us that dark shadow contact judgments are *bad* in an absolute sense. We employed a psychophysical technique for the reason that we were interested in determining whether light shadows were better than dark shadows. We observed a problem with objects being floaty in AR, we acknowledged problems with rendering shadows in AR, and then we attempted to understand it in the larger context of which tools are available to us in an MR setting to try to offer a solution.

Another limitation is that we did not manipulate object shading in our experiments because we were more interested in the effect of an object's shape on ground contact judgments. However, given the sensitive effects of shadow luminance values on surface contact judgments observed in our studies, it would be reasonable to claim that our results were influenced by the color of our target objects. Similar prior research on surface contact perception in MR has evaluated the effect of differences in object and shadow shading on people's surface contact judgments [5]. Based on this prior research, we elected to use a median gray value to shade all of our test objects to mitigate effects of object shading, although we recognize that this is an imperfect solution.

Neither the current work nor the prior work conducted by Adams et al. [5] manipulate the color of background surfaces on which target objects are displayed. Yet just as an object's shading may influence our spatial perception of that object in space [6, 102, 104, 372], so too may the backdrop upon which it is presented [352]. In our case, we used the same, median blue tablecloth throughout both experiments. Further work will need to test background surfaces of varying color and texture to fully understand how these may interact with shadows in MR devices.

This paper demonstrates that there are advantages to non-photorealistic rendering of cast shadows for surface contact judgments in MR. We found that judgments were better with light shadows in two types of AR devices and in VR under certain conditions. In arriving at this finding we experimented with a variety of object shapes, orientations, and complexities. Our findings suggest that under certain circumstances it may be desirable to use light shadows in MR applications.

Because the light shadows in our current study enhanced people's accuracy in surface contact judgments—it may be worthwhile to evaluate whether more ambitious non-photorealistic rendering approaches facilitate spatial perception in AR. Evaluating colorful shadows, like those designed by Ooi et al. [345], for AR may be a desirable starting point. If more research findings are able to confirm the benefits for non-photorealistic rendering for spatial perception in AR, then we may encourage the use of more stylized graphical elements for designers and developers of MR applications. However, we must be cautious about arbitrarily applying non-photorealistic effects to MR simulations before evaluating them, especially given the results of Cidota et al. [81], who showed that blur and fade may have adverse effects on action-based depth judgments in augmented reality in personal space.

Our current findings may have ramifications for egocentric depth judgments in MR displays given prior work [333], but at present we can only generalize our findings to surface contact judgments. In future research, it may be worthwhile to evaluate how object shape and cast shadow shading manipulations affect more direct measures of depth perception. Rosales et al. [397] demonstrated that, in the absence of cast shadows, people perceive an object that is placed above the ground incorrectly as farther away. This may help explain some of the effects of overestimation found in prior AR depth perception research, especially given that many assessments use floating objects for assessment [360, 429, 456]. In future work, we intend to evaluate how photorealistic and non-photorealistic shadows affect egocentric depth judgments in MR displays.

Part IV

Depth Perception

6 Egocentric Depth Perception in MR

The ability to perceive absolute distances to objects in AR is critical to many applications that require an understanding of scale, such as environmental simulations for training and architectural design. Nonetheless, open questions remain about human spatial perception in augmented reality and the depth cues that are used when virtual and real worlds coexist. For example, researchers have yet to isolate what properties of the visual information provided by augmented reality displays affect people's ability to correctly perceive virtual objects in real spaces (See Section 2.4). In Aim 2 of the dissertation we hope to address these concerns.

Our investigation into the influence of realistic and nonrealistic rendering approaches on depth perception in AR began in Chapters 4 and 5 with a series of studies that investigated the influence of shadow shading and other object properties on surface contact perception. In these studies, we found that the absence of a cast shadow resulted in less confident surface contact judgments between an object and a surface beneath it (Section 5.1). We then found evidence that photorealistic rendering approaches for cast shadows did not further enhance people's surface contact judgements. On the contrary, a non-photorealistic cast shadow shading method—a white shadow—consistently improved people's surface contact judgments in virtual reality, optical see-through AR, and video see-through AR HMDs. This research, which was conducted for Aim 1 of the dissertation, also raised two questions: (1) to what extent does the vertical displacement of an object above the ground influence people's depth judgements and their ability to use differently rendered cast shadows as a depth cue? and (2) do our results on surface contact perception generalize to other measures of depth perception?

Prior research has shown that the vertical position of an object (e.g., whether an object is positioned on or above the ground) can influence where people perceive the object in space [332, 333, 397]. Even when objects are floating well above the ground such that they are disconnected from their cast shadows, cast shadows still play an important role in determining the position of the object in space since they allow a viewer to estimate the position of the object relative to other geometric features and surfaces within the environment [228, 255, 332].

In our previous research, we only evaluated subtle vertical displacements of objects since surface contact judgments were our measure of interest. By using absolute egocentric distance estimations (i.e., judgements of distances from the viewer to an object in space) we may also extend our understanding of the influence of cast shadows on depth perception to objects that are positioned above the ground. For user interaction in AR HMDs, it is particularly important to understand how mid-air virtual objects are perceived given the prevalence of user interfaces in AR that are positioned above the ground for ease of viewing. Distortions in depth perception of virtual objects may make interactions more challenging or frustrating for AR users.

Evaluating surface contact judgments is not an approach that is commonly used to investigate depth

perception or the perception of spatial layout. Our first experimental protocol was based off the work of a single publication, Madison et al. [295]. And the experimental protocol we employed for the subsequent studies relied on a psychophysical approach (specifically a temporal 2AFC procedure). Although alternative forced choice experiments have been employed in the past to study depth perception judgements from a viewer to objects in space [48, 104, 161, 451, 457], they have not been used to study surface contact perception. The dearth of research that directly evaluates surface contact perception makes it difficult to determine how our results may generalize to other measures of depth perception. As such, in Aim 2 we extend the results of our prior research on surface contact perception to a more prevalent measure of depth perception by conducting two evaluations on absolute egocentric distance judgements in AR. The investigation of absolute distance perception may allow us to draw comparisons between our research on surface contact perception and non-photorealistic shadow shading to a larger body of literature.

The current work was the first to evaluate absolute distance perception in either device. In our first study (Section 7.1), we asked participants to verbally report distance judgments to grounded and floating targets that were rendered either with or without a cast shadow along the ground. Then, in our second depth perception study (Section 7.2), we evaluated people's distance judgements to the same targets with either realistic shadows or non-photorealistic graphical elements beneath them.

The findings from our initial depth perception study suggest that currently available video see-through displays may induce more distance underestimation than their optical see-through counterparts. Interestingly, the presence of cast shadows improved depth judgments for both grounded and floating target objects, although floating targets were still consistently perceived as farther away even when a shadow was present.

Informed by the outcomes of our prior surface contact studies, in the second depth perception experiment we hypothesized that target objects presented with non-photorealistic graphical elements, in lieu of shadows, would elicit distance judgments with equivalent accuracy to target objects that were presented with realistic shadows. Our findings offer a two-fold contribution that strengthens the foundation for developing more inclusive augmented reality (AR) applications. First, they reinforce the validity of our prior research on surface contact perception. This is achieved by demonstrating that the observed effects in those studies were not simply due to artifacts of experimental design (e.g., the experimental task or characteristics of the virtual stimuli)—but rather reflect a genuine relationship between surface contact perception and spatial perception. Second, our findings provide compelling new evidence that achieving accurate spatial perception in AR environments does not require highly realistic graphics. This has significant implications for accessibility, as prior research by Zhao et al. [517] suggests that individuals with vision impairments often benefit from the use of stylized renderings, such as edge enhancements and contrast amplification techniques in AR.

6.1 Motivation and Goals

Few comparisons of depth perception between stereoscopic VST AR and OST AR head-mounted displays exist [6, 41, 311]. However, direct comparisons of how different AR displays influence spatial perception can provide important insights into how the technical tradeoffs between AR HMDs influence perception. Direct comparisons are also beneficial for establishing developer guidelines for where and how virtual objects should be rendered in real spaces to enhance depth perception.

As such, in the following experiments we evaluated people's depth perception using absolute ego-

Table 6.1: Specifications from the two augmented reality displays employed in the current study are presented.

Display Specifications	Varjo XR-3	HoloLens 2
field of view (FoV)	115 × 90°	43 × 29°
focus: resolution per eye (pixels)	1920 × 1920	2048 × 1080
focus: pixel density (PPD)	70	47
peripheral: resolution per eye (pixels)	2880 × 2720	-
peripheral: pixel density (PPD)	30	-
refresh rate (Hz)	90	60
weight (g)	980	566
price (USD)	5500	3500

centric distance judgments in two augmented reality displays: the Microsoft HoloLens 2 and the Varjo XR-3. Although both devices provide augmented reality via HMD, the technology behind them considerably differs.

The Varjo XR-3 functions both as a pure virtual reality device and a video see-through AR display. As a result, it shares similar ergonomic and display properties to contemporary virtual reality displays—although the XR-3 boasts exceptional video see-through capabilities. In contrast, the HoloLens 2 is a lightweight display that relies on optical see-through technology to create augmented reality. Instead of using a display panel to render virtual overlays, the HoloLens 2 projects overlays onto a plastic shield in front of the viewer's eyes. As a result, in the HoloLens viewers have an unaltered view of the real world, but the augmented field of view in OST AR is much smaller. For reference, a more thorough comparison between the two device specifications can be seen in Table 6.1.

The unique mechanical properties and rendering approaches of these devices have ramifications for how virtual objects are rendered and integrated into real world scenes. One consequence of employing video feed to capture real world images for VST AR is that cameras can introduce optical aberrations, like lens distortions, to the real world image. In contrast, although OST AR displays do not distort real world images, the use of additive light to render virtual overlays causes virtual objects to appear transparent. The darker the color value; the more transparent the overlay. In certain lighting scenarios, like outside on a sunny day, this may cause pictorial depth cues, like cast shadows, to be less salient or even imperceptible.

Figure 6.1: The optical see-through (OST) and the video see-through (VST) augmented reality display evaluated in Experiments 5 and 6. On the left is the Varjo XR-3 (VST AR) and on the right is the Microsoft HoloLens 2 (OST AR).

In addition to AR device, we carefully selected two attributes of virtual targets to better understand how these attributes influence the perception of absolute distances. The first was the vertical position of a virtual object. In order to make gazing at overlays more comfortable, AR applications may present virtual objects as floating, or vertically displaced above the ground (e.g., Google Maps AR). Perhaps because of this, much of the prior research investigating depth perception in AR has used floating virtual targets for assessment [102, 135, 175, 255, 360, 361, 429]. However, the decision to evaluate floating targets may have an undesired effect on people's depth judgments.

The distal horizon as well as the height of an object relative to the ground influences depth perception judgments [314, 315]. As such, the position of an object relative to a surface, like whether the object is floating in air or anchored on the ground, alters where we perceive that object to be positioned in space [332, 352]. Because the human visual system treats floating objects as though they are located on the ground plane, floating targets are typically perceived as farther away [144, 146]. In augmented reality, as well, the influence of optical contact on distance perception has been demonstrated in the Microsoft HoloLens 1 [397]. Specifically, Salas-Rosales and colleagues demonstrated that floating virtual targets were perceived as on the ground but farther away in AR when no surface contact information, like cast shadows, was present.

The second factor we evaluated was cast shadow. In this dissertation, we have found that the presence of a cast shadow improves the confidence of people's surface contact ratings [5]. For floating targets, prior research by Ni et al. has shown that the presence of a cast shadow can mitigate the influence of optical contact in virtual environments [333]. As a result, people give more accurate egocentric distance judgments to floating targets when cast shadows are present. Motivated by these findings, a growing body of depth perception research in augmented reality has looked at the effects of cast shadow on floating targets [102, 135, 175]. However, none of this prior research has looked at the effects of cast shadow for targets placed on the ground. As such, it is difficult to extrapolate the results of this prior research to general design guidelines for AR applications, which may be expected to place virtual objects at various heights.

In addition, all of the prior research on the influence of cast shadows on distance perception have relied on perceptual matching, a relative—not an absolute—measure of distance perception [121, 132]. Conducting distance estimation research using different evaluation methods is important for generalizing effects, since distinct experimental paradigms require participants to encode different spatial information [153, 232]. We bridge this gap in literature by evaluating the influence of cast shadows on depth perception using an absolute measure of distance: verbal report.

6.2 General Methods

In order to evaluate how shadow shading methods affect a viewer's distance estimations in head-mounted mixed reality displays, we conducted two experiments across two unique display conditions: an optical see-through augmented reality display (the Microsoft HoloLens 2) and a video see-through augmented reality display (the Varjo XR-3). The subsequent sections discuss the technical setup of our experiments (Section 6.2.1) as well as the procedure of our experiments (Section 6.2.2). For rendering shadows, we employed the same methods that were used in our surface contact perception experiments (Section 4.2.2).

6.2.1 Materials

The experiment was conducted in a 36 x 26 x 9 ft room that provided a 36 ft linear distance forward for placing targets. A university classroom was reserved throughout the duration of the experiment and tables aligned both sides of the participant during the experiment. Images of the room can be seen in Figure 7.2.

We conducted the experiment in an optical see-through augmented reality HMD and a video see-through augmented reality HMD. For the OST augmented reality display condition, we employed the Microsoft HoloLens 2. The HoloLens 2 weighs 566g and has a field of view (FoV) of $43° \times 29°$. Position tracking in the HoloLens 2 is performed by its native inside out spatial tracking method. For the VST augmented reality condition, we used the Varjo XR-3. The XR-3 weighs 980g and has a FoV of $115° \times 90°$. For position tracking, the tethered display used the SteamVR 2.0 tracking system in conjunction with the Varjo's native depth sensors, which relied on LiDAR and RGB camera fusion.

The Varjo XR-3 functions both as a pure virtual reality device and a VST AR display. As a result, it shares similar ergonomic and display properties to contemporary virtual reality displays—although the XR-3 boasts exceptional video see-through capabilities. In contrast, the HoloLens 2 is a lightweight display that relies on OST technology to create augmented reality. Instead of using a display panel to render virtual overlays, the HoloLens projects overlays onto a plastic shield in front of the viewer's eyes. As a result, in the HoloLens viewers have an unaltered view of the real world, but the augmented field of view in OST AR is much smaller and virtual overlays appear translucent since these displays rely on additive light for rendering. Without the ability to subtract light, OST AR displays are unable to render dark color values, which instead become increasingly transparent as they approach black. A more thorough comparison between the two device specifications can be seen in Table 6.1.

Applications for both devices were developed in Unity version 2020.3.13f1 with the C# programming language. Shaders to render hard shadows were programmed using a variant of the HLSL language that is compatible with the Unity game engine. The cast shadow shader was developed to render shadows with specified color values. Because the HoloLens 2 is unable to render black, a shadow with a grayscale RGB value of 36 was selected. The same shaders were used for both devices.

6.2.2 Procedure

Participants were met at the door of the classroom, where they were given a description of the experiment, an informed consent form, a proof of payment form, as well as their monetary compensation for volunteering to participate in the study. The study followed Covid-19 safety protocols set by the university. As such, all participants were required to wear masks and equipment was sanitized between participants.

Before introducing the volunteer to the augmented reality equipment, the experimenter familiarized the participant with basic distance measures in an adjacent hallway. Depending on the participant's preference, either metric or imperial units of measure were reviewed using a retractable tape measure. Reviewed distances did not exceed 1 meter or 1 yard. After the participant expressed that they were comfortable with the distances, the experimenter guided them back into the classroom.

The participant was then outfitted with the first head-mounted display, and the protocol was described to them. The were told that "target object would appear at various distances" along the floor relative to the viewer. Each target object appeared for five seconds before disappearing. At which point, the participant called out the estimated distance to the target. After the experimenter transcribed the

participant's response, the next trial commenced. The beginning of a subsequent trial was denoted by the sound of a beep. The participant was given no feedback on their performance during the experiment.

Isolated facts and experiments have in themselves no value, however great their number may be. They only become valuable in a theoretical or practical point of view when they make us acquainted with the law of a series of uniformly recurring phenomena, or, it may be, only give a negative result showing an incompleteness in our knowledge of such a law, either by its not being perfect

Hermann von Helmholtz

7 Investigating Egocentric Depth Perception in MR

This research investigates how realistic and non-realistic rendering affect depth perception in augmented reality. It builds upon the findings of our earlier research on surface contact perception, in which we demonstrated that non-photorealistic rendering approaches for cast shadows could improve surface contact judgments compared to judgements made with realistic or absent shadows. In the second aim of this dissertation, we extend this work to measures of egocentric distance perception to better determine if this phenomenon generalizes to another dimension of spatial perception.

Over two experiments, we measure absolute egocentric distance judgments (estimating distances from the viewer to an object) to targets with different shadow rendering conditions in an optical see-through and a video see-through head-mounted display. In doing so, we hope to expand our understanding of how shadows affect depth perception in general and how the unique graphics limitations of these different AR devices impacts the utility of shadows for depth perception.

Figure 7.1: In Study 5, participants viewed spherical targets that were positioned on and above the ground. These targets were either rendered with or without a cast shadow.

7.1 Study 5 - Distance Perception in Mixed Reality

In our first egocentric distance perception study, we used verbal report to obtain absolute measures of people's distance perception judgments in two AR devices: the Microsoft HoloLens 2 and the Varjo XR-3 (Figure 6.1). We present targets at various distances within action space, and targets are either presented on or above the ground (Figure 7.1). Each target object was rendered either with or without a cast shadow. We anticipated that judgments of distance would be underestimated in both AR displays, but that there would be more underestimation in VST AR than in OST AR. Based on prior research on the influence of height in the visual field and on the influence of cast shadows, we anticipated that floating objects without a cast shadows would be perceived as farther away than targets on the ground when shadows were absent.

In total, we developed four hypotheses for the current experiment:

H1: Targets will be underestimated in both AR devices.

Table 7.1: Independent and dependent variables of our experiment

Independent Variables			
	observers	24	(random)
H1	distance	3	3, 4.5, 6
H2	shadow	2	yes, no
H3	height	2	0, 0.2m
H4	display	2	vst ar, ost ar
	repetition	3	1,2,3
Dependent Variables			
distance judgments (meters)			

H2: When a shadow is present, people's distance judgments will become more accurate.

H3: There will be an interaction between shadow and target height, such that there will be a difference between floating and grounded objects without shadows, but no difference when shadows are present.

H4: The video see-through display will induce more distance underestimation than the optical see-through display.

7.1.1 Participants

Twenty-four students and staff from Vanderbilt University were invited as volunteers for this experiment in exchange for 10 USD and 45 minutes of their time. The average age was 28.2 ± 8.75 years (Min: 21, Max: 68). Sixteen volunteers were male and nine were female. All participants experienced both the optical see-through AR display and video see-through AR display for a within-subjects experimental design. Our methods were approved by the local institutional review board, and written consent was obtained from volunteers prior to participation. All participants had normal or corrected-to-normal vision.

7.1.2 Design

To address our hypotheses, we utilized a 2 (display) × 2 (shadow shading) × 2 (target height) × 3 (target distance) within-subjects factorial design such that all conditions were presented to every participant. Distance judgments were obtained through verbal report.

The order in which the participants experienced the AR displays was counterbalanced such that half of the volunteers experienced the HoloLens 2 first and half of the volunteers experienced the Varjo XR-3 first. To reduce potential learning effects between display conditions, participants were moved to the opposite side of the room and rotated 180° before beginning the second part of the experiment with the other AR display. A participant standing on opposite sides of the room with each AR display can be seen in Figure 7.2.

We selected a sphere to be the virtual target (Figure 7.1). The virtual sphere measured 20 cm in diameter and was rendered with a middle gray RGB color value of 128. Participants viewed the sphere presented at three distances ($3m$, $4.5m$, and $6m$). Spheres were either placed on or above the ground at $0.2m$. A height of $0.2m$ was selected to draw comparisons between the current study and that con-

ducted by Salas-Rosales and colleagues [397], which also presented targets at 0.2m above the ground plane.

Prior research has shown that the angle at which a virtual light is positioned influences distance perception judgements [102, 135]. Therefore, we displayed spheres with or without a cast shadow, which was placed immediately beneath the object when it was present. This approach to rendering cast shadows is referred to as "drop shadow" in game development and in prior AR research [102]. For each display, this resulted in twelve unique combinations of stimuli.

Except for device, all other factors (i.e., shadow shading, target height, target distance) were pseudo-randomized so that a participant viewed each unique combination once before experiencing the same combination again. All unique combinations were repeated three times, which resulted in a total number of 36 trials per display. Each participant completed a grand total of 72 trials across both displays. With 24 subjects, a total of 1, 536 trials were collected overall.

7.1.3 Statistical Analysis

Before performing statistical analyses on our data, we converted all recorded distance estimates to meters. The distribution of participants' continuous response data was non-normal, which was exhibited by a positive skew in the data distribution. A Shapiro-Wilk test ($W = 0.943, p < 0.001$) and QQ plot inspection of the residuals from our model further revealed that the distribution of the residuals was not normally distributed. Fortunately, linear mixed models are robust to violations of distributional assumptions [408].

Because the population's responses for verbal reports of distance should have a Gaussian distribution, we assume that the underlying distribution of responses from the population is normal for our analyses [11]. Furthermore, to avoid overfitting our predictor values to the current data set, we do not fit our sample data to another distribution [12] nor do we transform our observed data to ensure the results of our analysis are interpretable [50]. Overfitting could compromise our ability to generalize the current results to other samples of the population.

We used a linear mixed-effects model (LMM) to investigate the influence of shadow, target height, and distance on people's distance judgments. Linear mixed models are a form of generalized linear regression that assume a normally distributed dependent variable. They are appropriate for repeated-measures designs because they allow for accounting of both within- and between-participant variability [386]. This is particularly important for examining verbal reports of distance estimates, which can be variable across individuals [248, 283, 327]. LMMs also permit model specification, so our analysis included only the interactions that were hypothesized a priori. This increased our power to detect differences.

Outside of factors related to our hypotheses, we also included experimental block order and the visual context of the room—from participants standing on either the left side of the room or the right side of the room to view targets—in our LMM to better understand how these counterbalanced, experimental factors would influence people's judgments. Both block order (2 levels: first, second) and visual context (2 levels: left, right), were treated as categorical factors. Further, we included interactions between device and experimental block as well as between device and visual context to ensure that these factors did not distort our results pertaining to device differences.

To account for individual variability in distance judgment behavior over repeated measures, we included a random intercept (μ_0). We then used Satterthwaite approximation via the *lmerTest* package

Figure 7.2: A participant views the experiment in the OST AR condition (left) and the VST AR condition (right). For each device, the experiment was conducted on opposite sides of the classroom.

[250] to calculate significance levels. The general regression equation is depicted in Equation 7.1 below:

$$\begin{aligned} \Upsilon = B_0 &+ B_1(\text{device}) + B_2(\text{shadow}) + B_3(\text{height}) \\ &+ B_4(\text{distance}) + B_5(\text{order}) + B_6(\text{context}) \\ &+ B_7(\text{shadow} \times \text{height}) + B_8(\text{device} \times \text{order}) \\ &+ B_9(\text{device} \times \text{context}) + \mu_0 \end{aligned} \tag{7.1}$$

We used the results of this analysis to answer our research questions: 1) whether people's distance judgments would be underestimated (i.e., **H1**); 2) whether peoples distance judgments would improve with the presence of shadows (i.e., **H2** the main effect of shadow); 3) interactions between target height and shadow (i.e., **H3**); and 4) whether distance misperception would be more severe in the Varjo XR-3 than the HoloLens 2 (i.e., **H4** the main effect of device).

7.1.4 Results

Participants' distance judgments were recorded and statistically analyzed in meters. However, in the following section we also report these values when converted into ratios to facilitate comparisons between the current work and prior research. To create ratios, participants' verbal distance estimates were divided by the actual distances to the target for a given trial. A ratio less than 1 indicates underestimation of distance, and a ratio greater than 1 indicates overestimation. Overall, distance judgments were somewhat variable with a mean distance estimate across participants of 3.722m ($SD = 1.566$, $Min = 0.914$, $Max = 10$). A mean estimate of 3.722m corresponds to a distance ratio of 0.827 or 17.3% underestimation.

The influence of experimental design: block order and environmental context

Although we counterbalanced our experimental factors, we nonetheless wanted to account for any variance in people's responses that was due to the order of experimental block experienced or due to the visual context provided by the room (i.e., whether the participant viewed the space from the left or right side of the room) by including them within our LMM. Based on prior depth perception studies, effects of experimental block order [132, 518] and visual context may be expected [132, 258, 464, 465].

Table 7.2: Mean egocentric distance judgments in meters for each device. Values in parentheses are standard errors.

	Each Distance			All Distances
Display	3	4.5	6	
All	2.43 (.03)	3.71 (.05)	5.03 (.06)	3.72 (.04)
HoloLens 2	2.47 (.04)	3.83 (.07)	5.24 (.09)	3.85 (.06)
Varjo XR-3	2.39 (.05)	3.78 (.06)	4.82 (.08)	3.60 (.05)

However, a significant interaction between device and either of these experimental design factors could convolute our planned analysis of device differences.

The analysis revealed a significant effect of block order ($B = 0.126, SE = 0.035, p < 0.001$) and a significant effect of visual context provided by viewing the room at two different locations ($B = 0.353, SE = 0.034, p < 0.001$). The effect of order indicates that people's responses were on average 0.126m farther in the second block of trials. The effect of environmental context indicates that people's responses were 0.353m farther given the environmental context provided by standing at the right side of the room (Figure 7.2, right) than the context provided by the left side of the room (Figure 7.2, left). Fortunately, our analysis did not show an interaction between device and order nor did it show an interaction between device and visual context.

Distance judgments will be underestimated (**H1**)

Participants underestimated distances to all targets by 17.6% on average ($M_{Ratio} = 0.824$, $SD = 0.176$, $Min = 0.305, Max = 2.54$). As shown in Figure 7.3, participants increased their egocentric distance judgments to virtual targets as the actual distance to the targets increased, supported by the significant main effect of distance ($B = 0.869$, $SE = 0.014$, $p < 0.001$). Participants estimated distances to be approximately 0.87m farther for every meter of increase in actual distance to the sphere, on average.

Distance judgments will be more accurate when a shadow is present (**H2**)

Our analysis supported this prediction. Distance judgments were significantly more accurate when spheres were rendered with a shadow when compared to judgments to spheres without a shadow. This was shown by a significant main effect of shadow ($B = 0.083$, $SE = 0.034$, $p < 0.05$), which indicated that participants estimated distances, on average, were 0.08m farther when a cast shadow was present. Overall, participants underestimated distances to targets rendered with shadows by 16.7% and they underestimated distances to targets without shadows by 18.6%.

There will be an interaction between shadow and target height (**H3**)

A priori, we anticipated that distance judgments to floating spheres would be similar to those positioned on the ground when shadows were present. Conversely, we predicted that distances to floating spheres would be judged as farther than those positioned on the ground when shadows were absent. However, our analysis showed no significant interaction between sphere height and the presence of a shadow.

Instead, our linear mixed model revealed main effects of both shadows (Section 7.1.4) and height ($B = 0.349$, $SE = 0.174$, $p < 0.05$) with no interaction between the two. When spheres were

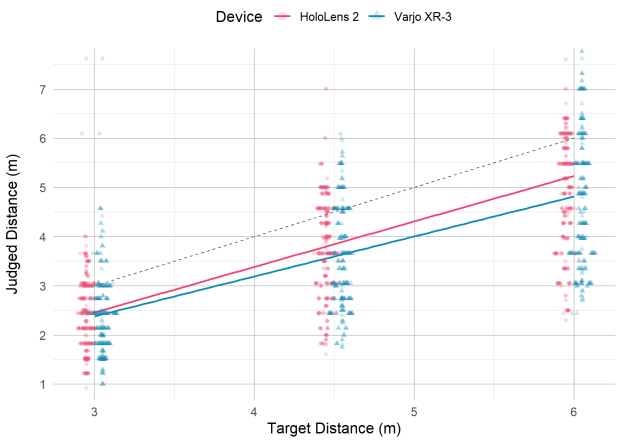

Figure 7.3: Judgments for the HoloLens 2 and Varjo XR-3 at each target distance fit with linear regression, featuring swarm plots of raw datapoints. The dotted line represents veridical performance.

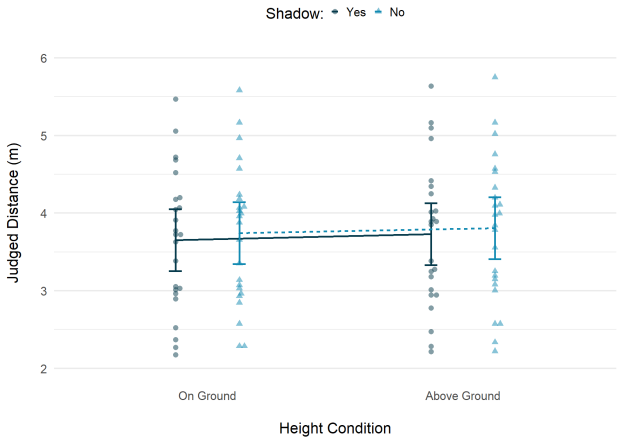

Figure 7.4: Estimated marginal means (EMMs) for Shadow & Height condition contrasts with raw EMM values plotted

floating above-ground they were perceived as farther away, with 16.72% underestimation for above-ground spheres and 18.5% underestimation for on-ground spheres. The differences in judgments based on height and on shadows are illustrated by Figure 7.4 Because we found main effects of shadow and height but no interaction between the two, we can conclude that target shadows influenced distance judgments, regardless of target height.

There will be more distance underestimation in the VST AR display than the OST AR display (**H4**)

As shown in Figure 7.3, egocentric distance judgments to virtual spheres were underestimated in both devices; distances were underestimated by 15.1% in the HoloLens 2 ($M_{Ratio} = 0.849$, $SD = 0.253$, $Min = 0.305$, $Max = 2.54$), and distances were underestimated by 20.2% in the Varjo XR-3 ($M_{Ratio} = 0.798$, $SD = 0.245$, $Min = 0.333$, $Max = 2.54$). Our statistical analysis indicated that this difference was significant. We found a main effect of device such that participants estimated distances to be 0.25m farther in the HoloLens 2 compared to the Varjo XR-3 ($B = 0.250$, $SE = 0.035$ $p < 0.001$). As such, distance underestimation was less severe in the optical see-through display (HoloLens 2) than the video see-through display (Varjo XR-3).

7.1.5 Discussion

The perception of scale through AR displays is an important problem that should be understood if AR is going to be successfully deployed in applications involving action that takes place over several meters. In this paper, our goal was to understand how AR displays affected the perception of scale, and to understand which characteristics of virtual objects affected that perception when those objects are seen in the context of the real world. In particular, we used the Microsoft HoloLens 2, an optical see-through display, and the Varjo XR-3, a video see-through display, to attempt to understand the perception of scale and characteristics of virtual objects. These two display types are both state-of-the-

art for their respective categories, so if we find characteristics that are similar for both displays, then we likely have a better understanding of the perceptual issues around AR more broadly.

First, distance judgments across both displays were underestimated, a finding that supports our first hypothesis (**H1**). We found 17.6% underestimation, on average. This result reinforces a growing body of literature evaluating egocentric distance perception in AR head-mounted displays that has found distance estimates in action space (2m - 30m) to be underestimated [109, 132, 222, 368, 397, 455, 478, 479]. However, participants were more accurate at estimating distances in the HoloLens 2 than in the Varjo XR-3, with an average of 15.1% underestimation in the HoloLens 2 and 20.2% underestimation in the Varjo XR-3, supporting hypothesis (**H4**).

This latter result confirms prior work done in both VR and AR. First, the Varjo XR-3 is heavier than the HoloLens 2 (980 g vs. 566 g), and weight of devices is a known factor in distance underestimation in VR [501]. Second, the field of view of the Varjo XR-3 is narrower than the field of view of the HoloLens 2, which allows nearly unobstructed viewing of the real world scene. Field of view is also a factor in distance underestimation in VR [64, 211, 212] and AR [368]. Finally, any misalignment of the cameras used in a video see-through display may cause disparities resulting in the depth of the scene being distorted [31]. Likewise, a magnification or minification of the scene seen through these cameras could cause misperception of depth [246]. The latter problems might also exist for graphically displayed objects in an optical see-through display but would not affect the real world objects seen through this type of display.

Additionally, we found a statistically significant effect of shadow presence or absence on distance judgments, confirming (**H2**). However, the improvement in distance perception was small, about 2%. This improvement is smaller than prior work in the real world and what our knowledge of graphics would predict [9, 102, 135, 175, 198, 492]. This finding is important regardless of the size of the effect because, insofar as we are aware, we are the first to use an absolute measure of distance perception in judging the effect of shadows on distance perception in AR (perceptual matching was used in prior work [102, 135, 175]). It is important to confirm effects through a variety of means and measures, and this result thus represents that. However, the small magnitude of the effect is surprising and a topic we return to below.

Finally, we did not find an interaction of height and shadow as we hypothesized (**H3**). The interaction of height and shadow is an effect we hypothesized based on prior work by Salas-Rosales et al. [397] in AR, and this interaction is one predicted by Gibson's ground theory [144]. We did find an effect of height on depth judgments, which is consistent with prior work both in the real world and AR [100, 255, 346], but the failure of rendered shadows to pin the location of objects down, i.e., the lack of a significant interaction, must be seen as a limitation of our experiment. This limitation should be noted since prior work has demonstrated that what people regard as a shadow in the real world is flexible [296, 333].

We conjecture two possible reasons for this lack of an interaction. As mentioned previously, the overall effect of shadow on distance underestimation was smaller than we expected *a priori*, particularly based on prior work [397]. It is possible that we did not have enough experimental power to detect an interaction given this. It may be that both these of these results are an outcome of using verbal reports. Verbal report measures are often more variable than other distance measures [13, 247, 508] and they can be susceptible to anchoring effects [353, 464, 465]. Thus, participants may not make verbal distinctions beyond the nearest 0.25 m (for example) or may repeat common responses. Secondly, the choice of a sphere may have made it harder for participants to judge the effect of shadows and ground contact than other shapes that are more commonly used in distance estimation studies in VR and AR, such as

a cube [397], traffic cone [224], or hockey puck [64]. Although the use of a sphere with shadows in distance estimation studies has significant antecedents [514], we have prior work that shows prediction of surface contact is harder for a sphere than other shapes (like a cube), and this may have also affected results [Anonymize citation]. Future work should explore these questions more thoroughly.

Another potential limitation of the current work is that we do not directly compare people's distance reports between real and augmented reality targets in the current study. As reported in Feldstein et al. [121], people's distance judgments in real spaces are typically accurate (M=92%), but the variance across these reports is considerable. One way to obtain more accurate comparisons of distance judgments between real and augmented reality targets is to evaluate people's judgments in both.

Although our current study provides a foundation for future investigations to evaluate how technical trade-offs between optical see-through and video see-through displays influence depth perception, it is impossible to infer what specific differences in these displays cause dissimilar degrees of underestimation with the current study alone. Part of the reason for this is that there are a large number of hardware and software differences between the two devices. Additional research is required to isolate influencing factors.

7.2 Study 6 - Distance Perception with Stylized Shadows

Early research in augmented reality suggested that virtual objected rendered by AR displays needed to match the appearance of their real world environment for accurate spatial perception [107]. For example, the shading of a virtual object should be visually realistic such that virtual lighting models should imitate lighting information found in the natural world. However, in recent years, augmented reality researchers have begun to argue against this theory [2, 5, 81, 441, 479]. This shift in thought developed in tandem with a rise in commercial AR applications—many of which rely on non-photorealistic (NPR) graphics to create engaging interactions between real and virtual spaces.

Niantic's Pokémon Go is an augmented reality game in which players encounter fantastical creatures in real world spaces. Google Maps AR presents viewers with floating arrows and informative icons showcasing local businesses to guide users to their destinations by using geospatial information. Even face overlays from popular platforms like Meta's Instagram and ByteDance's TikTok often rely on stylized overlays to exaggerate or create impossible 3D fixtures in people's physical space. A partial explanation for the ubiquity of stylized graphics in AR applications is that rendering photorealistic graphics is computationally expensive, and most current AR applications are deployed on mobile devices, which have a limited computational budget. However, it is uncertain how stylized graphics affect our visual perception of space when they are superimposed into reality.

In this dissertation, we begin to investigate this question by looking at a single graphical element: cast shadow. In our initial research studies, we provided evidence that a non-photorealistic rendering approach for cast shadows (in this case a white shadow) enhanced surface contact perception in MR. We then evaluated egocentric distance perception in two augmented reality devices to targets with and without cast shadows to understand how the absence of cast shadows and the presence of realistic cast shadows affected people's ability to accurately perceive depth in AR.

In our second distance perception study, we again use verbal report to obtain absolute measures of people's distance perception judgments in the Microsoft HoloLens 2 and the Varjo XR-3. We present targets at various distances within action space, and targets are positioned on and above the ground. However, in this second experiment we compare how distance judgements to objects rendered with

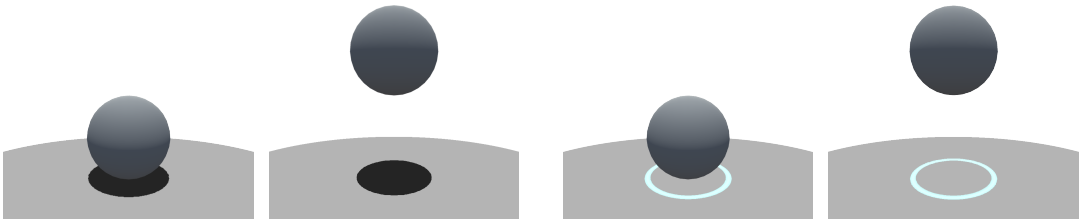

Figure 7.5: In Study 6, participants viewed target spheres that were positioned on and above the ground. These targets were either rendered with a cast shadow or with a glowing ring on the ground beneath them. The gray ground surface is provided here for illustration purposes; the actual ground was shown in the AR devices

a non-photorealistic surface contact cue compare to distance judgements to objects rendered with visually realistic cast shadows With this comparison, we seek to understand how the use of a non-photorealistic surface contact cue influence people's egocentric depth perception.

We predicted that depth judgements between the non-photorealistic surface contact cue and the realistic cast shadow would be similar. Furthermore, unlike our previous experiment, we anticipated that on and above ground judgements would be similar in the current study. By design, the effect of height observed in our prior study was largely driven by the no shadow condition, in which the position of objects in space was ambiguous due to a lack of surface contact information. Because both the on and above ground targets in the current study are presented with surface contact cues and because we anticipate similar performance between the two surface contact cue conditions, we predicted that judgements to objects of different height would be comparable in our second experiment. Because we evaluated the same two AR HMDs in our current study, we anticipated that distance judgements in both devices would be underestimated and that distances would be further underestimated in the video see-through condition.

In review, our hypotheses include:

H1: Distances will be underestimated, but there will be less underestimation (greater accuracy) with the HoloLens 2 compared to the Varjo XR-3.

H2: Distance estimations will increase as actual target distance increases.

H3: (a) Distance estimations will not differ between realistic versus stylized shadows, but (b) shadow type will have a greater influence in the OST vs VST device (shadow *times* device interaction).

H4: Distance estimations will not change with target height.

7.2.1 Participants

A different set of twenty-four students from Vanderbilt University participated in exchange for 10 USD and 45 minutes of their time. The average age was 27 years (Min: 21, Max: 59). Fifteen volunteers were female and nine were male. All participants experienced both the optical see-through AR display and video see-through AR display for a within-subjects experimental design. Our experimental methods were approved by the local institutional review board, and written consent was obtained from all volunteers prior to participation. All participants had normal or corrected to normal vision.

Figure 7.6: A participant views the experiment in the OST AR condition (left) and the VST AR condition (right). For each device, the experiment was conducted on opposite sides of the classroom.

7.2.2 Design

The experimental design of our second experiment is nearly identical to the design used in our first distance estimation experiment. We again utilized a 2 (display) × 2 (shadow shading) × 2 (target height) × 3 (target distance) within-subjects factorial design such that all conditions were presented to every participant. Distance judgments were again obtained through verbal report. In the distance estimation experiment, the shadow shading conditions referred to either a realistic cast shadow or a stylized graphical element along the ground. The four target height and 'shadow' shading conditions are shown in Figure 7.5.

The realistic shadow was the same as the cast shadow from the previous distance estimation study. The stylized graphical element, however, was rendered as a glowing white ring underneath the target object. This stands in contrast to the solid white shadow employed in our surface contact perception research (Section 4). We selected a glowing ring to replace the regular cast shadow, because this kind of graphical element is commonly used in AR applications as well as video games to indicate the position of objects of interest. The glowing ring also provides a more useful analogue to common interface elements used in AR actual applications.

The materials and procedure for our second experiment were the same as our first distance estimation study (See Section 6.2).

7.2.3 Statistical Analysis

Data were analyzed with mixed models, which are appropriate for the nested structure of the data in this experiment. Mixed models allow for the partitioning of variance both within and between participants. All analyses were performed in R [379]. Mixed models were run using the `lmer` function from the lme4 package [40], and the intraclass correlations (ICC) were calculated with the performance package [288].

Due to some of our hypotheses aligning with a null effect, we conducted our analyses in both a frequentist and Bayesian framework. Bayesian models were run using the `rstan`, and `brms` packages [66, 437]. For the Bayesian analyses, we report betas, credible intervals, and Bayes factors. Bayes factors were computed using bridge sampling and thus the reported values are approximations as a direct calculation is not possible within the context of multi-level Bayesian models. A credible interval is a probability statement that the true parameter would lie within an interval a certain percent (e.g.,

Table 7.3: Summary of outputs from our Frequentist and Bayesian statistical analyses

	Frequentist		Bayesian	
Predictors	*Estimates*	*Confidence (CI)*	*Estimates*	*Credible (CI)*
(Intercept)	2.24 ***	[1.56 2.91]	2.24	[1.53 2.96]
Device	- .44 ***	[-.64 -.24]	- .44	[-.64 -.23]
Shadow	- .01	[-.18 .16]	-.01	[-.18 .17]
Height	.01	[-.13 .15]	.01	[-.13 .15]
Distance (4.5m)	.58 ***	[.46 .70]	.58	[.46 .70]
Distance (6m)	1.14 ***	[1.01 1.26]	1.13	[1.01 1.26]
Order	.75 *	[-.07 1.57]	.76	[-.07 1.57]
Room	.44	[-.33 1.21]	.44	[-.39 1.24]
Shadow × Device	.05	[-.15 .25]	.05	[-.15 .25]
Shadow × Height	- .05	[-.25 .15]	- .05	[-.25 .15]
Device × Room	.10	[-.10 .31]	.10	[-.09 .30]
Device × Order	.24 **	[.03 .44]	.24	[.04 .44]

* p <.05 ** p <.01 *** p <.001

90%, 95%, 99%) of the time, given the observed data. For example, a 95% credible interval provides a probability statement that given the observed data, the true parameter would fall within the given interval 95% of the time. Bayes factors are mathematically defined by dividing the likelihood of the data under one model/hypothesis by the likelihood of the data under another model/hypothesis, such that a Bayes factor of exactly 1 suggests the data are equally likely under either model/hypothesis.

7.2.4 Results

Results from both frameworks are presented in Table 7.3. During posthoc assumption checking, we found that there was some slight heteroscedasticity when plotting residuals vs. fitted values. Thus, we re-ran the analysis using robust standard errors (using the `robustlmm` package in R) [236], while there were some slight changes in standard errors and p-values most changes were in the second or third decimal places and there were no changes in the direction or significance of effects that would change our interpretation. For brevity, only the original analyses are reported.

One note for interpretation, we used a null/uninformative prior in our Bayesian analysis, thus the Bayes factors presented should be interpreted with caution because Bayes factors are highly sensitive to prior specification [8, 138, 158, 277]. Different priors could lead to a dramatic shift in the Bayes factor (see Liu and Aitkin [277], for an example). Accordingly, our interpretation in the Bayesian analysis leans more on the information provided by the credible interval, as the posterior distribution summarized by the credible interval is notably more stable and less sensitive to the prior, especially as sample size increases.

The primary dependent variable was distance estimates (converted to meters). The intraclass correlation (ICC) for distance estimates was .44, indicating that 44% of the variance in distance estimates was between participants and 56% was within participants. Regardless of condition, on average participants in this experiment underestimated distance by ~27%. All variables were dummy coded in our

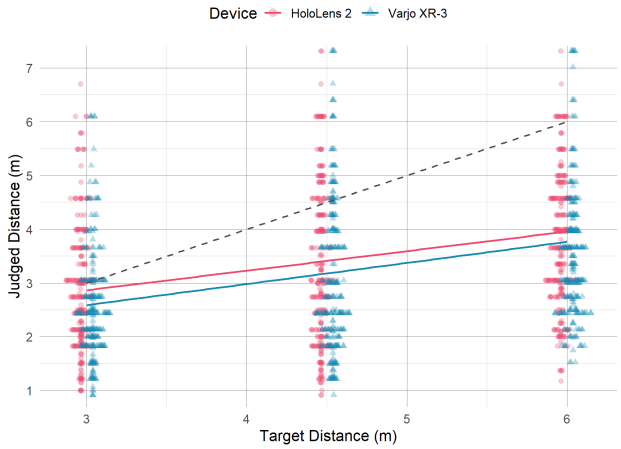

Figure 7.7: Judgments for the HoloLens 2 and Varjo XR-3 at each target distance fit with linear regression, featuring swarm plots of raw datapoints. The dotted line represents veridical performance.

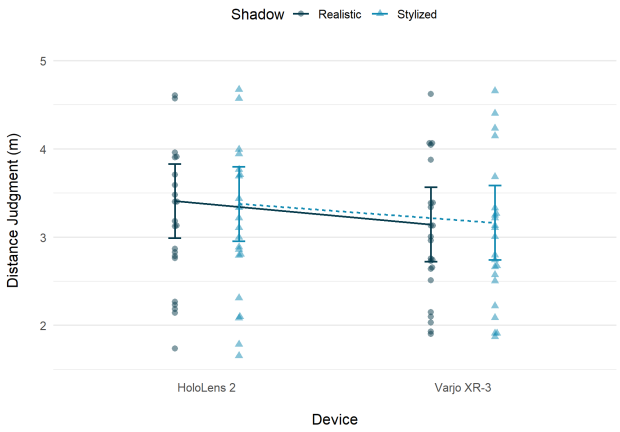

Figure 7.8: Estimated marginal means (EMMs) for Device & Shadow condition contrasts with raw EMM values plotted

analyses, which means the intercept represents the average distance estimate when all other variables are set to 0 (i.e., Device=HoloLens, Shadow=realistic shadow, Height=on ground, Distance=3m, Order=HoloLens first, Room=HoloLens on left side of room) and thus all betas (B) are relative to this intercept value.

H1: Effect of device

On average, participants underestimated distances by ~24% in the HoloLens 2 and ~29% in the Varjo XR-3. Hypothesis 1 was supported (via the frequentist analysis) in that distance estimates in the Varjo XR-3 were shorter ($B = 0.44, SE = .10, p < .001$), compared to distance estimates in the HoloLens 2 (See Figure 7.7). This means that across participants, distance estimates in the Varjo XR-3 were 44cm shorter. The Bayesian analysis further confirmed Hypothesis 1 with an identical beta/slope value of $-.44$ and a 95% credible interval that ranged from .64 to .23. While credible intervals are not designed for hypothesis testing [45], they do provide intuitive and interpretable estimates of uncertainty. Thus a 95% credible interval ranging from .64 to .23 suggests that the range of plausible values for the effect of device span from 23cm to 64cm of underestimation in the Varjo XR-3 compared to the HoloLens 2. The Bayes factor when comparing the full model to a model with device and all device interaction terms removed was > 100 $(2, 526.26)$, suggesting the data is more than 100 times more likely under the alternative hypothesis than the null hypothesis. However, the effect of device is qualified by a significant device x order interaction ($B = .24, SE = .10, p = .02$). This suggests there was a larger effect of device ($B = .44$) when participants ran through the experiment in the HoloLens 2 first and the Varjo XR-3 second and smaller effect of device ($B = .20$) when participants ran through the experiment in the Varjo XR-3 first and the HoloLens 2 second.

H2: Effect of distance

Hypothesis 2 was supported by both the frequentist and Bayesian analyses. In the frequentist analysis, both the 4.5 ($B = 0.58, SE = 0.06, p < 0.001$) and 6 meter ($B = 1.14, SE = 0.06, p < .001$) distances were estimated as farther than the 3m distance, as seen in Figure 7.7. This means that on average, at 3m (represented by the intercept in our models) participants estimated the sphere to be 2.24m away. At 4.5m estimates increased by .58m, meaning that on average at 4.5m participants estimated the sphere to be 2.82m away. While at 6m estimates increased by 1.14m relative to the intercept (3m), meaning that on average at 6m, participants estimated the sphere to be 3.37m away. The Bayesian analysis further confirmed Hypothesis 2 with identical beta/slope values of .58 and 1.14 and 95% credible intervals that ranged from .46 to .70 and 1.01 to 1.26. The range of plausible values range from 46cm to 70cm and 101cm to 126cm—meaning there is almost certainly a positive relationship between target distance and estimated distance. The Bayes factor for effect of device was > 100 (1.27×1029), suggesting the data is more than 100 times more likely under the alternative hypothesis than the null hypothesis.

H3: Effect of shadows across devices

Hypothesis 3 was partially supported by our frequentist and Bayesian analyses. In the frequentist analysis, our prediction was that there was no effect of shadow (H3.a) ($B = .01, SE = .09, p = .92$), but our prediction failed since the shadow x device interaction term was non-significant ($B = .05, SE = .10, p = .52$) (H3.b). The Bayesian analysis resulted in an identical beta/slope value of .01 and .05 and 95% credible intervals that ranged from 0.18 to 0.16 and .15 to .25, respectively. Given 0 is near the center of both 95% credible intervals above, we interpret an effect of 0 as plausible for the effect of shadow and the shadow x device interaction. The estimated marginal means for this interaction are visualized in Figure 7.8.

The Bayes factor when comparing the full model to a model shadow and all shadow interaction terms removed was ~0.01, suggesting the data is approximately 99 times more likely under the null hypothesis than the alternative hypothesis. The Bayes factor for the shadow x device interaction effect was ~0.13, suggesting the data is approximately 8 times more likely under the null hypothesis compared to the alternative hypothesis.

The "~" indicates the potential slight variability of this estimate, given that bridge sampling was used to calculate Bayes factor. A direct calculation is not possible within the context of multi-level models.

H4: Effect of height

Hypothesis 4 was supported by both the frequentist and Bayesian analyses. In the frequentist analysis, there was not a significant effect of height ($B = .01, SE = .03, p = .88$). The Bayesian analysis resulted in an identical beta/slope value of .01 and a 95% credible interval that ranged from .13 to .15. Given 0 is near the center of the 95% credible interval and the interval is fairly symmetrical, we interpret an effect of 0 as highly plausible.

Even if the effect is non-zero, it is likely small, as the tails of the 95% credible interval lie around $\pm .15$. Even at the extremes of plausible values (.13, .15) this would translate to spheres positioned .2 meters off the ground being either under or overestimated by $13 - 15$cm, with an effect close to 0(.01) being most likely. The Bayes factor for the effect of height was ~.02, suggesting the data is approximately 62 times more likely under the null hypothesis compared to the alternative hypothesis.

7.2.5 Discussion

The current experiment tested whether absolute egocentric distance judgments would be influenced by the type of shadow and the type of AR device used. Previous work had suggested that there would be overall underestimation of distance to AR targets, but how this underestimation is influenced by different shadow rendering techniques in mixed reality devices that use different methods for presenting AR graphics (OST vs. VST) was unknown. In a completely within-subject design, observers verbally reported perceived distance to spherical targets $3 - 6$ meters away, on or above the ground, with realistic and non-realistic shadows, using two different AR devices. Our results mostly supported our hypotheses. **H1:** We found underestimation of distance in both devices but less underestimation with the HoloLens 2 (24%) than with the Varjo XR-3 (29%). **H2:** We found the expected effect of distance–as target distance increased, verbal estimates increased. **H3:** We found no difference due to shadow type, and this did not interact with the device type, suggesting that the stylized shadow was as effective as the realistically rendered shadow. **H4:** We found that estimations did not change with location on or above the ground, also providing support for the use of both shadows to specify location relative to the ground plane. We discuss each of these effects (or lack of) further in the context of prior work and potential implications.

Distances were underestimated, as expected, in AR, but participants were more accurate with the HoloLens 2 (OST) compared to the Varjo XR-3 (VST) displays. This finding replicated of our previous experiment (Section 7.1 [4]), and is an important result given there are few other studies on distance perception using the relatively new Varjo displays. There are several factors that could possibly contribute to the differences found between the devices in this work. One notable factor is the camera-based systems used in VST displays. In the Varjo XR-3, misalignment of the cameras with the eyes could distort depth in the scene [77, 174, 458, 507]. Another factor could be the different FOVs; a smaller FOV is associated with greater distance underestimation in VR [64, 214, 303, 304] and AR [133, 368]. Although the HoloLens 2 has a smaller FOV for the presentation of virtual objects, the viewer sees their full real-world FOV through the OST display, which is larger than the FOV seen through the Varjo XR-3. The weight of the Varjo XR-3 is also greater than that of the HoloLens 2, a factor shown to influence distance underestimation (see Kelly [223] for a recent review).

We found no effect of shadow type on distance estimations, across the two devices. Unlike in our previous distance estimation experiment, we included shadows on all trials, but varied whether they appeared realistic or stylized as shown in Figure 7.5. We predicted that the stylized shadow would serve to provide information for ground contact equally well as the realistic shadow, supporting the findings of our surface contact perception research (Section III) [2, 5]. We also considered the possibility that the shadows would have different effects given the two different display types (i.e., particularly the OST might benefit more from a lighter colored shadow), but we found no evidence for an overall difference or an interaction with device. Why might the stylized, non-realistic shadow be so effective? As suggested by our previous work, it could be that the bright ring provided such a salient cue for ground contact that its benefit outweighed any costs of appearing unnatural. The ring used in the current study differs from the solid white shadow used in our previous experiments yet it still had similar effects. Future work should examine whether other creative forms of nonrealistic shadows [209, 221] would also match performance of realistic shadows in distance perception tasks.

Consistent with Gibson's ground theory![144], prior work has shown that AR objects are perceived as farther away when they are located off the ground, given no additional information for ground contact [4, 397]. However, when shadows are provided as a ground contact cue, we would expect this

effect to be reduced. Thus, our current finding that there was no effect of height above the ground on perceived distance with both types of shadows provides further support that the stylized shadow serves as an effective cue for location, similar to the realistic shadow. We tested only one relatively small (20cm) vertical displacement above the ground. Future work could assess the utility of different types of shadows at different heights and distances to further generalize these effects.

This study used verbal reports as the response measure for distance. Distance underestimation is consistently seen in the real world when using verbal reports [283]. Because we did not run a matched real world condition, we cannot make strong claims about whether the current AR results show greater or comparable amounts of underestimation to the real world. However, it is notable that the magnitude of distance underestimation increased with increasing distance in our data (~20% at 3m, ~37% at 4.5m, and ~44% at 6m), a result that is not typically seen in the real world at this range of distances [283, 370]. Studies that have directly compared AR and real world estimates show mixed results, which is likely due to multiple factors including the type of measure and the type of AR display. Action measures, such as blind walking [281, 392] are typically accurate in the real world and some studies have shown comparable distance estimation with real and AR targets (e.g., Stefanucci et al. [438]). There has been significant work in virtual reality studying distance estimation and response measures as well (see Creem-Regehr et al. [93] for a recent review), but less work has been done in AR (however, see Jamiy et al.![109] for a video see-through virtual reality comparison). It would be interesting to compare these AR devices in a study with action measures to see if it leads to an improved understanding of the differences between OST and VST devices.

7.3 Summary

In Aim 2 the dissertation, we examined the impact of augmented reality display on depth perception with two major goals in mind. First, we wanted to compare depth perception in optical see-through (OST) and video see-through (VST) displays to better understand their perceptual limitations and the trade-offs between them. Second, we wanted to investigate the influence of realistic and unrealistic graphics on people's depth perception, especially as applied to cast shadows and other visual cues that connect virtual objects to real surfaces in space.

We conducted two experiments in which verbal reports were employed to evaluate absolute distance perception in two commercially available HMDs: the Microsoft HoloLens 2 (an OST AR HMD) and the Varjo XR-3 (a VST AR HMD). Consistent with prior research, people exhibited significant distance compression in both devices, although people experienced less severe distance compression in the HoloLens 2. Our studies were the first to evaluate absolute measures of distance perception in either the HoloLens 2 or the Varjo XR-3. They are also some of the few research studies that have compared depth perception between OST AR and VST AR HMDs.

Our research furthers our understanding of the influence of virtual object height on distance perception judgments in AR, it extends our prior research on the influence of cast shadow on spatial perception by connecting surface contact perception research to distance perception research, and it provides evidence that stylized graphics do not detract from accurate spatial perception in AR.

Beyond the limitations discussed previously, future research should consider generalizing the findings across various AR displays and employing diverse methods (e.g., blind walking, perceptual matching) to evaluate depth perception. Additionally, given the influence of height in the visual field found in the current work, it may be beneficial to conduct a more in-depth investigation of the influence of

target height on depth perception in AR. Finally, it may also be interesting to look at how the unique rendering properties of the two displays influence depth perception, given that OST and VST AR devices integrate virtual and real environment information in unique ways.

Acknowledgment I would like to thank Mirinda Whitaker for taking the time to discuss Bayesian statistics with me and for her tremendous help with the statistical analysis for this experiment.

Part V

Vision Simulation & Eye Tracking

What we cannot imagine cannot come into being
bell hooks

8 Vision Simulation

In the final aim of this dissertation (Aim 3), we extend our investigation into spatial perception by evaluating distance judgements made by people with simulated vision loss in MR. People with impaired vision report difficulties performing a range of daily tasks that require an accurate encoding of 3D space [382, 404]. For example, people who experience severe vision loss to their peripheral or central vision express difficulties performing tasks like driving, navigating stairs, and avoiding obstacles while walking [116, 181]. In traditional two-dimensional (2D) media, it has been demonstrated that people with vision loss benefit from nonrealistic image enhancements like edge and contrast enhancement, since these methods improve the visibility of items on a screen [70]. However, it is unknown if stylized graphics and image effects will aid or impede people with visual impairments as they navigate complex 3D environments [403]. Vision enhancements that interfere with a patient's residual vision or interfere with their ability to accurately perceive the distance to objects in space may exacerbate preexisting mobility difficulties.

To facilitate our study on the influence of visual field loss on spatial perception and accessibility, we create an immersive low vision simulation. This simulation provides control over the extent of visual field loss experienced by the user. The control provided by our simulation is useful for better understanding the influence of specific vision loss symptoms, in our case visual field loss, on people's perception and behavior. As such, we use this simulation to study the influence of visual field loss on spatial perception in general and to better understand the influence of realistic versus stylized graphics on spatial perception judgements.

In this chapter, we describe our vision impairment simulation. Our implementation pays special attention to the reproduction of symptoms associated with visual field loss and reduced visual acuity. Both symptoms are inversely correlated with vision-related quality of life in patients with low vision [23, 181, 350, 378]. Further, central vision loss and peripheral vision loss have important, differential consequences for spatial perception [231, 356, 431, 482]. To create more accurate visualizations, we also draw inspiration from the most common descriptors used by low vision patients to describe their experience with vision loss [87, 124, 194, 462]. Before introducing our vision simulation, we briefly describe the characteristics of low vision and vision impairment in Section 3.1. In Section 8.2, we then explain the image processing pipeline used for our simulation. We also provide context for each feature of the pipeline by discussing how our implementation differs from that found in prior vision simulation work. Finally, we discuss at a higher level how our simulation compares with previous work, some limitations of our current approach, and how future iterations of our system may include other symptoms of vision impairment.

8.1 Motivation & Goals

Low vision is not the same as having poor visual acuity —- nor is it the same as having a complete lack of light perception (*total blindness*). The functional definition of low vision defines low vision as a visual impairment that cannot be corrected by medical treatment or conventional eyeglasses [328]. Typically, individuals with low vision experience partial vision loss within their visual field. As vision impairments become more severe over time, these localized areas of vision loss, or *scotomas*, can result in a complete loss in light sensitivity and reduced visual acuity can be experienced in larger areas of the visual field. Common eye diseases associated with low vision include: glaucoma, age-related macular degeneration, and diabetic retinopathy. The symptoms that individuals experience in low vision are highly heterogeneous, since the area and the extent of the visual field affected can vary. Multiple areas of the visual field can also be affected.

This heterogeneity makes classifying and studying how low vision affects visual function, visual perception, and behavior challenging. As a result, traditional classification schemas for estimating visual function in people with low vision can be vague or inaccurate. For example, population-based surveys often exclusively rely on visual acuity to categorize the severity of vision impairments [201]. However, ophthalmologists often discourage the use of visual acuity tests alone to assess visual function in people with low vision [84, 207, 262] as visual acuity has poor diagnostic accuracy for low vision [80, 201]. Since the 10th revision of the *International Classifications of Diseases (ICD)*, an internationally used diagnostic tool for categorizing disease, the importance of considering other characteristics of vision loss has been formally codified [97, 201, 357]. The ICD-10 puts a particularly strong emphasis on visual field extent. For example, they classify low vision as having a visual acuity of less than 6/18 but equal to or better than 3/60 or a corresponding visual field loss to less than 20 degrees in a person's better eye (ICD-10: H54.9) [201].

How to best render visual field loss and scotomas for simulation is an active area of research [364]. Even though immersive low vision simulations exist, most are based on simplified symptoms of eye diseases [512] and are unable to produce the irregular scotomas that individuals experience in reality.

8.2 Our Vision Simulation

In this section we describe the image processing pipeline for our proposed gaze contingent simulation of visual field loss. For each eye, the algorithm takes the monocular, digital image produced by a display and a 2D matrix that contains visual field loss information (the visual field map) as inputs. It then performs the following operations: (1) it matches the visual field map (VFM) to the screen resolution and field of view (FoV) of the display, (2) it positions the VFM texture relative to the center position of a person's gaze, and (3) it performs image processing operations to simulate the symptoms of visual field loss at pixels on the display's screen where vision loss is indicated by the VFM. For rendering symptoms of vision loss, we design two methods that may be used together or in isolation to simulate different eye disease symptoms: (3a) rendering vision loss with opacity and (3b) rendering vision loss with blur. Figure 8.1 displays a schematic of our image processing pipeline for clarity.

The visual field map input is a 2D matrix that indicates where vision loss is present as well as the severity of vision loss at specific areas within the visual field. Our simulation is designed to be data agnostic. As such, it can accept arbitrary 2D matrices (e.g., 2D images) as VFM inputs. However—in the interest of creating realistic reproductions of visual field loss—in the current work we also introduce a

data-driven solution for generating VFMs from clinical perimetry data (Section 8.2.1). The patient data we use to generate VFMs is from a specific perimetry test: the Humphrey Visual Field test. Integrating VFMs that are generated from different perimetry tests in future research is a straightforward task since the simulation itself accepts arbitrary VFMs as inputs.

We provide an explicit port of our simulation for the HTC Vive Pro Eye. We planned to include a port for the Varjo XR-3, as well. Unfortunately, complex post-processing operations, like those employed in our simulation, are prohibited by the Varjo SDK*.

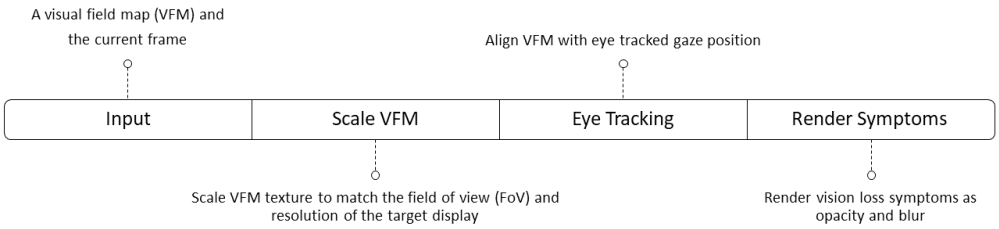

Figure 8.1: Schematic of the vision simulation pipeline

8.2.1 Visual field map

Our simulation uses visual field maps to determine where and how image processing techniques that simulate vision loss should be applied within the field of view of a head-mounted display. These visual field maps are 2D matrices in which each value corresponds to linear scaled retinal sensitivity values in the range [0,1]. Within these maps, black (0) indicates a complete loss of retinal sensitivity. This approach allows our simulation to accept arbitrary visual field maps as inputs, since any 2D matrix may be used to map where vision loss occurs. For example, VFMs can be generated as single channel grayscale textures in which each pixel corresponds to a retinal sensitivity value. Our simulation permits different VFMs to be used as inputs for the left and right eyes.

The ability to accept arbitrary visual field maps as inputs and the ability to specify different maps for each eye are desirable in vision simulations for a number of reasons. The appearance of scotomas is highly heterogeneous, such that the shape, the size, and the number of scotomas vary across individuals—even when they have similar medical diagnoses [86, 196, 462]. An individual may also have different severities of vision loss between their two eyes [16, 61]. As such, being able to capture these characteristics is important to ensure more realistic visualizations of vision loss.

Even more realistic representations of diseased visual fields can be created by using patient data captured through visual field tests (*perimetry tests*) as inputs. By using patient data, we may better replicate the extent of a particular individual's vision loss as well as the shapes and sizes of any scotomas in their vision. Because our simulation is data agnostic—it can process data across a wide range of different visual field test data formats. However, one should always consider how data is collected for a given perimetry test before naively interpolating values to generate a visual field map. In the following subsections, we describe a principled approach for converting patient data into visual field maps for our system.

*Varjo Technologies Oy (2024). https://developer.varjo.com/docs/get-started/Post-processing

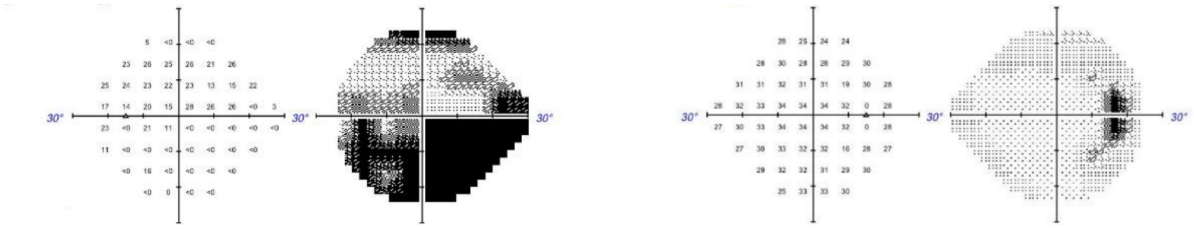

Figure 8.2: Visual field examination results using the Humphrey field analyzer (HFA) [72] 24-2 SITA-Fast protocol for a patient with glaucoma. The patient has visual loss indicative of advanced glaucoma in their left eye (left) and nonspecific changes in their right eye (right) [393]

Visual field data

In clinical settings, perimetry tests are useful for diagnosing a range of medical conditions (e.g., retinal diseases, neurological diseases, glaucoma). These assessments detect vision loss within the visual field, and they provide a map of where scotomas occur. In the context of vision simulation, the resulting matrices of perimetry data can be useful for determining where and how image processing effects should be applied within a display to mimic patient symptoms, since perimetry data captures both where vision loss occurs as well as the severity of vision loss at a given location within the visual field [302, 367, 445].

A number of different types of perimetry tests are used in medical practice, including Humphrey Field Analyzer (HFA), Goldmann, and Octopus perimetry. All perimetry tests produce coordinates of data in which each datapoint corresponds to a person's sensitivity to visible light (or retinal sensitivity) at a specific region within their visual field—although the way datapoints are measured, the number of data points, and the extent of the visual field tested vary. For example, automated static HFA perimetry produces data as a discrete grid of logarithmic sensitivities and manual kinetic Goldmann perimetry produces data as a circular isopter on a polar coordinate grid [391]. Accordingly, how patient visual field data should be processed to generate visual field maps is dependent on the specific perimetry used to measure vision loss. To illustrate how patient perimetry data can be converted into a visual field map in practice, we look at a specific example: Humphrey field analyzer.

Generate a visual field map from Humphrey Field Analyzer

Humphrey Field Analyzer data is useful for generating data-driven visual field maps in vision simulations on account of its common use in clinical practice. Further, Montesano et al. [320] have published an invaluable dataset of visual field data collected from 3,895 individuals using the HFA 24-2 protocol that we may exploit to visualize a range of different visual field symptoms. The HFA is conducted by presenting lights with different levels of brightness to a patient while they fixate their eyes at a target at the center of a display. The patient is asked to respond each time they detect the presence of a light in their visual field. The test is conducted for each eye, individually. The HFA allows for different testing protocols, in which the number of points tested and the extent of the field of view measured vary [338]. An example of the data produced by the 24-2 protocol, which measures 54 points that are measured 24 degrees temporally and 30 degrees nasally, is shown in Figure 8.2.

Retinal sensitivity is measured in reference to luminance intensity. The HFA, like most perimetry tests, does not shine light directly into a patient's eyes, but rather projects light onto a reflecting surface. The light that reaches the patient's eyes is thus measured by candela per square meter (cd/m^2), which

is the luminous intensity over a projected surface area [410]. In perimetry tests, luminance intensity is also often reported in apostilb (asb), where 1 asb = 0.3183 (cd/m^2).

There is an inverse relationship between the luminance of the stimuli in the test and a patient's light sensitivity—such that high sensitivity is shown by responses to lights with low brightness and low sensitivity is shown by responses to lights with high brightness.

Retinal sensitivity is reported as the negative log of the threshold luminance, and it is given the arbitrary units decibels (db) [39] (*Appendix, p.319-323*). A retinal sensitivity change of +1 db equals a threshold change of –0.1 log (asb). Logarithmic scaling follows from Weber's law, which states that perceived brightness is proportional to the logarithm of the actual intensity measured [119, 472]. Light sensitivity values near 30 dB indicate normal light sensitivity and values near 0 dB indicate little to no light sensitivity [59, 79, 218, 505]. Individual points with a value of 40 dB or greater are considered hypersensitive results caused by patient overreaction or guessing during the exam [73, 172]. Because our VFM is designed to simulated impaired vision, we clip all values greater than 30 dB, since areas of normal vision should not be rendered with vision loss symptoms.

To ensure that our visual field map can be utilized with a wide range of image processing operations, we normalize sensitivity values and convert them to a linear scale in the range [0,1]. Converting retinal sensitivity values to a linear scale is important for ensuring that image processing methods function correctly since most of these operations (e.g., resizing, filtering) assume linear inputs. Performing linear operations on nonlinear inputs can introduce undesirable image artifacts [155]. For example, lighting and resizing operations performed in nonlinear space (e.g., gamma-corrected sRGB images) can cause shadows and dark color values to become abnormally dark. Converted values as stored as floating point numbers instead of 8-bit numbers, which are common in RGB images and other types of texture maps, to ensure that a wide range of retinal sensitivity values are represented. The human eye is capable of adapting to a dynamic range of luminance intensities beyond the range provided by 8 bit quantization [122].

8.2.2 Match to HMD screen and FOV

To ensure that visual field maps are rendered correctly within the HMD, they are upsampled to match the per eye resolution and monocular field of view of the HMD. Specifically, our solution uses the per eye (monocular) resolution and the FoV of the display as inputs to ensure that visual field loss is rendered correctly. As such, our vision simulation may be applied to any eye-tracked head-mounted display, where the resolution and FoV of the display are measurable.

By default, our simulation assumes that the VFM corresponds to the full monocular FoV of the HMD. This behavior can be overridden by specifying a smaller horizontal and vertical FoV for the VFM. When FoV information is provided, the simulation upsamples the VFM to match the pixel count of the display that corresponds only to that portion of the display's FoV.

This functionality is especially important for VFMs that are generated from patient data, because the visual field extent tested in perimetry protocols is often smaller than the extent viewable in contemporary VR HMDs. For example, the HFA 24-2 protocol captures 54° visual field horizontally, but contemporary VR HMDs provide approximately 100° horizontally. In the current version of the simulation, upsampling is performed using hardware accelerated bilinear interpolation [484]. However, the final version of the simulation will employ Gaussian upsampling to better mimic the human visual system [139, 515]. When the field of view of the VFM is smaller than the field of view of the display, the last pixel at the border of the visual field map is extended outward to complete the periphery us-

ing bilinear interpolation. This approach allows for flexible simulation of both central and peripheral scotomas.

For the current work, we estimate field of view for an HMD using the field of view that is rendered to its display panel. These values are obtained using the OpenVR API [474] and the HMDQ command line tool [324]. We use this approach, which queries FoV data from the headset driver rather than the manufacturer's estimates, since manufacturers often do not report the FoV data we require for this simulation (e.g., monocular, horizontal FoV) and since manufacturers typically do not report how FoV is measured in their assessments. Leveraging the OpenVR API also ensures that our FoV calculations and any resulting adjustments to visual field maps for our simulation are reproducible across HMDs of the same make and model.

8.2.3 Gaze contingent rendering

Gaze contingent rendering is achieved in our simulation by positioning the visual field map texture relative to the center of a person's gaze. All image processing methods for simulating symptoms of vision loss use the position shifted visual field map as an input. Eye tracking integration is imperative for the simulation of low vision, since vision impairments for these conditions often affect specific areas of a person's visual field. For example, vision loss from glaucoma is most prominent in peripheral vision and vision loss from age-related macular degeneration (AMD) is most prominent in central vision. To simulate these vision conditions correctly, a virtual scotoma must be rendered relative to where a person is looking——not as a static image on a display.

8.2.4 Rendering vision loss symptoms

A variety of image processing techniques may be employed with our visual field map to create different vision loss symptoms. In the current work, we prioritize visual field loss simulation. And we build two unique screen shaders that can be used to render vision loss symptoms. The first method renders vision loss as an opacity, and the second method renders vision loss by applying image blur.

The two shaders work in concert with the visual field map so that additional image processing is performed only when vision loss at a given pixel is indicated by the VFM. The vision loss shaders may be employed individually or in combination—depending on the symptoms that one wishes to simulate. The final image rendered by the display is the result of alpha blending operations at each pixel between the original image and images produced by the low vision shaders. Blending weights for each pixel are informed by the retinal sensitivity values stored within the VFM, where a value of 0 (which corresponds to no vision impairment) indicates that only the unaltered image color value should be used for rendering.

Rendering vision loss with opacity

Rendering vision loss with opacities is the most traditional method for simulating visual field loss [30, 152, 481]. This approach is also what the National Eye Institute (NEI) and National Institute of Health (NIH) used to generate their low vision example visualizations (Figure 8.3). Although most low vision patients do not describe their vision loss symptoms as dark holes or tunnels in vision, there are several vision conditions in which vision loss can be perceived as an opacity or dark shadow by the patient, including: retinitis pigmentosa, retinal detachment, and certain forms of AMD [462, 503]. It

Figure 8.3: Images provided by the National Eye Institute Media Library that are commonly used to demonstrate symptoms of different eye diseases: (left) age-related macular degeneration, (center) glaucoma, (right) diabetic retinopathy [329].

is likely that simulating visual field loss with opacity is such a common approach—despite its lack of realism for most low vision conditions—because visualizations created with this approach are easy to interpret for the naive viewer and because they are easy to generate.

Accordingly, a number of researchers have implemented methods for rendering vision loss as opacities for realtime vision simulation [21, 30, 243, 477, 496, 512]. Most methods involve similar post-processing effects. For example, Velazquez et al. [481] simulated central vision loss using binary masks to produce sharp onsets of vision loss, which is an approach that resembles the dense scotomas exhibited in severe cases of age-related macular degeneration (AMD) [462]. Krösl et al. [243] developed a method to simulate the dark shadows in vision described by some patients with cataracts [316] as well as retinitis pigmentosa [503]. To achieve this effect, they linearly interpolate between a source image and a dark mask with an alpha channel to create variable levels of opacity in the final image.

For our first vision loss shader, we present a solution similar to what has been developed in prior research. For each image rendered in the display, we apply a post-processing effect that renders opacities in a person's vision, where vision loss is indicated by the visual field map. By default, opacities in our simulation are rendered as black. At each pixel, linear interpolation is performed between the pixel value of the original image and the color of the opacity. The weight of this interpolation is decided by the visual field map. When the VFM indicates no vision loss, the color value of the original image pixel is used. When complete vision loss has occurred, the pixel is completely replaced by the color of the opacity. All other values within the VFM result in blending between the original image and the opacity color.

Rendering vision loss with blur

Traditional visualizations that render scotomas as dark blind spots in vision, like the images used by the NEI and NIH for AMD Figure 8.3 (left) and glaucoma Figure 8.3 (center) are not accurate for most low vision patients. For patients with glaucoma, Crabb et al. [87] found that vision loss descriptors were most often reported as blur and missing features (as opposed to black tunnel effects or black patches). Fletcher et al. [124] reported similar findings in patients with AMD. When scotomas were detected, they were registered as experiences in which items in the environment "disappeared"—rather than as large opacities in vision. Similarly, a number of other low vision patient studies have found that the most common descriptors for vision loss symptoms used by patients are "blur" and "missing" (or "disappearing") information [87, 124, 194, 462]. Disappearing effects often occur when there is complete loss of vision (blind spots) within the visual field. This phenomenon can be explained by the

Figure 8.4: Examples of simulated perceptual filling with extreme blur: (left) unaltered image of table with candle visible, (right) a central scotoma completely obscures the candle.

perceptual "filling in" effects in the visual system [99, 237, 520]. Rather than ignoring the absence of information when blind spots occur, the human visual system constructs a complete representation of the affected area by blending proximal background information [237, 520]. Perceptual filling in can even occur in the presence of artificial scotomas [381], although it is more robust in the presence of real scotomas [520]. The underlying processes that permit active filling in are not yet well understood [99, 237, 520].

For our second vision loss shader, we propose a method to simulate visual field loss using these two most frequently described symptoms: blur and missing information. We render both blur and perceptual filling in effects with a single shader that introduces variable amounts of blur to an image. This approach is desirable since extreme instances of blur may be used to approximate the blending effects described by the perceptual filling phenomenon [380, 520] and since filling in effects are often co-morbid with visual acuity loss [134, 215]. Perceptual filling is approximated in our proposed shader by introducing extreme blur to regions of the visual field with complete vision loss. This blur is severe enough to obscure all visual detail, such that only color and luminance information from nearby background information remains (e.g., Figure 8.4).

Generating Blur Although different smoothing methods exist, the most commonly employed for simulating reduced visual acuity through blur is the Gaussian filter. In 2006, Hogervorst & van Damme. [182] noted a linear correlation between visual acuity and threshold for Gaussian blur filters. Due in part to this finding, many of the vision simulations that followed have also employed different implementations of Gaussian smoothing [21, 240, 241, 290]. However, as pointed out by Jones et al. [215], the large kernel sizes required to simulate severe blurring, like that which is experienced by people with moderate and severe low vision, can make traditional Gaussian convolution operations too computationally expensive for real time rendering. This is problematic, since real time performance and the maintenance of high frame rates is paramount to ensure that simulated vision impairments are rendered at the correct positions within the field of view when eye tracking is employed.

A solution to this problem is to use multi-resolution image representations. These data structures are often called *image pyramids* or *mipmaps*. Image pyramids are stacks of images with different resolutions, in which the highest resolution (original) image is at the bottom and the lowest resolution image is at the top—a structure that resembles the shape of a pyramid. Image pyramids are constructed

through two linear operations: filtering and sampling. For each level in the pyramid, the source image is filtered and then downsampled to produce a smaller version of the image from the previous level. In realtime vision simulations, this source image is the digital image captured by the display at each frame.

Image pyramids are computationally efficient and their development was initially motivated by insights from human vision [68, 459], which is encouraging for applying this technique for vision simulation. Tanimoto and Pavlidis's seminal paper on the use of hierarchical data structures for image processing was inspired by the visual encoding performed by the human retina [459]. The authors argued that hierarchical data structures, as opposed to uniform arrays, better mimic the array structure of the retina in which wide angle, low-resolution images (peripheral vision) and high-resolution foveal images (central vision) operate together to create efficient visual representations of an environment.

Image Pyramids in Low Vision Simulations The idea of employing an image pyramid to produce symptoms of vision loss was first proposed by Perry and Geisler in 2002 [137, 367]. Their method uses a Laplacian pyramid with a Gaussian weighting function to simulate reduced visual acuity. Yet, they do not take full advantage of the efficiency of the image pyramid data structure. In their solution, after the pyramid is constructed, they again process each image in the pyramid by upsampling all images to match the resolution of the original source image. These expanded images are used in lieu of the downsampled ones for blur computations. The authors then use a custom blending function to sample between adjacent levels within the image stack as needed to determine the value of each pixel in the final image [367].

Perry and Geisler's simulation operated between 20-60 frames per second (FPS). The poor runtime performance of their system may be explained by hardware and software limitations at the time of publication (2002) as well as by the additional operations they perform on the image stack after the pyramid is generated. Their approach deviates from the more computationally efficient sampling methods employed by standard and contemporary image pyramid algorithms [117, 263].

Recently, Jones et al. [215, 216] revisited the idea of employing multi-resolution pyramids to generate symptoms of low vision. The final version of their simulation was developed with the Unity game engine (version 2017.4.1f1), and their blur method employed the image pyramid method provided by the engine's mipmap generation feature [216]. The filtering method employed by their simulation is not explicitly stated in the published system description; however, they most likely employed a box filter, since this is the default filter employed to generate mipmaps in Unity 2017.4.1f1 *. They then used bilinear sampling to sample between adjacent levels of the pyramid as required to produce image blur. Jones et al.'s [215, 216] blur method forfeited the perceptual benefits of Gaussian smoothing in the interest of computational efficiency. But these two features need not be mutually exclusive.

Our Gaussian Pyramid Method In the current work, we take advantage of the benefits of both Gaussian blurring and image pyramid data structures by employing a Gaussian Pyramid to render visual field loss. We employ a Gaussian pyramid—as opposed to a Laplacian one [137, 367]—to better capture the concomitant reduction of contrast sensitivity and visual acuity that often accompanies vision loss [194, 394, 471]. Our system is also explicitly coded to handle complete blind spots in vision. These "disappearing" or perceptual "filling-in" effects that often occur with severe and complete visual field loss [87, 124, 462] can be simulated by inducing blur to the such severity that visual information from environmental features is completely lost.

*https://docs.unity3d.com/2017.4/Documentation/Manual/TextureTypes.html

An image pyramid is generated for each new image rendered by the display. The Gaussian pyramid consists of a stack of images where, at each level of the pyramid, the next image is smoothed with a Gaussian kernel and downsampled by a factor of 2 [68, 137]. The image from the previous level of the pyramid is used as input to generate the image at the next level. At the lowest resolution image of the pyramid (the maximum mipmap level), where the image resolution is mathematically zero, we include a base case in which the image is a 1×1 texel image. This value stored in this texel represents the average color value of the entire image.

After the Gaussian Pyramid is constructed, our algorithm uses the values within the visual field map to determine how the image pyramid is sampled to produce variable blur. When the VFM indicates no vision loss, the original image is sampled. When complete vision loss has occurred, the maximum level of the pyramid is sampled. The values stored within the VFM are continuous, but the image pyramid contains discrete levels of resolution loss. To compensate for this, when the value of the VFM corresponds to a level of vision loss that is between levels of the pyramid, trilinear interpolation is used to sample between the two nearest levels. Because the images stored within the pyramid are of a lower resolution than the output image, texture coordinates are also scaled to the dimensions of the selected levels during this operation.

Both bilinear interpolation (e.g., Jones et al. [215, 216]) and trilinear interpolation are common sampling methods for mipmaps [117]. We use trilinear interpolation in the current work, because it permits weighted averaging between mipmap levels. Although bilinear interpolation is more efficient [326], it can result in missed texels and abrupt switches between pyramid levels [117].

For image processing, Gaussian filters are controlled with two blur parameters: sigma and kernel size. Sigma is the standard deviation of the Gaussian distribution. The kernel is a small matrix of numbers that can be convolved with a source image to perform filtering operations. By default, our blur shader uses a small, 3×3 Gaussian kernel for smoothing to ensure reasonable runtime performance.

8.3 Summary

Our simulation is not the first to employ visual field data to visualize impairments on 2D displays [88, 148, 367] or on stereoscopic displays [127, 215, 445]. However, to the best of our knowledge, our system is the first to explicitly outline a principled approach for matching visual field data onto the field of view of a head-mounted display (Section 8.2.2). How developers measure and use display FoV in their calculations has ramifications for visual field mapping accuracy [292, 405]. As such, it is somewhat surprising that this has been left unreported in prior vision simulation research.

Our simulation is also one of the first in HMDs to render complete blind spots, not as misleading dark spots within the visual field, but as areas of missing information that are affected by perceptual filling. At present, we are only aware of one other HMD-based simulation that addresses perceptual filling as a symptom of low vision. Jones et al. [215]'s solution for perceptual filling used a simple "inpainter" approach, in which the linear-weighted sum of the four nearest pixel values were used to fill in areas where blind spots occurred. We argue that the approaches employed by the current work and by Jones et al. [215] provide more accurate representations of the symptoms described by low vision patients [87, 124, 194, 462] and observed by vision and optometry researchers [99, 237, 520]. This approach stands in stark contrast to a long line of vision simulation research in which blind spots are rendered as dark opacities in vision [69, 98, 127, 134, 140, 210, 269, 329, 340, 444, 477, 481].

For the next version of this simulation, we hope to better map the relationship between blur sever-

ity in our simulation and experienced visual acuity. Such a mapping would offer two key advantages. First, it would enable us to render blur with more precise control, in a way that corresponds to specific visual acuity values. Second, it would provide a deeper understanding of how users perceive and interact within the simulation under varying degrees of visual blur, ultimately leading to a more realistic and informative user experience. For example, we may better decide the value of sigma employed by our Gaussian filter at each level of the image pyramid by conducting a small user study in which participants with normal vision perform visual acuity tests with different blur parameters for the simulation. This approach for determining Gaussian filter parameters has also been used in prior vision simulation research [183, 241, 512]. For example, Wu et al. [512] asked 8 participants to report letters on a Snellen chart in VR to relate visual acuity to their blur factor. They found that acuity estimates fit an exponential curve. Using Landolt-C tests for visual acuity assessment, Hogervorst & van Damme [183] noted an inverse, linear relationship between visual acuity and the threshold values used in their simulation. The aforementioned research found different relationships between blur parameters and visual acuity, because the publications employed different displays as well as different methods for inducing blur. This underlies the need to perform a similar evaluation to determine the relationship between blur parameters and visual acuity in the current work.

In the interest of studying the relationship between visual field loss and space perception, we prioritize the simulation of visual field loss in this dissertation. However, vision impairments are heterogeneous. Patients with low vision may experience a wide range of other symptoms in conjunction with visual field loss, such as glare [215, 290], clouded lenses from cataracts [244], visual distortions [215, 310] and more. It will be fruitful to consider visualizing additional vision symptoms in future iterations of the simulation. This can be accomplished through the development of additional post-processing effects and shaders. At present, for visual field map input, our simulation uses 2D matrices in which a single value is stored at each datapoint. In future work, additional channels of information may be added to encode information for different vision impairment symptoms (e.g., visual distortions).

9 Eye Tracking

In this chapter, we describe our eye tracking data quality assessment, in which we evaluate the performance of the underlying hardware of two eye-tracked HMDs to determine if these devices are appropriate candidates for deploying our vision simulation. Understanding the capabilities and the limitations of an eye tracking system is imperative for gaze contingent applications [430, 439]. Understanding data quality is especially important for eye-tracked vision simulations, since data quality determines if virtual scotomas are correctly aligned with where a person is looking in the display [243, 430]. Another benefit to understanding the capabilities and limitations of an eye tracker is that one can use this information to design more appropriate applications that consider the limitations of the technology. For example, eye tracking researcher have developed methods to model appropriate area of interest (AOI) sizes, based on the data quality and filtering algorithms employed by eye trackers [348, 415, 480].

9.1 Motivation & Goals

Data quality is particularly important for gaze contingent rendering applications, like the vision simulation described in the current work, since correct eye position data is required to ensure that rendered images are correct when a person performs rapid eye movements. Poor data quality can influence fixation and saccade classification algorithms [52, 186, 188]. And poor data quality from eye tracking devices can result in inaccurate interpretations of results in studies that look at gaze behavior [52, 186, 188].

In the current work, we evaluate data quality in two candidate eye-tracked HMDs: the Varjo XR-3 and the HTC Vive Pro Eye. Both displays use a camera based pupil-corneal reflection approach for eye tracking, which is the dominant approach employed by commercially available eye-trackers [188, 189, 197]. With this approach, high speed video cameras in conjunction with infrared (or near-infrared) spectrum illuminators are used to track the position of the pupil during eye tracking. The infrared (IR) spectrum light creates highly visible reflections (or 'glints') on the cornea of the eye that remain relatively fixed on the camera sensor. The relative positions of the pupil and corneal reflections can then be used to determine eye movements. The difference in position between the pupil and reflections can also be used to help compensate for head movements. In enclosed HMDs, IR spectrum is particularly useful since IR light exposure can be controlled within the sealed headsets and since the human eye cannot see IR light. The Varjo XR-3 uses a custom solution for their integrated eye tracking, and the HTC Vive Pro Eye employs a Tobii integrated eye tracker. Both displays use two cameras within the headsets, with one for each eye. We discuss the manufacturer specifications for data quality in both displays in more detail in Section 9.

Krösl et al. [243] provide a concrete example of how poor data quality can negatively impact vision

impairment simulations. Data quality was not explicitly measured in Krösl et al.'s work; however, they documented several incidents that compromised the performance of their vision simulation during its application. During a behavioral evaluation of their cataract simulation in the HTC Vive Pro Eye, it was discovered that participants were able to look around simulated scotomas due to eye tracking errors and latencies within the eye tracking system.

Manufacturers of eye-tracking devices often provide specifications of system and eye tracking capabilities to showcase the best possible performance of their equipment. These statistics are typically recorded under ideal laboratory scenarios (e.g., with appropriate lighting, with a chin rest, etc.). Accordingly, manufacturer specifications are useful for understanding the best-case performance of eye-tracked devices, but they can be poor representatives of how the devices perform in real use cases [82, 430]. In addition, manufacturer specifications may not report added latencies from the display of the HMD, which is important to consider for gaze contingent rendering applications like the one we present in the current work. As such, before deploying our vision simulation, we evaluate three of the most common measures for evaluating data quality: (1) accuracy, (2) precision, and (3) data loss [187, 188].

9.2 Eye Tracking Methods

9.2.1 Data quality measures

In this section, we define and operationalize the data quality measures of interest. We also report the findings of prior research on data quality assessments in eye tracked head-mounted displays. Few empirical data quality assessments of eye tracked HMDs exist, and it is difficult to generalize the results of the current research due to differences across evaluated displays and evaluation protocols.

Accuracy

Accuracy, or *spatial accuracy*, is the closeness of the gaze position reported by an eye-tracking system to the actual gaze position of a user. It is operationalized as the difference between the true gaze position and the gaze position measured by the eye tracker. The smaller the offset; the better the accuracy. Figure 9.1 (left) provides a visualization of eye tracking data with high and low accuracy. For evaluations of accuracy, a participant is requested to look at a series of points positioned on a screen. Data is then recorded when the participant's eyes are fixating at each point. Accuracy is most often measured as the average angular offset in degrees of visual angle between measured fixation locations and the corresponding locations of the fixation targets [185, 463]. Accuracy can be expressed with the formula below (Equation 9.1), where θ_i represents the angular offset of a given sample and n represents a subset of samples. Accuracy can be computed separately for horizontal and vertical dimensions.

$$\theta_{Offset} = \frac{1}{n}\sum_{i=1}^{n}\theta_i \tag{9.1}$$

Manufacturers often report accuracies close to 1.0° visual angle for mixed reality head-mounted displays with integrated eye-trackers (e.g., Table 9.1). However, in practice, the actual accuracy of eye-trackers is often only comparable to manufacturer specifications for fixation targets near the center of

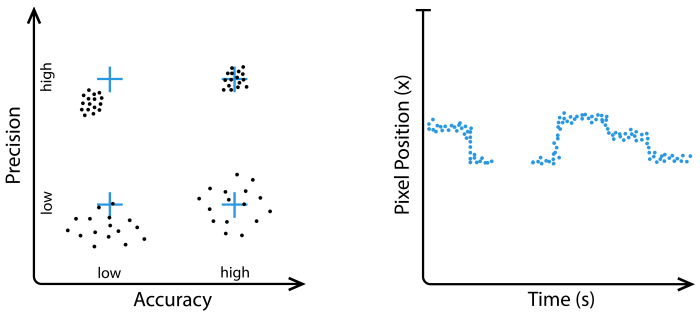

Figure 9.1: Visual representations of accuracy and precision (left) and data loss (right)

a display [280, 358, 430]. This degradation of eye tracking quality as eye movements approach the periphery of a display is a well documented problem in display-based eye trackers [188, 334].

Precision

Precision, or *spatial precision*, is the variable error (or the "closeness" of a set [334]) of the reported gaze positions. In eye tracking research, it is often also described as the reproducibility of a gaze position from one sample to the next when gaze position is stable [177, 188]. Small deviations between individual data points indicate high precision; large deviations indicate low precision. Figure 9.1 (left) provides a visualization of eye tracking data with high and low precision. Accuracy and precision are typically measured concurrently using the same procedure. Eye tracking precision is most frequently operationalized as either the standard deviation (STD) of the gaze position data samples or the root mean square of inter-sample distances (RMS-S2S) [185, 335]. STD measures the spatial dispersion of eye movement position data, whereas RMS-S2S measures the magnitude of displacement between subsequent gaze samples [335]. In other words, RMS-S2S captures the sample-to-sample jitter present within the system.

Reporting both measures is necessary to create a more complete picture of precision, since the two precision measures capture different aspects of variance in eye tracking data accuracy [188, 335]. There are ongoing debates within the eye-tracking research community about the trade-offs between these two approaches for measuring precision. For example, some researchers have argued that RMS-S2S, which relies on sample-to-sample distances, is inherently biased towards eye trackers with high sampling rates. Possible evidence for bias have been demonstrated in prior investigations by both Blignaut and Beelders [51] and Wang et al. [490]. However, Niehorster and colleagues [335] have pointed out that there are fundamental issues with the referenced eye tracking assessments, such as the downsampling of already filtered data, that undermine the claim of bias.

Precision can be computed separately for horizontal and vertical dimensions. Mathematically, STD is the square root of the sum of variance over a subset of data, where variance is the squared difference between datapoints and the mean of those datapoints. Equation 9.2 expresses the formula for the STD [187] for a set of n data samples, where x_i represents a data sample and $\bar{x}$ is the average of the data samples.

$$\theta_{STD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \tag{9.2}$$

Equation 9.3 expresses the formula for RMS-S2S [187], where θ_i denotes the angular distances between successive samples for a set of n data samples.

$$\theta_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\theta_i^2} = \sqrt{\frac{\theta_1^2 + \theta_2^2 + \ldots + \theta_n^2}{n}} \tag{9.3}$$

For accurate calculations of common eye tracking behaviors, an RMS-S2S lower than $0.05°$ is typically recommended [184, 493]. However, larger precision errors may be permissible for gaze contingent vision simulation since scotomas typically cover areas of the visual field as large or larger than the fovea, which spans $1.5 - 2°$ of the visual field.

Manufacturers rarely report eye tracking precision in HMDs (See Table 9.1). However, given the importance of precision for interpreting gaze position data, XR researchers have investigated a number of methods of quantifying precision in eye tracked HMDs. Thus far, RMS-S2S has been the most common measure used to report precision. Pastel et al. [358] reported an average precision of $0.07°$ in an HTC Vive retrofitted with an SMI eye tracker, while Sipatchin et al. [430] reported an average of $2.17°$ in the HTC Vive Pro Eye.

Data loss

Data loss refers to the percentage, or portion, of samples during which no gaze coordinate was reported [187, 188, 493] during a session. Figure 9.1 (center) provides a visualization of data loss in eye tracking data. Data loss is reported relative to the sampling frequency of the eye tracker. For example, it is expected that a 200 Hz eye-tracker, like the one employed by the Varjo XR-3, will deliver 200 gaze coordinates per second. If 20 frames of that data are missing, then 10% data loss is observed. Data loss is another important measure to include in data quality evaluations since data loss can greatly affect dependent eye tracking measures like fixations [494].

Some data loss during eye tracking is inevitable. Blinks alone can account for about 2% of data loss in a given data set [185]. Data loss can also occur when glasses, eyelashes, makeup, or other factors prevent the video camera from capturing an image of the person's eye. In the HTC Vive Pro Eye, Sipatchin et al. [430] reported 3.69% data loss on average when head movements were kept still and 7.76% data loss when head movement was encouraged. In the current work, we use the formula for data loss from Nystrom et al. [341], in which the proportion of valid data samples P_v is defined as:

$$P_v = \frac{N_{valid}}{N_{all}}, P_v \in [0, 1] \tag{9.4}$$

where N_{valid} and N_{all} represent the number of valid samples and the total number of samples collected, respectively. This proportion is then converted to a percentage.

9.2.2 Data preprocessing

Data calibration and validation with fixation targets are contingent on the assumption that participants are able to maintain a consistent fixation towards a target. However, even during fixations, the human eyes are in constant motion (e.g., microsaccades, gaze drift). Therefore, data preprocessing and cleaning is often required to correctly interpret fixation data.

Table 9.1: Manufacturer provided specifications for the Varjo XR-3 and HTC Vive Pro Eye[*†‡]. The reports of accuracy for the HTC Vive Pro Eye are based on the central 20° of the display[†].

Varjo XR-3			
Display Specifications		**Eye Tracking Specifications**	
field of view (°)	115 × 90	accuracy (°)	1
IPD range (mm)	58-72	precision (°)	-
refresh rate (Hz)	90	sampling frequency (Hz)	200
HTC Vive Pro Eye			
Display Specifications		**Eye Tracking Specifications**	
field of view (°)	107 × 107	accuracy (°)	0.5-1.1
IPD range (mm)	61–72	precision (°)	-
refresh rate (Hz)	90	sampling frequency (Hz)	120

Standard data cleaning protocols for eye tracking measures do not exist [150]. There are as many approaches to eye tracking data preprocessing as there are research groups analyzing that data. Often, the methods employed are dictated by the precedent set by a particular research group, by the experimental task, and by the eye movement measure of interest [74, 123]. Researchers may discard suspected, spurious data points based on distance thresholds, temporal thresholds, event detections, statistical outlier detection methods, fixation algorithms, manual inspection, or some combination of these methods. They may also completely remove a participant's data from an analysis if smoe percentage of their data is invalid. A recent review on data cleaning practices for eye movement data in reading research by Eskenazi et al. [115] found that 89% of 192 surveyed articles reported using at least on data cleaning method. In the last few years, a growing number of researchers have expressed concern over inconsistencies in both the reporting and the application of data cleaning methods across eye tracking publications [74, 123, 149, 150, 177].

One reason for this concern is that, although data processing is often necessary when analyzing eye-tracking data, what methods are chosen can impact on the results of one's analyses [177, 349]. Hessels et al. [178] demonstrated that how areas of interest (AOIs) are defined (e.g., size, position) can impact experimental outcomes in visual attention research. And Shic et al. [422] provide a particularly extreme example in which fixation-classification algorithm settings reversed the direction of differences between two experimental groups in an experiment where fixation duration was the primary measure.

At present, we are aware of three assessments of eye-tracking accuracy and precision in the HTC Vive Pro Eye (Adhanom et al. [25], Schuetz et al. [413], & Sipatchin et al. [430]) and two with the HTC Vive fitted with an SMI eye tracker (Lohr et al. [280], Pastel et al. [358]). As expected, inconsistencies in data processing reporting and methods can be found among these studies, as well. The publication by Adhanom et al. [25] reported a cursory assessment of eye tracking accuracy and precision with the HTC Vive Pro Eye using the author's tool, Gazemetric. However, the authors did not report what, if any, data cleaning steps were employed in their assessment. The other four articles employed multiple data processing methods–two of which were employed by all of the articles. This included:

[*]Varjo Technologies Oy (2024). https://developer.varjo.com/docs/get-started/eye-tracking-with-varjo-headset
[†]HTC Corporation (2024). https://developer.vive.com/resources/hardware-guides/vive-pro-eye-specs-user-guide/
[‡]VRCompare (2024). https://vr-compare.com

the removal of invalid data samples (data points where neither the left nor right eye were tracked) and the removal of data samples at the start of a fixation trial (400-700ms) [280, 358, 413, 430]. Eye tracking researchers (e.g., Holmqvist et al. [187]) often recommend removing samples at the start of fixation trials in order to avoid the influence of corrective saccades and post-saccadic oscillations when a participant initially transitions into fixation [42, 239].

Three of the four articles also used distance thresholds to remove spurious data points—although each research group adopted a different distance. Schuetz et al. [413] incorporated a threshold of 5° visual angle away from a fixation target; Pastel et al. [358] used a threshold of 3° visual angle; and a preprint by Lohr et al [280] used a threshold of 2° visual angle. Lohr et al. [280] additionally applied Tukey fences to remove potential outliers. Sipatchin et al. [430] did not employ any additional data cleaning methods.

Recall that data preprocessing decisions have a notable impact on measures like accuracy and precision. We see this play out in the studies conducted with the HTC Vive Pro Eye, as well. Schetz et al. [413] reported an average accuracy of 1.22° before and 1.08° after removing outliers. In contrast, Sipatchin et al. [430] reported substantially higher overall errors in gaze angle, with an average accuracy of 4.16°. One contributing factor to this discrepancy in results is data preprocessing. Schuetz et al. performed an additional data processing step to remove eye movements there were not fixations. Sipatchin et al. did not. Another contributing factor was related to experimental design. Sipatchin et al. [430] sampled a much wider span of the HTC Vive Pro Eye's field of view (53.2° × 53.2°) than the 30° × 30° field of view that was sampled in Schuetz et al. [413]. The degradation of eye tracking data quality as one's gaze moves farther from the center of a display has been well documented by previous display-based eye tracking research [120, 190, 334, 341].

Moving forward, it will be important to consider how our own experimental design and analysis decisions impact our results. Understanding this will allow us to better contextualize our data quality investigations with the HTC Vive Pro Eye and the Varjo XR-3.

9.3 Study 7 - Differences in data quality between eye-tracked MR HMDs

Understanding the data quality of eye tracked HMDs is imperative for determining their suitability for applications like simulated vision and foveated rendering, which rely on gaze-contingent rendering. Data quality is equally important for accurate interpretation of various eye-tracking behavior measures. As such, we evaluated data quality in the Varjo XR-3 and HTC Vive Pro Eye. The reported data quality values are critical considerations for both the application of our vision simulation and for future research utilizing the eye tracking functionalities of these devices.

Hypotheses

We predicted better data quality in the Varjo XR-3 for both accuracy and precision. This prediction was motivated by prior data quality assessments conducted in the HTC Vive Pro Eye, which have reported notably larger position errors than what was expected based on the eye tracking accuracy values provided by the manufacturer [25, 413, 424, 430]. Although we are not aware of any published eye tracking data quality assessments conducted to date in the Varjo XR-3, the manufacturer reported eye

tracking specifications provided by Varjo generally describe lower errors and higher sampling rates for eye tracking than those reported for the HTC Vive Pro Eye.

9.3.1 Materials

We employed two eye-tracked head-mounted displays for our investigation: the Varjo XR-3 and the HTC Vive Pro Eye. The Varjo XR-3 has differing maximum resolutions for the center and peripheral regions of the display. Within the center (27° x 27°) of the screen, images are rendered with a pixel resolution of 1920 x 1920 (70 ppd uOLED) per eye. Beyond this region of the screen, images are rendered with a pixel resolution of 2880 x 2720 (30 ppd LCD) per eye. The HTC Vive Pro Eye maintains the same resolution across an entire display. Its maximum per eye resolution is 1440 × 1600 (14 ppd OLED). The HMDs employ different eye tracking solutions. Both eye-trackers are binocular. The maximum sampling frequency of the Varjo XR-3 is 200 Hz, while the maximum sampling frequency for the HTC Vive Pro Eye is 120 Hz. Both systems use eye tracking to automatically compute a user's IPD. Additional information about the display and eye tracking specifications for these devices can be found in Table 9.1.

For this experiment, the HMDs were powered by a Windows 10 machine with an AMD Ryzen 9 7950X 16-core processor, 128 GB RAM, and an NVIDIA GeForce RTX 4090 graphics card. Applications for the HMDs were developed in Unity version 2021.3.4f1 with the C# programming language. Position tracking for both displays was performed using SteamVR (v2.1.10) and the SteamVR Tracking System with two 2.0 Base Stations. The Varjo Base Station (v3.10.3) software was used to run the device on our computer. The Varjo XR Plugin (v3.5.0) was then used to extract eye tracking data. The HTC Vive Pro Eye required the SRanpal Runtime (v1.3.6.11) software to access eye tracking data. The OpenXR Plugin (v1.4.2) and the VIVE OpenXR Plugin-Windows (v1.0.13) packages were then used within Unity to extract eye tracking data.

9.3.2 Participants

Twenty-two individuals (12 female, 10 male) volunteered for our experiment. The average age was 28.2 $\pm$ 4.5 years (min: 20, max: 38). Eye tracking studies often prescreen participants for vision correction, since glasses and contact lenses can degrade eye tracking data quality [341, 413]. We did not prescreen participants for vision correction during recruitment, because we were interested in measuring the data quality of these systems for a more representative population. Five participants wore glasses; one wore contact lenses.

Our methods were approved by the local institutional review board, and written consent was obtained from volunteers prior to participation. Each participant was given 10 USD for approximately 20 minutes of their time.

9.3.3 Design

Our experiment used a within-subjects design such that all participants performed the same data quality assessment in both the Varjo XR-3 and the HTC Vive Pro Eye. A within-subjects design was desirable for making comparisons between the two devices, because between participant factors—including differences in eye physiology (e.g., eye color, skin pigmentation, lash direction, baseline pupil size, etc.) and the use of vision correction (e.g., glasses, contacts)—can impact eye tracking data quality [176,

341]. The order in which participants experienced the two displays was counterbalanced to mitigate potential effects of order.

The processes for calibrating an eye tracking system and for validating an eye tracking system are similar. Eye-tracking validation is most often performed immediately after the eye tracking system has been calibrated. Both calibration and validation require participants to view a series of fixation targets. The positions of the fixation points used during validation should differ from the position of the points used in calibration to avoid underestimations of accuracy and precision [188, 341, 463]. Because the human eye is constantly in motion, data preprocessing is almost always required to remove measurements that were taken when a participant's eye moved away from the target due to inattention, microsaccades, or gaze drift.

To ensure fair comparison between the two systems, we employed eye tracking calibration routines with the same number of fixation targets for both the HTC Vive Pro Eye and the Varjo XR-3. Specifically, we used their native 5-point calibration solutions. The fixation targets for both displays are circular; however, the target for the HTC Vive Pro Eye is a solid circle while the target for the Varjo XR-3 is an animated concentric circle that becomes smaller as the participant's gaze nears it. Both devices use a rectangular layout to position the 5 fixation targets, with a single point at the center of the display. For both devices, the size and the positions of the fixation targets used during calibration were not accessible. Information about what filtering algorithms were applied to eye tracking data were also not accessible. All participants were able to calibrate using the eye-trackers. Both devices required users to have "valid" calibration data before they could continue to use eye tracking; however, the criteria for acceptable eye tracking calibration was not provided for either device.

For our data quality assessment, fixation targets spanned the central 20° of each display's field of view. We selected the central 20° to better compare the accuracy reported in our current evaluation against the values reported by the HTC Vive Pro Eye's specifications (See Table 9.1). The HTC Vive Pro reports an accuracy of 0.5°-1.1° within a 20° field of view. It does not provide a report of accuracy outside of this region. In contrast, the Varjo XR-3 reports an accuracy of 1°, but it does not specify what region of the display's field of view was assessed.

Our assessment used 9 fixation targets, which were positioned in a circular layout with a single point at the center of the display (See Figure 9.2). All other fixation targets were positioned 10° visual angle away from the center. We used a white, annuli (ring) as our target. Each annulus had an outer diameter that measured 1.0° visual angle and an inner diameter that measured 0.5° visual angle. The targets were rendered against a black plane that was positioned 1 meter in front of the viewer. This plane was fixed to always appear immediately in front of the viewer—even if their head moved. The background of the virtual environment beyond the plane was also rendered as black so that the black plane and the background were visibly indistinguishable.

Each target was presented twice during the experiment. As a result, participants completed 18 fixation trials for each device. The presentation order for the targets was pseudo-randomized to prevent the same target from appearing consecutively. Overall, data from 792 fixation trails were collected across all participants, with 396 fixation trials for each device.

9.3.4 Procedure

Before starting the experiment, the participant reported their basic demographic information. They were debriefed on the purpose of the experiment, and they provided written consent for their participation. Then, the experimenter guided the participant to a chair where they would remain seated

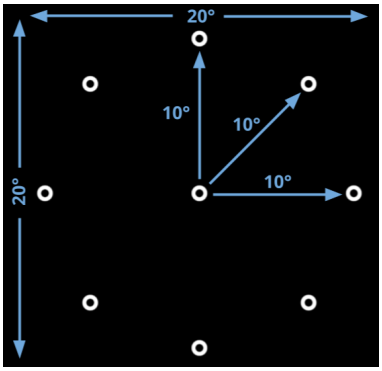

Figure 9.2: Circular layout of the 9 fixation targets used for the data quality assessment. Targets spanned 20° of each device's FOV. The eight exterior targets were positioned 10° visual angle away from the center. Each target was an annulus with an outer diameter of 1.0° visual angle and an inner diameter of 0.5° visual angle.

throughout the experiment. The participant was introduced to the two HMDs, and they were shown how to adjust the headset to fit and how to interact with the eye-tracking calibration system. The experimenter further specified that, during the calibration and validation tasks, participants should try to look at the "center-most" point of each target.

After the debriefing, the participant donned their first assigned HMD. The device's native eye tracking calibration procedure guided the participant through the headset adjustment task, in which it prompted the user to move the headset up or town to better capture their eyes if necessary. The calibration procedure also calculated participants' IPD. Then, the 5-point eye tracking calibration procedure commenced.

Immediately after calibration was completed, the experimenter gave the participant a wireless computer mouse, and they started the eye tracking data validation program. Unlike the calibration procedure, which progressed automatically, the validation procedure advanced to the next trial (the next fixation target) only after the participant pressed the left mouse button. At the start of every experimental trial, a single fixation target appeared. Once the participant had the target in sight and was ready to begin recording accuracy and precision, they clicked the mouse button. We relied on user input to ensure that a participant moved their gaze to a target before data recording was initiated [341]. After 3 s, the target disappeared and the next target was revealed. This process repeated until the end of the experiment. The same process was repeated for the second HMD.

Accuracy and precision were only recorded during the 3s capture window of each experimental trial. Data loss was recorded throughout the application's runtime.

9.3.5 Analysis

Preprocessing

The binocular eye trackers we employed for this study provide combined, left, and right eye-tracking information through their respective SDKs. However, our primary analysis only evaluates the combined gaze vectors, since the combined vector is what most gaze interactions employ by default. Information for the left and right eyes, separately, was collected but not formally analyzed.

In Section 9.2.2, we discussed common data preprocessing approaches for cleaning eye tracking data as well as several approaches employed by recent studies conducted with the HTC Vive [280, 358] &

Table 9.2: Samples removed, across all participants, at each step of the eye tracking data cleaning and data processing steps for the HTC Vive Pro Eye and the Varjo XR-3. Table rows indicate the different data processing steps, while the table columns report the total number of samples, the number of samples removed, the percentage of the remaining samples, and the percentage of the samples removed. The **Data Loss Dataset** included data from the start to the end of the application. The **Fixation Task Dataset** only included samples collected during fixation trials. Recall that the sampling rates of the eye trackers differ for each display (120Hz & 200Hz)

	HTC Vive Pro Eye				Varjo XR3			
Data Processing Step	*Total*	*Removed*	*% Total*	*% Removed*	*Total*	*Removed*	*% Total*	*% Removed*
All Data	369,193	-	100.0%	-	586,527	-	100.0%	-
Startup Compensation	359,070	10,123	97.26%	2.74%	569,327	17,200	97.07%	2.93%
Data Loss Dataset	359,070		97.26%		569,327		97.07%	
Performance Data Only	142,762	-	100.0%	-	234,570	-	100.0%	-
Event Contingency	119,225	23,537	83.51%	16.49%	196,368	38,202	83.71%	16.29%
Invalid Data	117,479	1,746	82.29%	1.22%	192,164	4,204	81.92%	1.79%
Outlier Removal	114,470	3,009	80.18%	2.01%	177,660	11,552	77.00%	4.92%
Fixation Task Dataset	114,470		80.18%		177,660		77.00%	

the HTC Vive Pro Eye [25, 413, 430]. The majority of this work relied on distance thresholds (e.g., the removal of all samples farther than 3° visual angle from a fixation target) to remove spurious eye movements made by participants during fixations [280, 358, 413]. However, distance thresholds would be an inappropriate choice for the current study, since we needed to compare the performance of eye tracking solutions with different specifications. The use of absolute distance thresholds could bias our analysis, since the spread of errors captured by the two devices were likely to differ [348]. In addition, we may be able to derive a reasonable distance threshold for the HTC Vive Pro Eye based on the results of previous data quality publications, but no reproducible data quality assessments of the Varjo XR-3 have been published to date.

For our data loss assessment (**Data Loss Dataset**, Table 9.2), the entire data set from the start to the end of the applications (All Data, Table 9.2) were used—except for the first 200ms of collected data (Startup Compensation, Table 9.2). This portion of the data was excluded to ensure that eye tracking registration and recording was not affected by initial program startup. The assessments of accuracy and precision (**Fixation Task Dataset**, Table 9.2) only included data collected during experimental trials, where each fixation trial consisted of 3s (3000ms) of data and where data recording was initiated by participant button press (Performance Data Only, Table 9.2). Then, the following steps were taken to clean the data:

- Event Contingency (Table 9.2). The first 500ms of data were discarded for each trial. These data were removed to prevent any potential interference with fixation behavior that may have been caused by the button pressing event [150]. The remaining 2500ms were included in the next data cleaning step.
- Invalid Data (Table 9.2). Invalid gaze data was excluded from analysis. Gaze samples in which the previous sample was invalid were also excluded, since sample-to-sample distances for RMS-

S2S could not be calculated for samples in which the previous data point was invalid. Discarding invalid samples is necessary since, as noted by Hessels et al. [177], accuracy and precision cannot be properly measured when the eye itself is not tracked. In this work, we considered gaze samples invalid not only when combined gaze data was invalid—but also when either the left or right eye was not tracked (as reported by the SDKs).

Here, we should note differences in how eye tracking status was reported in the two SDKs. HTC's SRanipal SDK provided a binary flag (a boolean) to describe the tracking status of the left, right, and combined eye data: [*True, False*]. Varjo's SDK similarly used a binary flag to report the status of combined eye data: [*Valid, Invalid*]. However, it provided more nuanced feedback for the left and right eyes: [*Tracking, Compensated, Visible, Invalid*]. Because we lacked insight into how and when eye tracking was compensated for in either device, we only removed eye data that was explicitly marked as *False* or *Invalid*.

- Outlier Removal (Table 9.2). Gaze fixation tasks assume that a participant's gaze is stable throughout an experimental trial. However, in reality, a subset of the data collected during this interval can be attributed to non-fixational eye movements (e.g., gaze drift). In this work, to identify and remove suspected, spurious gaze data, we utilized the modified Z-score, which is based on the Median Absolute Deviation (MAD) [165, 271]. Median-based outlier detection methods, like MAD and the modified Z-score, are less sensitive to outliers and cut-offs than those based on the mean [400, 420]. This property was desirable for our outlier detection method, since eye tracking offsets were strictly positive and had a prominent right skew (e.g., Figure 9.3). Kolmogorov–Smirnov tests further confirmed that data were not normally distributed for either device. We considered any datapoints with absolute modified Z-score exceeding 3.5 to be outliers. 3.5 was selected as a conservative threshold based on the recommendation from Iglewicz & Hoaglin [202].

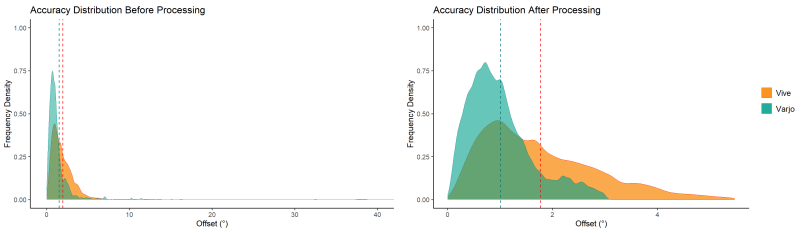

Figure 9.3: Frequency density plot of accuracy data in the HTC Vive Pro Eye and the Varjo XR-3 before (left) and after (right) eye tracking data processing.

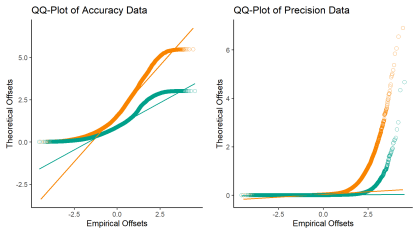

Figure 9.4: Q-Q plots of empirical vs. fitted data for accuracy (left) and precision (right)

Table 9.2 reports the number as well as the percentage of samples that were affected by each data processing step. Note that the total number of samples collected for the HTC Vive Pro Eye and the Varjo XR-3 differs. This is due to the higher sampling rate of the eye tracker in the Varjo XR-3 (200Hz) in comparison to the 120Hz sampling rate of the HTC Vive Pro Eye's eye tracker (Recall Table 9.1). The impact of the data processing steps for each device can also be observed through the frequency density plots in Figure 9.3.

We used manufacturer provided callback functions to sample the eye tracking data at the appropriate rates for each device (120Hz for the HTC Vive Pro Eye, 200 Hz for the Varjo XR-3). Timing for events

(e.g., registration of mouse click events) and rendering, however, were handled within the main Unity thread. Unity is not thread-safe, so functions from Unity's API cannot be called within the eye tracking callback thread. Because of this, there may be slight differences in event timing. Given the experimental task (fixation task) and the data processing steps employed, we do not suspect that this had a notable impact on our analysis.

In this section, we have provided a detailed discussion of our data preprocessing steps. In recent years, several research groups have called for better reporting and rational for the selection of data processing steps taken when analyzing eye tracking data [74, 123, 149]. And, as we have seen from the prior data quality experiments conducted with the HTC Vive Pro Eye, data processing can dramatically impact experimental outcomes. We believe reproducible and transparent reporting is particularly important for the current work, since our work is the first to report eye tracking data quality for the Varjo XR-3.

Statistical Analysis

Eye-tracking data is not always well described by a Gaussian distribution, since the variability of this data is not only explained by human gaze behavior–which may be argued as Gaussian in accordance to the Central Limit Theorem [251]—but also system measurement noise. For example, when feature localization of an eye tracker becomes unstable during fixation tasks, rapid spikes in gaze position signals can occur due to rapid changes in detected pupil locations. In the current work, we can observe these incidents of measurement errors in the dramatic spikes of our raw gaze position scatter plots (Figure 9.2). To complicate this matter further, the characteristics of measurement noise vary based on the eye tracker [335] and based on the eye tracking filter employed [334].

The Gamma distribution provides a reasonable match for the properties of the angular offset's theoretical distribution, since the angular offsets are strictly positive. In the current work, we can visually observe the abrupt cutoff at zero and the heavy right tail of distributions in the density plots of gaze position offsets for accuracy measurements in Figure 9.3. Mathematically, the Gamma distribution also provides an appropriate explanation for the relationship between the mean and variance in the current work, since the variance of the Gamma distribution increases proportionally with the mean. Accordingly, the Gamma distribution has been employed in prior eye-tracking research to simulate gaze positions during fixation [480], to model saccade movement [130], and to model fixation duration [416]. This decision is reinforced by Kolmogorov-Smirnov tests and quantile-quantile (QQ) plots (Figure 9.4), which indicate that accuracy measurements do not conform to a normal distribution for either the HTC Vive Pro Eye ($D=.611$, $p<.001$) or the Varjo XR-3 ($D=.539$, $p<.001$).

Accordingly, to understand differences in data quality between the HTC Vive Pro Eye and the Varjo XR-3, we employed a generalized linear mixed-effects model (GLMM) for our analysis. GLMMs are a form of generalized regression that is appropriate for repeated-measures designs, because they allow for accounting of both within- and between-participant variability [386]. GLMMs are well-suited to dealing with unbalanced datasets [26], which is of particular relevance for eye tracking data analyses between systems with different sampling rates. In addition, GLMMs allow for the specification of specific data distributions and link functions in the event that data does not meet Gaussian distribution data assumptions.

Significance levels were calculated using Satterthwaite approximation via the `lmerTest` package [250]. To interpret the interaction between device and fixation target, we calculated planned contrasts on the estimated marginal means using the `emmeans` package in R. To mitigate the risk of Type I error from

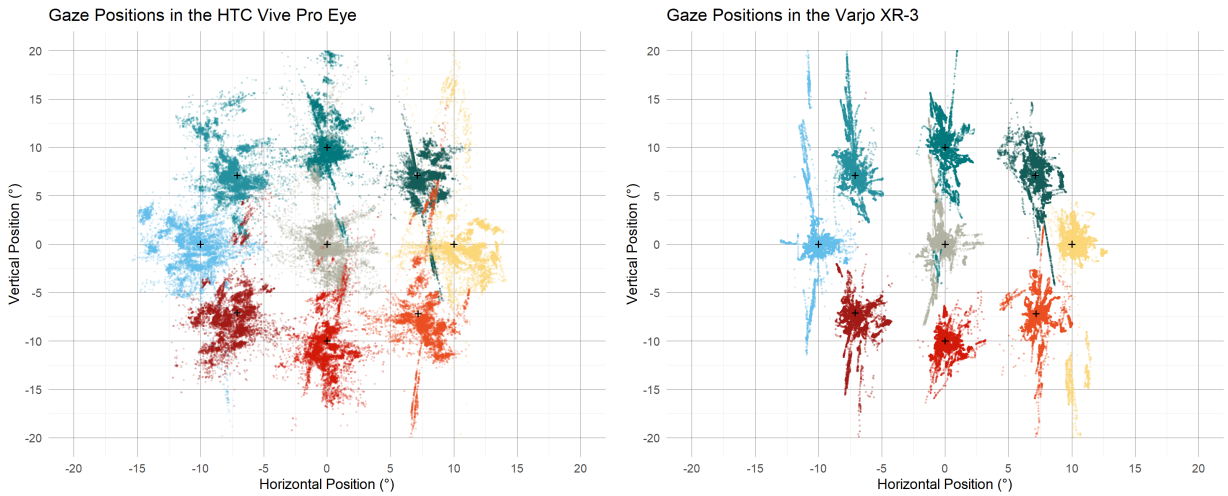

Figure 9.5: Scatterplots of the gaze positions in degrees visual angle of the combined gaze vectors for each of the nine fixation targets. Data are plotted for the HTC Vive Pro Eye (left) and the Varjo XR-3 (right). Black cross points indicate the true positions of fixation targets. Gaze positions beyond 20°, which corresponds to 2% of the total data for each device, have been clipped to improve the visibility of the plot.

multiple comparisons, we employed Bonferroni correction to adjust for pairwise comparisons.

Accuracy Accuracy was operationalized as the position offset in degrees visual angle between a gaze sample and a fixation target. To better understand eye tracking accuracy, we fit our GLMM with a Gamma distribution function (identity link), using the `glmer` function from the `lme4` library [40] in R version 4.3.2 [379]. The model included angular offsets as continuous outcomes.

Our experiment was designed to evaluate differences in eye tracking data quality between two devices. Accordingly, device (2 levels: HTC Vive Pro Eye, Varjo XR-3) was treated as a categorical predictor in our model. The order devices were experienced was also included as a categorical predictor (2 levels: first, second). Although we counterbalanced the order in which the devices were experienced, we anticipated that our participants, whom had little prior experience with eye trackers, may improve at the fixation task over time. The fixation targets themselves were also included as categorical predictors (9 levels), since data quality varies with respect to gaze angle [188, 334]. To account for individual variability in gaze behavior over repeated measures, we included participants as a random intercept $(1|Id)$. The formula for our GLMM in `glmer` model syntax is expressed by Equation 9.5:

$$\begin{aligned} \Upsilon = Offset \sim\ & Device + Target + Device : Target \\ & + Order + (1|Id) \end{aligned} \tag{9.5}$$

Because our predictors were categorical variables, an effects coding scheme [83] was used to ensure that parameter estimates corresponded to main effects [58]. We used simple contrast coding to compare each level of our categorical variables. Contrast coding required setting reference levels for our categorical factors. For device, the HTC Vive Pro Eye was set as the reference level. For order, the first block of trials was set as the reference level. For our fixation target predictor—which had 9 levels—the center fixation target was set as the reference.

Table 9.3: Accuracy and precision for targets, collapsed across participants, are reported. Accuracy is presented with both the mean and median angular offset for the combined gaze vectors. Precision is reported using standard deviation (STD) of angular offset samples as well as the root mean squared of inter-sample distances (RMS-S2S).

	HTC Vive Pro Eye				Varjo XR3			
	Accuracy		*Precision*		*Accuracy*		*Precision*	
Target	*Mean*	*Median*	*STD*	*RMS*	*Mean*	*Median*	*STD*	*RMS*
Center	1.55	1.17	1.07	0.0948	0.92	0.82	0.56	0.0100
Left	2.05	1.86	1.22	0.0995	0.94	0.82	0.64	0.0119
Right	2.03	1.73	1.17	0.0736	1.04	0.98	0.54	0.0094
Top	1.40	1.09	0.91	0.0584	1.03	0.98	0.57	0.0095
Down	1.79	1.74	1.04	0.0665	0.93	0.77	0.59	0.0096
Top Left	1.69	1.38	1.06	0.0760	0.99	0.86	0.63	0.0113
Top Right	1.53	1.38	0.97	0.0535	1.16	1.01	0.66	0.0099
Bot Left	1.93	1.68	1.09	0.0826	0.97	0.82	0.68	0.0103
Bot Right	1.96	1.68	1.11	0.0799	1.07	0.86	0.71	0.0096
TOTAL	1.77	1.57	1.10	0.076	1.01	0.88	0.62	0.0101

Precision We report both RMS and STD precision for our predictors in the results section. However, for our statistical analysis of precision, we evaluated RMS-S2S (or just RMS) precision. RMS precision was selected to better draw comparisons between our work and previous eye tracking data quality research in eye-tracked HMDs, since this measure has been the most commonly employed by MR researchers [22, 220, 358, 430]. To analyze RMS precision data, which had numerous zero and near-zero measurements, we required a method that could handle zero-inflated data [278]. Kolmogorov-Smirnov tests and quantile-quantile (QQ) plots (Figure 9.4) further support that precision measurements did not conform to a normal distribution for either the HTC Vive Pro Eye ($D=.5$, $p<.001$) or the Varjo XR-3 ($D=.5$, $p<.001$). The sharp upward curve of the precision QQ plot is a particularly strong visual indicator that the log link function is appropriate for this data. As such, for our analysis we employed the `glmmTMB` package in R [60] to specify a zero-inflated Gamma distribution (`ziGamma`) with a log link function. The log link function is commonly employed with Gamma fit mixed models, since it naturally corresponds to the exponential family distribution and since the log link helps to ensure that predicted values remain positive. The predictors for our model were the same as used previously (Equation 9.5).

9.3.6 Results

In this section, we report the descriptive statistics of our data quality measures as well as the results of our statistical analyses for accuracy and precision.

Figure 9.5 shows the raw gaze position offsets in degrees visual angle, collapsed across participants, for each fixation target in the HTC Vive Pro Eye (left) and the Varjo XR-3 (right). Examining the raw data through plots provides valuable insights beyond statistical analyses. For example, the plots reveal a significantly smaller spread in data points for the Varjo XR-3 compared to the HTC Vive Pro Eye. This visual confirmation supports the interpretation of our statistical results that we report in the subsequent sections. In addition, visualizing the raw data allows us to observe artifacts that are likely due to

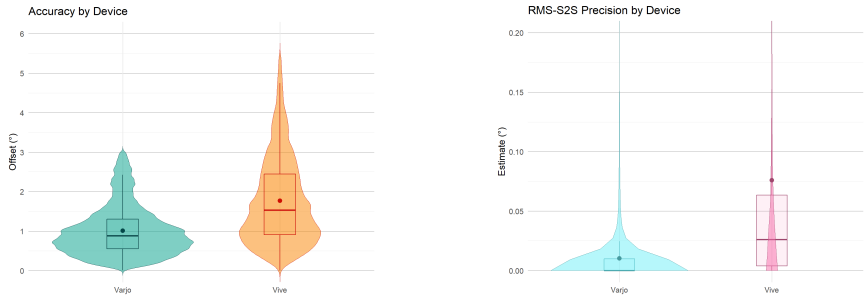

Figure 9.6: Violin plots for accuracy data (left) and for precision data (right) by device. The mean is indicated by a point.

eye tracker measurement errors (e.g., the sharp spikes in gaze offsets seen in the bottom-right target of both plots).

Accuracy

The average and median accuracies for the HTC Vive Pro Eye and the Varjo XR-3, as well as the average and median accuracies for each fixation target for each display, are reported in Table 9.3. The results of our statistical analysis for accuracy are reported in Table 9.4. We report regression coefficients (*B*), which provide an estimate of effect size; the confidence interval of the coefficients (*CI*); the ratios of estimated fixed effect coefficients to their standard error (*t-value*); and significance levels (*p-value*). For interactions, z ratios are reported.

Device Our statistical analysis supported that the difference in accuracy we found between the HTC Vive Pro Eye and the Varjo XR-3 was significant. We found a main effect of device in which gaze position offsets were .77° larger in the HTC Vive Pro Eye ($B = -.77, t = -228.78, p < .001$) than in the Varjo XR-3. Accuracy measurements for the two devices are visualized through violin plots in Figure 9.6

One HMD's eye tracking accuracy matched manufacturer specifications, while the other deviated (See Table 9.1 for device specifications). In our evaluation, the HTC Vive Pro Eye exhibited gaze position offsets exceeding the expected range ($.5 - 1.1°$) with an average of 1.77° ($Min = .002, Med = 1.53, Max = 5.47$) across trials and participants. Conversely, the Varjo XR-3 displayed accuracy measurements of 1.01° on average ($Min = .010, Med = .88, Max = 3.01$) that closely matched the manufacturer's specified value of 1.0°. These findings highlight the importance of independent evaluations of eye tracker data quality, since manufacturer provided specifications may not reflect the real-world performance of these systems.

Order There was a small but significant effect of order, in which accuracy was better by .03° on average in the second block of experimental trials ($B = -.03, t = -8.67, p < .001$). On average, accuracy was $1.34°(SD = .90)$ in the first block of trials and $1.27°(SD = .93)$ in the second block of trials. It is perhaps not surprising that there was a small improvement in eye tracking accuracy in the second block of trials, given that we did not actively seek out participants with prior experience with eye tracking during recruitment for this study.

Table 9.4: Results of planned comparisons using a Gamma regression model are displayed for accuracy data. *Estimates* represent the regression coefficients, *CI* is the confidence interval of the regression coefficient. The *Intercept* value is the grand mean position offset in degrees visual angle across both devices. Negative values for *Estimates* indicate that the first level of the predictor in the comparison had a larger average position offset (i.e., lower accuracy) than the second factor. Positive values indicate that the second level of the predictor had a larger average position offset (i.e., lower accuracy). Interactions are the result of planned contrasts on the estimated marginal means with Bonferroni correction.

Predictors		*Estimates*	*CI*		*t-value*	*p-value*
(Intercept)		1.41 ***	[1.27	1.54]	20.72	<.001
Device	vive – varjo	-.78 ***	[-.78	-.77]	-228.78	<.001
Order	first – second	-.03 ***	[-.04	-.02]	-8.67	<.001
Target	center – left	.21 ***	[.20	.22]	30.63	<.001
Target	center – right	.30 ***	[.29	.32]	43.38	<.001
Target	center – top	-.07 ***	[-.09	-.06]	-12.99	<.001
Target	center – bottom	.16 ***	[.15	.17]	24.49	<.001
Target	center – top left	.03 ***	[.02	.04]	5.24	<.001
Target	center – top right	.06 ***	[.05	.07]	9.52	<.001
Target	center – bottom left	.17 ***	[.16	.18]	25.34	<.001
Target	center – bottom right	.26 ***	[.25	.28]	37.93	<.001

Random Effects	*Variance*
σ^2	.31
τ_{Id}	.03
ICC	.09
Model Fit	
Marginal R^2	.327
Conditional R^2	.390

Interactions		*Estimates*	*CI*		*z-ratio*	*p-value*
Device×Target						
	center - left | vive	-.41 ***	[-.45	-.37]	-31.95	<.001
	center - right | vive	-.48 ***	[-.52	-.44]	-37.19	<.001
	center - top | vive	.18 ***	[.14	.21]	17.20	<.001
	center - bottom | vive	-.34 ***	[-.38	-.30]	-27.85	<.001
	center - top left | vive	-.06 ***	[-.10	-.03]	-5.78	<.001
	center - top right | vive	.06 ***	[.02	.09]	5.31	<.001
	center - bottom left | vive	-.38 ***	[-.41	-.34]	-30.44	<.001
	center - bottom right | vive	-.48 ***	[-.52	-.44]	-37.48	<.001
	center - left | varjo	-.01	[-.03	.00]	-2.78	.122
	center - right | varjo	-.13 ***	[-.14	-.11]	-23.72	<.001
	center - top | varjo	-.03 ***	[-.04	-.01]	-5.50	<.001
	center - bottom | varjo	.02 *	[-.00	.03]	3.17	<.05
	center - top left | varjo	.00	[-.02	.02]	.06	1.00
	center - top right | varjo	-.17 ***	[-.19	-.15]	-30.67	<.001
	center - bottom left | varjo	.04 ***	[.02	.05]	7.70	<.001
	center - bottom right | varjo	-.04 ***	[-.06	-.03]	-8.40	<.001

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

A priori, we did not predict an interaction between device and order. However, given the significant main effect of order in our analysis, we decided to perform a model comparison to test for the presence of a Device×Order interaction, since this interaction could have notable impact on the interpretation of our results. Of concern, this interaction between device and order could indicate that some aspect of our experimental design influenced experimental outcomes, as well. The `performance` package in R was used to compare models and produce a model performance score [288]. This score is calculated by normalizing all fit indices of a model (i.e., Akaike Information Criterion or AIC, Bayesian Information

Criterion or BIC, conditional and marginal R2, ICC, Root Mean Square Error of Approximation or RMSE, and Bayes Factor or BF), taking their mean, and then returning a score from 0% to 100%. Higher scores indicate better model performance.

Model comparison revealed the model with the interaction included performed no better than the original model without the interaction, with both receiving performance scores of 50%. Because model fit was not improved by including this interaction and because the Device×Order interaction reported in this model was not significant, we opted not to include this interaction term in our final model.

Target It is well documented that in display-based eye trackers, data quality decreases as gaze eccentricity from the center of the screen increases. We can see this trend in our main effects of fixation target position in Table 9.4, in which accuracy was significantly better at the center target compared to all other fixation targets—except for the one positioned at the top ($B = -.07, t = -12.98, p < 0.001$). While the main effects analysis reveals significant differences between target position conditions, it is crucial to consider the potential interaction between target position and device type. Eye tracker precision is known to vary across the visual field, and these variations might differ between the two devices under investigation. Therefore, interpreting the main effects without considering this interaction could be misleading.

Device×Target Because these HMDs employ different solutions for eye tracking, there are a number of reasons we can suspect that precision will vary between the two displays. For example, the physical position of the eye tracking sensors within the display chassis are different. We provide the statistical results for our analysis of Device×Target interaction in Table 9.4. However, it may be easier for the reader to interpret the interactions by observing the forest plot of these interaction effects in Figure9.7.

For the HTC Vive Pro Eye, accuracy was significantly different between the center target and all other fixation targets. Accuracy was better at the center fixation target compared to 6 of the targets. And accuracy was worse at the center fixation target compared to two of the other targets: the top fixation target ($B = .18, z = 17.20, p < .001$) and the top right fixation target ($B = .06, z = 5.31, p < .001$).

For the Varjo XR-3, accuracy was significantly different between the center target and 5 of the other fixation targets. Accuracy was better at the center fixation target compared to 4 of the targets. And accuracy was worse at the center fixation target compared to one other fixation target: the bottom left target ($B = .04, z = 7.70, p < .001$). There was no significant difference in accuracy between the center fixation target and 3 of the other fixation targets: the left, bottom, and bottom left fixation targets.

From the size of coefficients produced by our analysis and from visual inspection of these coefficients in Figure9.7, it is apparent that accuracy varies more across the screen in the HTC Vive Pro Eye than it does for the Varjo XR-3. We can see this in the comparatively larger distances between fixation targets and the center fixation target in the Vive Pro. From an applied perspective, developers should consider that eye tracking accuracy will vary across the screen in this device. To facilitate tasks like gaze interactions with virtual interfaces and correct interpretation of gaze behavior during visual search, it may be beneficial to use larger areas of interest (AOIs) in this display. The Varjo XR-3 has comparatively smaller differences in accuracy across the screen–at least within the range of the screen that we evaluated in this study. It will be beneficial to look at the Varjo XR-3's eye tracker performance across a wider span of the display's field of view given its performance in this experiment.

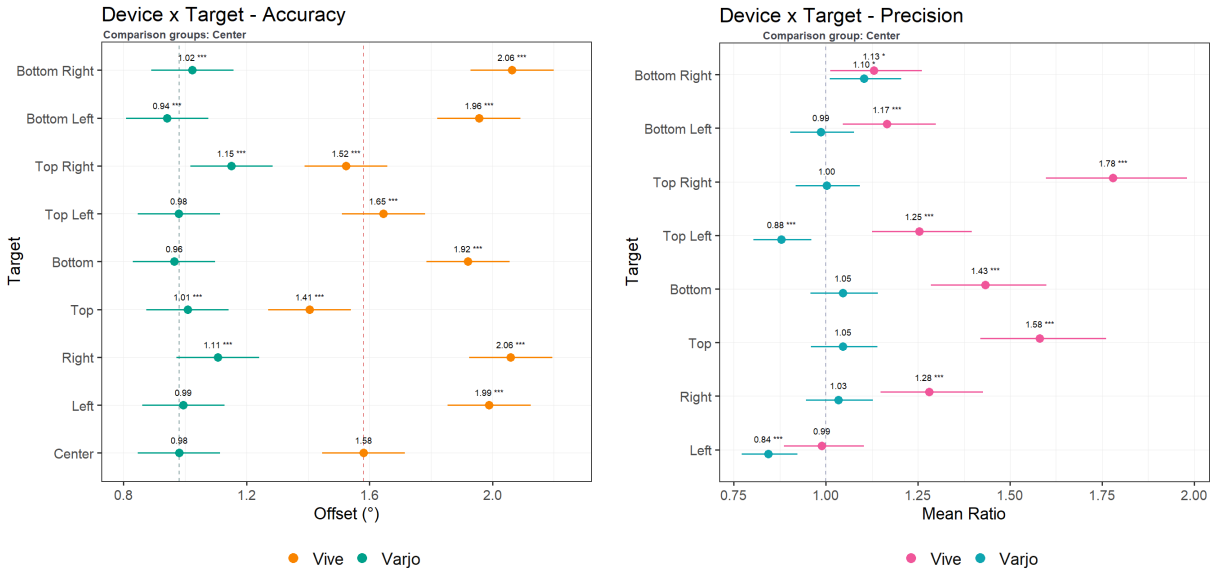

Figure 9.7: Forest plots depicting Device×Target interactions for accuracy (left) and for precision (right). For precision data, we plot the mean ratios from our model for easier interpretation. All statistical comparisons are made relative to the center fixation target. Significance is Bonferroni corrected for multiple comparisons.

Precision

The output of our GLMM for precision is reported in Table 9.5. To facilitate interpretation of model estimates on the original scale of the response variable, estimates (B) from the Zero-Inflated Gamma GLMM with a log link function were exponentiated. These exponentiated coefficients represent the multiplicative effects on the expected value for a one-unit increase in the corresponding predictor. As such, in Table 9.5 estimates > 1 indicate that the second level of the categorical factor is associated with larger position offsets (i.e., worse precision); coefficients < 1 indicate that the second level of the factor is associated with smaller offsets (i.e., better precision) compared to the first level.

We report both the root mean squared of inter-sample distances (RMS-S2S or RMS) and the standard deviation (STD) of angular offsets for the both devices and for each of the fixation target positions in these devices in Table 9.3.

Device We could not compare our measurements of precision against the manufacturer specifications for precision, since neither manufacturer provided reports on the precision of their eye trackers in the device specifications. However, given the high sampling rate and comparatively better accuracy of the Varjo XR-3, we predicted better precision in this display. Our statistical analysis indicates that RMS precision was significantly better in the Varjo XR-3 compared to the HTC Vive Pro Eye ($B = .23, z = -195.90, p < .001$). For precision, we report an RMS of .07° (STD=1.10°) across all fixation targets for the HTC Vive Pro Eye and an RMS of 0.01° (STD=.62°) across all fixation targets for the Varjo XR-3. RMS precision for both devices is depicted in Figure 9.6.

Order We found a main effect of order, in which precision was significantly better in the second block of trials than the first block of trials ($B = .96, z = -3.47, p < 0.001$). However, this effect was

Table 9.5: Results of zero-inflated Gamma regression model with log-link are displayed for precision data. *Estimates* represent the regression coefficients, *CI* is the confidence interval of the regression coefficient. Values for *Estimates* < 1 indicate that the first level of the predictor in the comparison had a larger RMS precision offset (i.e., lower precision) than the second factor. Values > 1 indicate that the second level of the predictor had a larger RMS precision offset (i.e., lower precision). Interactions are the result of planned contrasts on the estimated marginal means with Bonferroni correction..

Predictors		*Estimates*	*CI*	*z-value*	*p-value*
(Intercept)		.03 ***	[.02 .03]	-42.60	<.001
Device	vive – varjo	.12 ***	[.12 .13]	-195.90	<.001
Order	first – second	.96 ***	[.94 .98]	-3.47	<.001
Target	center – left	1.09 ***	[1.05 1.14]	4.02	<.001
Target	center – right	.87 ***	[.83 .91]	-6.31	<.001
Target	center – top	.78 ***	[.74 .81]	-11.42	<.001
Target	center – bottom	.82 ***	[.78 .85]	-9.07	<.001
Target	center – top left	.95 *	[.91 .99]	-2.1	< .05
Target	center – top right	.75 ***	[.72 .78]	-13.10	<.001
Target	center – bottom left	.93 **	[.89 .97]	-3.13	<.01
Target	center – bottom right	.90 ***	[.86 .94]	-4.91	<.001

Random Effects	*Variance*
σ^2	7.64
τ_{Id}	.16
ICC	.02
Model Fit	
Marginal R^2	.120
Conditional R^2	.138

Interactions		*Estimates*	*CI*	*z-ratio*	*p-value*
Device × Target					
	center - left | vive	- .01	[- .12 .10]	- .31	1.00
	center - right | vive	.25 ***	[.14 .36]	7.14	<.001
	center - top | vive	.46 ***	[.35 .57]	13.32	<.001
	center - bottom | vive	.36 ***	[.25 .47]	10.34	<.001
	center - top left | vive	.23 ***	[.12 .34]	6.59	<.001
	center - top right | vive	.58 ***	[.47 .69]	16.78	<.001
	center - bottom left | vive	.15 ***	[.04 .26]	4.45	<.001
	center - bottom right | vive	.12 *	[.01 .23]	3.48	<.05
	center - left | varjo	- .17 ***	[- .26 - .08]	- 5.96	<.001
	center - right | varjo	.03	[- .06 .12]	1.18	.960
	center - top | varjo	.05	[- .04 .13]	1.64	.784
	center - bottom | varjo	.04	[- .04 .13]	1.61	.801
	center - top left | varjo	- .13 ***	[- .22 - .04]	- 4.53	<.001
	center - top right | varjo	.00	[- .09 .09]	.07	1.00
	center - bottom left | varjo	- .01	[- .10 .08]	- .49	1.00
	center - bottom right | varjo	.10 *	[.01 .19]	3.53	<.05

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

quite small. This is corroborated by an RMS precision of .039° ($STD = .93°$) in the first block of trials and .033° ($STD = .90°$) in the second block of trials, when we collapse across device and target.

Because there was a main effect of order, we again evaluated the effect of Device×Order interaction through model comparison to ensure that this interaction did not impact the interpretation of our main results with the `performance` package. Model comparison revealed that our original model had a better performance score (100%) compared to the model with Device×Order interaction as a predictor (0%). Therefore, we opted not to modify our original model to include Device×Order interaction.

Target Just as accuracy can be expected to degrade with increasing eccentricity from the center of a display based eye tracker, so can precision. We can see this in our significant main effects of fixation target position in Table 9.5, in which precision was significantly better at the center target compared to all other fixation targets—except for the one positioned at the left ($B = -.07, t = -12.98, p < 0.001$). However, to properly interpret the influence of fixation target position, we must analyze the interaction between device and target.

Device×Target To better understand this main effect, we again evaluate the Device×Target interaction, since precision should vary based on screen position between the two devices. Table 9.3 reports the RMS and STD precision values for each target. And the results of our statistical analyses for Target as well as the Device×Target interaction are reported in Table 9.5. The interaction is then visualized through forest plots to facilitate ease of interpretation in Figure 9.6.

For the HTC Vive Pro Eye, precision was significantly better at the center target compared to 7 of the other fixation targets. There was no significant difference between the center and left fixation target in our assessment ($B = . - 01, z = -.31, p = 1.00$).

In contrast, for the Varjo XR-3 precision was only significantly different from 3 of the fixation targets. Precision was better at the center target than at the bottom right target ($B = .10, z = 3.53, p < .05$). Precision was worse at the center fixation target compared to two of the other targets: the left fixation target ($B = -.17, z = -5.96, p < 001$) and the top left fixation target ($B = . - 13, z = -4.53, p < .001$). This outcome is interesting, because it provides an indication that precision did not decrease with eccentricity within 20° of the center of the display (our sampled area for fixation targets) in our assessment. This is an encouraging outcome for the Varjo XR-3. It is also impressive how close to zero the precision measurements are for this device (See Table 9.3). Future research should consider sampling a larger potion of the display's field of view to determine if this pattern persists across the farther extents of the display. In contrast, the precision values reported for the HTC Vive Pro Eye in this study may present issues for applications where very small eye movements are required. As mentioned previously for accuracy, people who employ this eye tracker should carefully consider the size of AOIs and other visual targets to ensure that gaze is tracked sufficiently.

Data Loss

Based on the criteria we established in the data preprocessing section (Section 9.2.2), we counted samples as invalid when either the left, the right, or the combined eye sample was marked *False* for the HTC Vive Pro Eye or *Invalid* for the Varjo XR-3. With this criteria, we measured 6.28% data loss of eye tracking samples in the HTC Vive Pro Eye and 6.10% data loss of eye tracking samples in the Varjo XR-3.

We did not statistically analyze differences in eye tracking data loss between the HTC Vive Pro Eye and the Varjo XR-3 since the systems use quite different reporting mechanisms to convey the eye tracking status. Instead, in this section, we provide detailed descriptions of the different approaches employed by the two systems for reporting eye tracking states.

The HTC Vive Pro Eye uses a boolean flag [*True, False*] to describe the tracking status of the left, right, and combined eye data. The Varjo XR-3, in contrast, appears to use a more probabilistic approach for reporting eye tracking status. For combined eye data, the Varjo XR-3 uses a boolean flag [*Valid, Invalid*] similar to what is employed by the HTC Vive Pro Eye. However, Varjo's eye tracking SDK provides a total of four tracking states for the left and right eyes:

- *0 – Invalid* - Eye is not tracked and not visible (e.g., the eye is shut)
- *1 – Visible* - Eye is visible but not reliably tracked (e.g., during a saccade or blink)
- *2 – Compensated* - Eye is tracked but quality is compromised (e.g., the headset has moved after calibration)
- *3 – Tracked* - Eye is tracked

While the Varjo XR-3 provides some additional information about tracking states, neither of the display systems evaluated in this research clearly explain how eye-tracking status is measured or determined. In Table 9.6, we report the number of eye tracking samples as well as the percentage of eye tracking samples that were assigned to each of the available tracking states provided by the respective device's SDK. The breakdown of the Varjo XR-3's eye tracking states is particularly interesting, because a large portion of samples were labeled as *Compensated*: 13.46% for the left eye and 14.08% for the right eye.

By including this tracking state label, Varjo has indicated that eye tracking data is registered differently between the *Tracked* and *Compensated* states. This may also mean that eye tracking data is processed differently (e.g., different weights may be applied to data from the left and right eyes when one eye is *Compensated* for the calculation of the combined gaze vector or different filtering algorithms may be applied to eye tracking data when in a *Compensated* state). It may be of interest for researchers who conduct eye tracking research with the Varjo XR-3 to record the tracking states for eyes across trials to better understand how their experimental outcomes are affected by intermediate tracking states. This may be important to consider since eye tracking accuracy and precision are quite sensitive to changes in filtering methods [294, 335]. It is highly probably that the HTC Vive Pro Eye also employs some form of compensation algorithm to stabilize eye tracking data, which is a common practice in commercial eye tracked devices. Unfortunately, this information is not accessible through the eye tracking states provided by HTC's SRanipal SDK.

We believe it is also of interest to note eye tracking states and data loss information that was collected during fixation trials. In theory, eye tracking data should be relatively stable for this subset of samples, since participants were required to maintain fixations near the center of the display screens. When we looked at data loss that was incurred during fixation trials, when no samples were removed due to preprocessing, we measured 1.02% data loss in the HTC Vive Pro Eye and 1.01% data loss in the Varjo XR-3. Curiously, in the Varjo XR-3 we measured 39.86% and 41.70% of data samples were labeled as *Compensated*. In future work, it will be beneficial to evaluate how data quality differs between *Tracked* and *Compensated* eye tracking samples.

9.3.7 Discussion

This research investigated eye tracking data quality in two MR displays: the Varjo XR-3 and the HTC Vive Pro Eye. We conducted this evaluation to better understand the capabilities and limitations of these eye-tracking systems, since both were candidates for the deployment of our vision simulation.

As expected, angular offsets in the Varjo XR-3 were smaller than offsets in the HTC Vive Pro Eye. Our work is the first to evaluate eye tracking data in the Varjo XR-3. And it is one of the first controlled eye tracking data quality assessments conducted for any eye tracked Varjo display.

In theory, the work conducted by Keshava et al. [229] may provide useful insights for contextualizing our current investigation. The authors of this study evaluated data quality measures, including

Table 9.6: Eye tracking states and data loss reported by the HTC Vive Pro Eye and the Varjo XR-3 during the experimental session

	HTC Vive Pro Eye			Varjo XR3				
	Total	*True*	*False*	*Total*	*Valid/Tracked*	*Compensated*	*Visible*	*Invalid*
Samples								
Combo	359, 070	336, 515	22, 555	586, 527	534, 603	-	-	34, 724
Left	359, 070	333, 656	25, 414	586, 527	451, 497	76, 608	1, 497	39, 725
Right	359, 070	334, 745	24, 325	586, 527	446, 937	80, 135	1, 781	40, 474
Percent								
Combo	100 %	93.72 %	6.28 %	100 %	93.90 %	-	-	6.10 %
Left	100 %	92.92 %	7.08 %	100 %	79.30 %	13.46 %	.26 %	6.98 %
Right	100 %	93.22 %	6.77 %	100 %	78.50 %	14.08 %	.31 %	7.11 %

accuracy and precision, in the HTC Vive Pro Eye and the Varjo VR-2. Unfortunately, this work was featured in a short conference abstract that omits important information about their experimental design that is imperative for correctly interpreting their experimental results. An interesting outcome from this work is that they reported better average accuracy in the HTC Vive Pro (1.28° horizontal offset, .89° vertical offset) than the Varjo VR-2 (3.29°horizontal offset, 4.93° vertical offset) [229]. In a recent Bachelor's thesis, Heikkilä et al. [173] measured the accuracy and precision of the eye tracker in the Varjo Aero. Although this work was not peer-reviewed, the authors provide more details about their experimental protocol and analysis. Their results for accuracy are similar to what we found in the current work. They reported average accuracy of .85° (horizontal) and 1.17° (vertical) in the Varjo Aero after removing the first 500ms of fixation trial data in their experiment.

For the HTC Vive Pro Eye, our accuracy measurements were notably larger than the range of expected accuracy values provided by HTC: $0.5 - 1.1°$. This discrepancy in performance, at least for accuracy, may be expected based on previous data quality research investigations conducted with the HTC Vive Pro Eye. Schuetz et al. [413] and Sipatchin et al. [430], who conducted comprehensive investigations of eye tracking accuracy and precision in the HTC Vive Pro Eye, provide additional context for our study. The study conducted by Sipatchin et al. [430] measured unfiltered gaze data, while the study conducted by Schuetz et al. [413] measured filtered gaze data. Evaluations of unfiltered gaze data are important, because they can evaluate the performance and noise present in the eye tracker itself [120, 294, 436]. Filtering algorithms can introduce unanticipated biases in gaze data (e.g., Niehorster et al. [336]). However, unfiltered raw data are not always accessible, since eye tracking manufacturers often apply unspecified filters. In this context, understanding the performance of the entire end-to-end system, including the filtered data is important.

Sipatchin et al. [430] measured large errors in gaze positions relative to fixation target positions, with an average accuracy of 4.16° across all fixation targets. This result is unsurprising given that they evaluated unfiltered eye tracking data, performed minimal data preprocessing, and that they evaluated fixation targets across a wide FOV ($54° \times 54°$). When they constrained their analysis to a narrower subset of fixation targets (the center horizontal row) in their post-hoc analyses, they reported a better average accuracy of 2.26°.

The vast majority of data quality assessments, both formal and informal assessments, conducted

with the HTC Vive Pro Eye have evaluated filtered eye tracking data. The current work employed filtered eye tracking data, as does the investigation conducted by Schuetz et al. [413]. Schuetz et al. [413] conducted a data quality assessment with filtered eye tracking data in the HTC Vive Pro Eye with fixation targets that spanned 30° × 30°. They reported a mean accuracy of 1.22° before data processing and a mean accuracy of 1.10° across participants after data processing. When they constrained their calculations to only include fixation targets that fell within the inner 20° of FOV, they found a mean accuracy of 0.97°.

A number of informal eye tracking data quality assessments have also been conducted in the HTC Vive Pro Eye. Adhanom et al. [25] and Sidenmark et al. [424] measured average accuracies of 1.49° and 1.23°, respectively, in data quality assessments conducted using *GazeMetrcis* [25], an eye tracking validation tool. In a short conference abstract, Keshava et al. [229] reported accuracies of 1.28° for horizontal gaze offsets and .89° for vertical gaze offsets. Unfortunately, these documents failed to report important information about their experimental design (e.g., the extent of the visual field assessed, the size and shape of fixation targets) and data processing, so it is difficult to contextualize their results.

Much like this prior research on eye tracking data quality, our own work is not without limitations. For example, we only assessed data quality to fixation targets against a dark background. The amount of light provided by the display can impact eye tracking data quality, because the pupil becomes more difficult to track as it shrinks in brightly illuminated environments. For future data quality assessments, we suggest evaluating accuracy and precision against multiple backgrounds. Another limitation of our work with that we naively relied on the tools automatic (and semi-automatic) interpupillary distance (IPD) adjustment tools provided by the two displays. Because the IPD adjustment tools provided by these displays rely on eye tracking data to perform this adjustment, they may be prone to bias and noise just like any other eye tracking measure. To ensure that participant IPDs are correctly accounted for in the displays, we suggest manual IPD adjustment. It may also be of interest for future researchers to measure the discrepancies between these automatic IPD adjustments and participants actual IPDs, which can be measured via pupillometer.

Our investigation into data loss reveals another issue with interpreting eye tracking data in commercial devices. Both the Varjo XR-3 and HTC Vive Pro provide information about tracking states through their SDK. However, it is unclear how eye tracking status is determined for either device. We illustrate this issue in our discussion of the different tracking status labels provided by the two devices. Our research highlights the importance of recording eye tracking states beyond instances of complete data loss for more accurate interpretation of eye tracking data and gaze behavior.

Future Work - Latency For gaze-contingent rendering applications, another data quality measure is of great importance: latency. A recent data quality assessment conducted by Sipatchin et al. [430] on commodity-level HMDs suggests that system latency remains a primary concern for the development of gaze contingent rendering applications—especially for applications that manipulate images near the fovea. The simulation of central vision loss is particularly challenging since intrasaccadic perception can occur [136], which results in a misalignment between one's simulated visual field and their actual fovea. Misalignment could allow "peeking" at the percepts that are supposed to be masked. To prevent this effect, a low vision simulation that attempts to mask central vision should have an overall latency of 25 ms or less; higher latencies can be tolerated for peripheral field loss[406]. This overall latency is dependent on the time it takes for a system to register an eye movement event, as well as the displays' update rate. To better understand the impact of eye tracking data quality on gaze-contingent

rendering, we planned to conduct latency assessments for our eye-tracked head-mounted displays in Section 9. For this latency assessment, we wanted to include latency evaluations of the HMDs with and without the vision simulation, which would allow us to establish the baseline performance of the system and to measure any computational overhead introduced by our simulation. To ensure a focused investigation within the allotted time frame for the dissertation, the evaluation of system latency was deferred for future research.

9.4 Summary

In this chapter, we presented the framework for our vision simulation in technical detail, and we conducted a data quality assessment across two eye-tracked HMDs to better understand the capabilities and limitations of these devices so that we can design more effective eye tracking applications, like our gaze contingent vision simulation.

Our vision simulation has a wide range of applications beyond the current work, as well. For example, vision simulations can be useful design tools. By allowing developers with normal vision to experience vision impairments firsthand, they may better consider how design decisions affect users with vision impairments. They may also be useful in clinical settings to allow doctors and family members of patients to better understand the experience of a patient with severe vision impairment. However, there are inherent differences between the experiences felt by people with real versus simulated vision loss [469]. Simulations provide useful control and study of specific symptoms of vision loss. But they cannot recreate the social and environmental experiences of patients [409, 469]—nor can they recreate the adaptive behaviors that individuals with vision loss develop over time.

Our data quality assessment then revealed lower data quality for the HTC Vive Pro Eye compared to the Varjo XR-3. However, this finding doesn't negate the potential utility of the Vive Pro Eye. Many eye tracking applications can be—and have been—effectively implemented for this device. It is important to remember that we can often optimize the design of eye tracking application to function effectively within the capabilities of the eye tracker.

One example of this can be found in the work by Schuetz et al. [415]. In this work, the authors employed a psychophysics-inspired model to determine the optimal target size of UI elements for reliable gaze selection based on eye tracking data quality measures. Such research efforts, which encompass a range of complexities from simple to highly sophisticated methods, are not uncommon in eye tracking research. Better yet, the eye tracking data quality assessment from the current work, in conjunction with those collected in prior research, can inform decisions about optimal design parameters for eye-tracked applications.

For instance, we can use this information to determine the appropriate size for a simulated central scotoma. Given that the accuracy and precision of eye trackers can vary depending on the position of someone's gaze relative to the center of the screen, we suggest using larger scotomas (blind spots) in vision simulations when eye tracking accuracy and precision are poor. To determine the size of the scomota, One may consider using a similar approach to what eye tracking researchers have already employed for determining the size of targets for reliable gaze selection in VR [415], or for determining the size of AOIs for behavioral research [348, 480]. The data quality assessments conducted in this study, along with those reported by other researchers, will be instrumental in guiding the selection of the eye tracking device for the final dissertation experiment. This is particularly important because the HTC Vive Pro Eye will be used to deploy the vision simulation for this study. The primary factor influenc-

ing this decision was the Varjo XR-3's software development kit (SDK). Unfortunately, the Varjo SDK currently restricts the implementation of complex post-processing operations required by our application. Consequently, despite its superior data quality, the Varjo XR-3 was not a feasible option for the vision simulation and our final experiment.

Part VI

Enhancing Depth Perception for Vision Loss

Also note that invariably when we design something that can be used by those with disabilities, we often make it better for everyone.

Don Norman

10 Distance Perception, Vision Loss, & The Promise of Stylized Graphics

In this chapter, we describe our investigation into the influence of simulated visual field loss (VFL) on distance perception. For this, we employ our vision simulation in order to tightly control the extent of visual field loss for a within-subjects experiment in which individuals assess the distance to targets in space with normal vision, with central vision loss, and with peripheral vision loss. In addition, participants will be shown targets that are presented with either realistic, low contrast cast shadow connecting cues or non-photorealistic, high-contrast graphical elements along the ground. Through this investigation, we hope to generalize the results of our prior research on spatial perception and the influence of non-photorealistic graphics in HMDs to a more inclusive audience. The results of this research may have implications for both accessibility and the hardware design of immersive HMDs.

10.1 Motivation and Goals

The different areas of a person's visual field correspond to anatomically distinct regions that can be defined relative to the center of vision based on notable drop-offs in visual resolution. These regions are the *fovea*, the *parafovea*, and the *periphery* [375]. As eccentricity from the center of one's vision increases, visual acuity is reduced. The fovea is the most central region and it encompasses approximately 1–2° of the visual field [375]. It is responsible for providing sharp visual details in vision. The parafoveal region then encompasses approximately 5–8° visual angle from fixation (or eccentricity) [448]. The fovea and parafoveal regions together encompass *central vision* [259, 448]. The area of vision beyond the parafovea in which the reduction in visual acuity is most severe is commonly referred to as *peripheral vision* [259, 448].

The most common eye disease that results in loss of central vision is age-related macular degeneration (AMD). Near the onset of this disease, patients often experience scotomas with diameters smaller than 10° or less [450]. However, the extent of vision loss increases as the disease progresses. Often, central scotomas in patients with advanced AMD have diameters ranging between $10-20°$ [160, 168, 412]. In a study of 24 AMD patients, Guez et al. [160] reported a mean diameter of 10.3°. In a study of 21 AMD patients, Hassan et al. [168] reported a mean diameter of 14.8°. Schuchard et al. [412] evaluated a larger cohort of 255 AMD patients and reported median width and height of scotomas as 21.8° and 17.9°, respectively. In this work, we simulate a central scotoma with a 20° diameter.

10.2 The Impact of Vision Loss on Distance Perception

Central and peripheral vision provide different, complementary information about an environment as well as the objects and events that occur within it. As such, it is perhaps unsurprising that visual field impairments can affect people in different ways, depending on what part of their visual field is affected [116]. For example, patients with glaucoma and patients with age-related macular degeneration (AMD) report both common and distinct disabilities in self-report assessments of instrumental activities of daily living assessments [181] and in quality of life assessments [116]. In one report, Evans et al. [116] found that patients with AMD more frequently reported issues with their ability to perform physical day-to-day tasks and driving than did patients with glaucoma. They argue that one reason for this difference is that central vision remains intact for most patients with glaucoma.

At present, there is evidence that certain low vision conditions have a negative impact on tasks that require depth perception (e.g., difficulties with mobility [249], hazard detection[289], driving [365]). However, the relationship between visual field loss type and depth perception remains unclear.

10.2.1 Peripheral Vision Loss

Previous distance perception research in VR HMDs has shown that the reduced peripheral field of view (FoV), when combined with the added head weight of these displays, contributes to distance underestimation in virtual environments [64, 126, 211, 303, 501]. However, artificially restricting one's field of view when looking at real environments does not necessarily cause distance underestimation [95]. Further, the field of view restrictions in VR HMD studies are qualitatively different from than the peripheral vision loss (PVL) described by people with low vision. FoV restrictions in the aforementioned VR experiments completely block light at the periphery. However, low vision patients with PVL typically retain some light sensitivity even when functional vision at the periphery has been impaired [87]. Retaining light sensitivity in peripheral vision appears important for spatial perception. Evidence for this can be found in both VR display research [212, 214, 272, 273] and in research with low vision patients [238, 431].

For example, research in VR has found that depth compression in HMDs can be mitigated by adding light stimulation to the periphery [212, 214, 272] or by introducing low resolution graphics at the periphery [273]. Jones et al. [212, 214] reported this phenomenon with a constant white light added to the viewer's far peripheral vision. Li et al. [272] then demonstrated that adding a black frame, which completely blocked the participant's peripheral vision, decreased distance judgements. However, adding a solid white or middle gray frame improved distance judgments. In followup work, Li et al. [273] then demonstrated that applying a pixelated peripheral treatment could also improve distance judgments in VR.

Prior research on depth perception with low vision patients investigating the influence of peripheral vision loss (PVL) on depth perception has been more equivocal. Kotecha et al. [238] found that, compared to controls with normal vision, people with PVL are slower and make more tentative movements when reaching to targets; however, people with PVL exhibited few differences in grasp-posture and grip execution compared to controls. Glaucoma patients also exhibit more accurate and efficient movements in comparison to individuals with central vision loss [356, 470]. Although Fortenbaugh et al. [128] noted distance compression for patients with glaucoma in their research, their work suffers from a number of methodological problems that cast doubt on their statistical analyses and conclusions.

Patients with glaucoma express difficulties performing a range of spatial tasks that require accurate depth perception (e.g., walking, driving, obstacle avoidance) [20, 330]. However, prior investigations on the influence of simulated and real PVL have not provided a clear understanding of the relationship between PVL and depth perception. It is possible that depth perception errors in these spatial tasks are caused by an impaired ability to detect objects of interest and hazards—rather than an impaired ability to correctly perceive the distance to the objects [264, 473]. In our experiment, simulated peripheral vision will be rendered as a severe reduction in visual acuity—not an opacity—to better represent the vision loss experienced by individuals with PVL. In the current work, we predict that people's depth judgements to targets will be similar between simulated peripheral vision loss and control viewing conditions (**H1**). Our prediction is motivated by emerging evidence in prior research that the retention of light sensitivity in the periphery is more important than intact visual function for depth judgements.

10.2.2 Central Vision Loss

Impaired central vision makes distinguishing fine details about objects within an environment difficult, since central vision provides the sharpest visual acuity.It is likely that the inability to distinguish details about an object and its nearby surroundings may have an adverse effect on depth perception judgements to that objects. Evidence in support of this idea can be found in prior research on reaching and grasping behaviors in people with central vision loss (CVL).

For example, Verghese et al. [482] found that patients with CVL performed worse at a peg-placement task than age-matched controls. Further, among CVL patients, those patients with central scotomas that interfered with stereopsis performed even worse. Timberlake et al. [470] similarly reported that reach-to-grasp movements were less efficient and reaction times were slower for individuals with AMD when compared to age-matched controls. In comparisons of reaching and grasping behaviors between low vision patients with PVL and CVL, both Sivak et al. [431] and Pardhan et al. [356] found that central vision loss had a significantly greater negative impact on task performance. Although the results of research investigating the influence of PVL on depth perception is muddled, research on the influence of central vision loss (CVL) is more clear, especially for depth judgements at near distances (< 2m).

In the current work, we hypothesize that central vision loss will primarily influence distance judgments through an interaction effect with shadow contrast. Specifically, we predict that central vision loss will lead to a greater overestimation of distance for targets presented with low-contrast shadows compared to high-contrast shadows (**H3**). This is because impaired central vision may hinder the ability to detect and utilize subtle depth cues provided through surface contact, such as faint shadows, for accurate distance perception. Consistent with previous research, including the 5th experiment of this dissertation [4], failure to perceive these cues is expected to result in distance overestimation [314, 332].

10.3 Study 8 - Distance Perception with Simulated Vision Loss

Previously in this dissertation, we evaluated the influence of shadow-like elements along the ground for people's depth judgements to floating and grounded targets in AR HMDs. In this work, we found that the presence of cast shadows mitigated distance compression in AR for both grounded and floating targets in participants with normal vision. We also found that the appearance of a shadow could be manipulated to unrealistic extents and still preserve its role as a cue for depth perception. Our research

indicates that the visibility of a cast shadow is more important than its realism for enhancing spatial perception in MR displays.

This outcome is especially promising for the development of applications for optical see-through AR displays, since dark color values in these displays are highly transparent. As we have seen, connecting cues that attach objects in space to nearby surfaces—like cast shadows [102, 135, 175, 333] and nested contact relations between adjoining surfaces [314, 315, 383]—play an important role in depth perception for disambiguating the positions of objects in space. However, as shown by Rand et al. [383], if the element that connects an object to nearby surfaces is not visible, then the utility of that element is lost.

Over a series of experiments, Rand and colleagues studied distance perception in normally sighted individuals when global reductions in visual acuity were applied through blur goggles [383, 384]. They found that people's distance judgments to targets placed on the ground were accurate even when visual acuity was artificially degraded [384]; however, depth perception judgements became less accurate when the vertical position of a target [383] and when the connection between a target object and the visual horizon were manipulated [384]. In an especially relevant experiment to the current work, Rand et al. [383] compared distance judgments made to targets presented on visible stands (detectable stands) against judgments made to targets that were presented on stands with sufficiently low contrast relative to their backgrounds that they could not be seen through the blur goggles (undetectable stands). They found that people's distance judgements to targets in the undetectable stand condition were overestimated relative to judgments made to targets in the detectable stand condition. This phenomenon of overestimation is similar to what occurs in the total absence of surface contact information (e.g., no cast shadow) [314, 333, 352, 397].

It is possible, then, that an inability to detect important spatial information within an environment is a contributing factor to the depth perception difficulties experienced by individuals with low vision. In particular, depth perception errors may occur when scotomas (blind spots within the visual field) interfere with an individual's ability to detect important information about an object of interest, its environment, and the visual information (e.g., cast shadow) that connects an object to its environment. As such, in our proposed work, we wish to evaluate how central visual field loss (e.g., age related macular degeneration, diabetic retinopathy) and peripheral visual field loss (e.g., glaucoma, retinitis pigmentosa), influence people's depth judgements. For this assessment, we employ our low vision simulation for experimental control to draw comparisons between depth judgments made under central field loss (CVL), peripheral field loss (PVL), and normal vision conditions. Participants in our study will evaluate the influence of low and high contrast connecting cues, like the cast shadows from our prior experiments. Our research will provide insights into how visual field loss affects distance perception, and it will provide actionable guidelines for improving distance perception and visual accessibility in MR displays for people with low vision.

Hypotheses

More formally, our three hypotheses about the relationship between visual field loss, depth perception, and connecting cues are as follows:

H1. People with central vision loss (CVL) will underestimate depth judgement more in action space than people with normal vision

H2. Depth judgments to targets with low-contrast connecting cues will be overestimated relative to depth judgments made to targets with high-contrast connecting cues for people with central vision loss.

H3. People with central vision loss will have longer response times than people with peripheral vision loss and no vision loss.

10.3.1 Materials

We employed the HTC Vive Pro Eye head-mounted display for our investigation. Additional information about this display and its software dependencies can be found in the previous chapter (Section 9.3.1). Although we found that the Varjo XR-3 had significantly better accuracy and precision compared to the HTC Vive Pro in our data quality assessment, we were unable to use the Varjo XR-3 for our vision simulation, because the Varjo SDK prevents developers from interjecting complex post-processing operations into the graphics pipeline*.

Vision Simulation In our eye tracking data quality assessment with the HTC Vive Pro Eye, we found an average accuracy of 1.77° and precision measurements of .08° (RMS-S2S) and 1.10° (STD) (Section 9.3). Another data quality assessment conducted by Sipatchin et al. [430], which sampled a much larger extent of this HMD's FOV (54° × 54°), reported an average accuracy of 4.16°. Because we do not artificially restrict people's head and eye movements in our experiment, participants are allowed to move their eyes anywhere within the HMD's FOV. Considering the limitations of the eye tracker, we realized that it was necessary to use a central scotoma that extended beyond the parafoveal region (8° eccentricity). By rendering a scotoma beyond this region, we can minimize the risk of participants 'peaking through' the scotoma when moderate eye tracking errors occur, since the fovea and parafoveal regions of vision would still remain occluded. Rendering a central scotoma with a 20° diameter allowed us to meet both the theoretical and technical demands of our study, since a 20° eccentricity is representative of the vision loss experienced by a patient with advanced AMD (Section 10.1).

Our simulation rendered both central and peripheral vision loss relative to this boarder at 20° eccentricity. For the center vision loss condition, the center most 20° of a participant's visual field was obscured; for the peripheral vision loss condition, the visual field beyond 20° was obscured.

Our vision simulation employed an image pyramid with 7 levels. The resulting image pyramid for the HTC Vive Pro Eye had a resolution of 1400 × 1600 pixels per eye for the top layer and a resolution of 22 × 25 pixel for the bottom layer. A separable Gaussian filter with a 3 × 3 kernel ($\sigma = 0.8$) was applied to each downsampled layer. The area of the field of view that was affected by downsampling was dictated by a radial gradient area. The radius of the circle was set to match 20 degrees eccentricity from the center of the screen. Once eye tracking was enabled, the center of the circle was aligned to the center gaze position. Figure 10.1 depicts the three resulting vision loss conditions: no vision loss (left), center vision loss (center), and peripheral vision loss (right).

Environments The experiment was conducted in a 7.3 × 8.5 × 3.7m room that provided a 7.3m linear distance forward for blind walking. Two virtual environments were created for the experiment: one for vision simulation adaptation and one for the actual experiment. The adaptation environment resembled the interior of a café (Figure 10.2); the experiment environment resembled an outdoor garden (Figure 10.1). Both environments were rendered with detailed textures, and thematically appropriate 3d models were strewn across both environments.

*Varjo Technologies Oy (2024). https://developer.varjo.com/docs/get-started/Post-processing

Figure 10.1: Images captures from SteamVR's 'VR view' of the simulated vision loss conditions experienced by participants: no vision loss (left), center vision loss (center), peripheral vision loss (right). The floating cube target is shown is a low contrast connecting cue (a cast shadow)

10.3.2 Participants

Twenty-four individuals volunteered for this experiment in exchange for 10 USD and approximately 45 minutes of their time. The average age was 28.2 $\pm$ 4.4 years (Min: 20, Max: 38). Fourteen volunteers were female and ten were male. Six participants wore glasses and two wore contact lenses. Our methods were approved by the local institutional review board, and written consent was obtained from volunteers prior to participation.

We recruited 24 participants because we needed our participant count to be a multiple of 6 to balance our within-subjects experimental factors perfectly (See Design & Procedure). Prior research has demonstrated that data collected from 10 [205, 281] to 23 [232] participants is sufficiently powered to detect differences in distance judgements made through blind walking, in which distance perception is assessed by observers walking without vision to previously viewed targets. Further, in our prior distance perception experiments, we were able to detect differences with 24 participants using verbal report, which is a more variable and less accurate distance measure [13, 283].

10.3.3 Design

To address our hypotheses, we designed an experiment with a 3 (visual field loss) × 2 (shadow shading) × 4 (target distance) within-subjects factorial design with repeated measures. All conditions were presented twice to every participant, and distance judgments were obtained through blind walking [281, 283]. The three visual field loss conditions included: a peripheral vision loss (PVL) condition, a central vision loss (CVL) condition, and a control condition with no vision loss. The size of the scotomas used in this study were based partially on the eccentricities defined by central and peripheral vision, as discussed in Section 10.2 and partially on the data quality limitations of the HTC Vive Pro Eye's eye tracker 9.3. Central and peripheral scotomas corresponded to areas inside and outside of 20° eccentricity, respectively. Scotomas were rendered as low resolution blur fields rather than as pure opacities, since this provides a more accurate representation of vision impairments [87, 124, 503].

A virtual cube was presented at four distances from the viewer: 2m, 2.5m, 3.0m, & 3.5m. To im-

Figure 10.2: The virtual environment used for the adaptation phase of the experiment

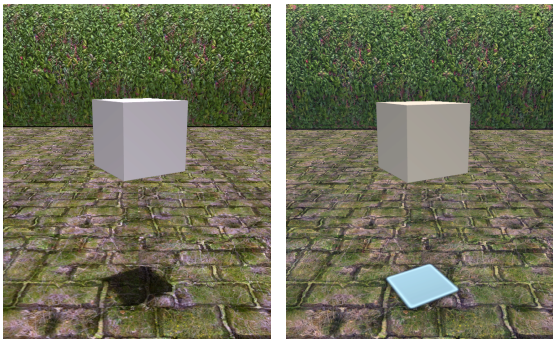

Figure 10.3: The low contrast (left) and high contrast (right) drop shadow connecting cues.

prove the generalizability of our experimental results, the cube was rotated randomly between 0 and 45°, horizontally [2, 135]. The virtual cube, depicted in Figure 10.3, measured 20cm in diameter, and it was rendered with a middle gray RGB color value of 128. Targets were positioned above the ground at 0.2m, and they were rendered with one of two shading conditions for connecting cues: realistic (dark, low-contrast) shadow or stylized (light, high-contrast) highlight along the ground (Figure 10.3). These two conditions were motivated by the prior research of this dissertation. Connecting cues were rendered immediately beneath each target during assessment.

The experiment was blocked by visual field loss condition, and the order in which visual field loss conditions were experienced was counterbalanced across participants. All other factors were pseudo-randomized within each block so that a participant views each unique combination once before experiencing the same combination again. Each participant completed a total of 48 trials during the experiment. with 24 participants, a total of 1, 152 trials was collected overall. For each participant, the experiment took approximately 45 minutes to complete.

The primary outcome measure for task performance in this experiment was distance judgment. However, we also report response time as a secondary outcome measure to better explain performance differences and the influence of scotomas on distance judgements. Response time is important when evaluating the influence of vision impairment on behavior, since often individuals with vision impairments are able to complete tasks—but at a slower rate and with more error [10, 181, 238]. Time may also be useful for understanding the difficulty or cognitive load of a task [38, 483].

10.3.4 Procedure

Upon arrival, participants were given a description of the experiment, an informed consent form, a proof of payment form, and monetary compensation for volunteering to participate in the study. Participants were outfitted with the eye tracked head-mounted display (the HTC Vive Pro Eye), and they performed the device's native 5 point eye tracking calibration. The experiment was blocked by vision condition, so the same procedure was repeated for each vision condition.

First, the participant was directed to the center of the room as the researcher launched the application for the adaptation phase of the experiment. During the adaptation phase, they were exposed to one of the three vision loss conditions (none, center, peripheral).

Adaptation Phase This adaptation phase served an important purpose: it allowed participants to familiarize themselves with the altered sensory input associated with the vision loss conditions. A person's perception of environmental scale—their ability to recover absolute distances and sizes in an environment [90]—is closely linked to their prior experiences and interactions with their environment. In virtual environments, there is evidence that a generalized rescaling of space occurs when appropriate perceptual-motor feedback is provided [225, 226, 319]. Most frequently, this feedback is provided through locomotion [3, 319]. Adaptations from walking interactions happen quickly. For example, two experiments by Kelly et al. [225, 226] demonstrated calibration effects during exposure times that ranged between approximately 2 and 3 minutes. In addition, calibration effects and the resulting rescaling of perceived space from walking interactions are robust and can transfer across different virtual environments [426].

An image of the virtual environment used for the adaptation phase can be seen in Figure 10.2. The participant was asked to explore and walk around this virtual environment for a minimum of 2 minutes, within the area indicated by the blue carpet, which corresponded to a cleared tracking space in the physical world. This adaptation phase was provided so that participants could update their understanding of environment scale, given the different simulated vision impairments. It was also included to help mitigate any novelty effects incurred during initial exposure to the simulated vision loss. A minimum interval of 2 minutes was chosen because we believed this amount of time would be sufficient for recalibration. It was also chosen to minimize participant fatigue and to encourage participant compliance by keeping the total time required for the experiment under 45 minutes. Because we employed a within-subjects experimental design, it was necessary to use a separate virtual environment for the adaptation phase to prevent participants from receiving feedback about their blind walking performance between the experimental blocks.

Experiment Phase Following the adaptation phase of the first experimental block, the participant completed a short, guided practice walk with their eyes closed, and they were guided in a backwards, circuitous route to the designated starting position for the subsequent test trials. The starting point was immediately next to a wall in the physical room. To prevent motor feedback from walking to the same starting point, the participant was again guided back to the starting point without vision and via backwards walking through a circuitous route at the end of each trial. Backwards walking was also conducted to avoid putting twisting strain on the HMD's tethered cable over numerous experimental trials.

For the testing phase, participants were told that target objects would appear at various distances and that the participant needed to walk to where they believed the target was positioned in space. At the start of each trial, participants were allowed to view the target for as long as necessary. They verbally indicated to the researcher when they were ready to begin the blind walking trial. Once the screen of the HMD went black, the participant initiated the blind walking trial. Once the participant finished their blind walk, the researcher guided the participant back to the starting point, again via a backwards, circuitous walk. The participant was given no feedback on their performance during the experiment.

The researcher used a wireless mouse click to advance the experiment. The start of the experiment, the start of an experimental trial, the start of the blind walking phase, and the end of an experimental trial were all advanced using a single mouse click. An audio cue was provided for each mouse click event to help avoid experimenter error (e.g., double clicks). At each mouse click event, position, performance, and time stamp data was collected for later analysis.

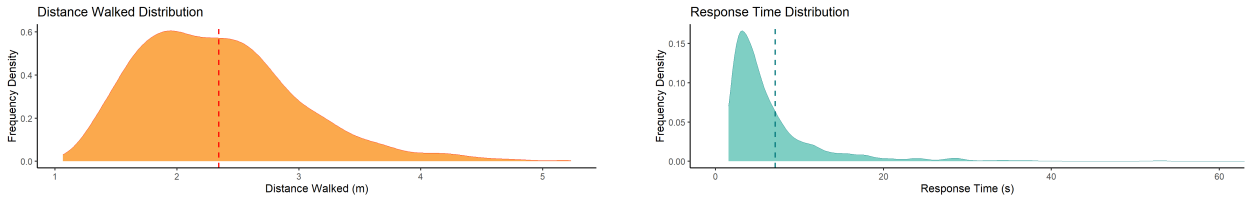

Figure 10.4: Data distributions for distance judgements as measured via blind walking (left) and for response time measurements (right) for the distance estimation experiment. The right tail of the response time (RT) distribution has been clipped at 60s to improve plot visibility. This distribution has a maximum value of 122s.

Within an experimental block, participants completed 16 trials. The same process was then repeated for all 3 experimental blocks.

10.3.5 Data Analysis

We statistically analyzed two measures in this report: distance judgements and response time. Distance judgements correspond to the actual distance walked by the participant for each trial. Blind walking distances were calculated as the Euclidean distance between the participants start and end positions, which were extracted from SteamVR's position tracking system via two SteamVR Base Stations (v2.0) in the corners of the room. Blind walking distances are reported in meters. Response times corresponded to the interval of time between when the target was presented to the user and when the user verbally indicated that they were ready to make a distance judgement through blind walking. Response times are reported in seconds.

A grand total of 1,152 trials were completed across all 24 participants, with each participant contributing data from 48 blind walking trials. Data from 13 trials (1.1% of the total data) were lost due to experimenter errors (e.g., prematurely exiting experiment application, accidentally double-clicking the mouse when progressing to the next part of a trial). As such, 1139 datapoints were evaluated in the final analyses.

To investigate the influence of our experimental factors, we analyzed people's responses with a mixed model approach. Mixed models are a form of generalized regression techniques that can account for both between-participant variability (i.e., the variability between different levels of categorical predictors across participants) and within-participant variability (i.e., the variability within each participant across conditions) in repeated measures research in which data are hierarchical, since they are collected under different conditions and nested within each study participant [386]. By employing statistical techniques that account for hierarchical data, effect size estimates and standard errors remain undistorted, and meaningful variance that may be overlooked when using aggregation is retained [351, 504].

We performed our statistical analyses using R version 4.3.2 [125]. Mixed models were specified using the `lmer` package [40]. Hypotheses were addressed through pairwise comparisons using the `emmeans` package. To interpret interactions between categorical factors, we conducted planned contrasts using estimated marginal means. To evaluate the effect of a continuous variable at different levels of categorical factors (i.e., the interaction between continuous and categorical factors), we calculated simple slopes through the `emtrends` function.

Satterthwaite approximation via the *lmerTest* package [250] was employed to calculate significance levels. Although Type I errors are somewhat reduced in hierarchical models compared to traditional

omnibus tests [159], they are not completely avoided [306]. Inflated type I errors can still emerge in the presence of complex covariance structures, small sample sizes, multiple comparisons, and imbalanced experimental designs [192, 291, 323, 407]. Because we conduct a confirmatory study with several pre-specified hypotheses analyzed through multiple pairwise comparisons in this study, we adjust significance levels for multiple comparisons with Bonferroni correction in our analyses [44, 323].

Distance Judgements

To answer our research questions on how depth perception judgements vary with visual field loss and the visibility of connecting cues, we used a linear mixed-effects model (LMM). For our LMM, we used the `lmer` function from the `lme4` package [40].

Our sampled distance judgement data was non-normal, as exhibited by a slight positive skew and a peaked distribution (See Figure 10.4 (left)). This was verified via Shapiro-Wilk test ($W=0.958$, $p<0.001$) and QQ plot inspection. However, because the population's distance judgement errors should have a Gaussian distribution, we assumed that the underlying distribution of distance judgements from the population was normal for our analyses [11]. Linear mixed models are robust to violations of distributional assumptions [408]. Further, to ensure that the results of our analyses were interpretable, we did not transform our observed data [50]. To avoid overfitting our predictor values to the current data set, we also did not fit our sampled data to another distribution [12].

For our LMM, we modeled continuous outcomes (distance walked) for our input variables (predictors). To answer our research questions, simulated vision loss, connecting cue, and distance from viewer were treated as factors. Simulated vision loss (3 levels: none, central vision loss, peripheral vision loss) and connecting cue (2 levels: low contrast, high contrast) were treated as categorical factors. We employed simple contrast coding to interpret the categorical factors in our experiment. The formula for our LMM in `lmer` model syntax is shown by Equation 10.1 below:

$$\begin{aligned}\Upsilon = \mathit{Response} \sim\ & \mathit{Distance} + \mathit{Vision} + \mathit{Shadow} + \mathit{Order} + \mathit{Vision} : \mathit{Distance} \\ & + \mathit{Vision} : \mathit{Shadow} + (1 + \mathit{Order}|\mathit{Id})\end{aligned} \tag{10.1}$$

Participants' distance judgments were recorded and statistically analyzed in meters. Target distance was treated as a mean-centered, continuous factor. In addition, because we anticipated that participants would not perfectly position themselves at the center of the starting point at the start of each trial, the target distance was calculated as the true distance between the viewer and the target at the start of a trial. This ensured that participants' walked distances were compared to the correct distances viewed during analysis. Our model also included an interaction between simulated vision loss and connecting cue conditions, since we predicted that connecting cue contrast would have an impact on participants' distance judgements in the central vision loss condition.

We anticipated that the experimental block order could have an impact on participants' performance and that this impact would vary across participants. or example, a participant that is prone to attentional lapses may behave differently than a focused participant on the first trial block. Accordingly, we included Order (3 levels: first, second, third) as a categorical main effect, together with by-subject random slopes for Order [36]. To account for individual variability in behavior over repeated measures, we also included a by-subject random intercept

We did not anticipate an interaction between simulated vision loss and the experiment block or-

der conditions (`Vision Loss x Order`). However, the presence of this interaction would have a pronounced impact on our interpretation of experimental results in this work. Of concern, the interaction between simulated vision loss and order could indicate that some aspect of our experimental design influenced experimental outcomes. As such, we conducted a model comparison between our planned LMM and the same model with `Vision Loss x Order` included. The `performance` package in R was used to compare models and produce a model performance score [288]. This score is calculated by normalizing all fit indices of a model (i.e., Akaike Information Criterion or AIC, Bayesian Information Criterion or BIC, conditional and marginal R2, ICC, Root Mean Square Error of Approximation or RMSE, and Bayes Factor or BF), taking their mean, and then returning a score from 0% to 100%. Higher scores indicate better model performance. Model comparison revealed the model with the interaction included performed no better than the original model without the interaction, with both receiving performance scores of 50%. Because model fit was not improved by including this interaction and because the `Vision Loss x Order` interaction reported in this model was not significant, we opted not to include this interaction term in our final model.

Response Times

To answer our research questions on the relationship between visual field loss and response times (RTs), we used a generalized linear mixed effects model (GLMM). Typically, RT distributions have a bell-like shape with a tail on the right side of the distribution [18, 499]. They also often have a minimum boundary on the left side of the distribution, since a small amount of time is required to encode a stimulus to execute a response [18].

The Gamma distribution is one of the common distributions used to describe response time data [18, 27, 309], since prior research has demonstrated that RT distributions are not Gaussian [309, 385]. For Gamma distributions, values are strictly positive and the variance of the distribution increases proportionally with the mean. In this work, we specified a Gamma distribution as an appropriate fit for our data. A density plot of our sampled response time data can be found in Figure 10.4 (right).

For our generalized linear mixed model (GLMM), we used the `glmer` function from the `lme4` package [40]. We specified a Gamma distribution and the identity link function for our model. The fixed and random factors for our model were similar to the LMM's described in the previous section. We employed simple contrast coding scheme for comparisons, and we conducted planned contrasts with estimated marginal means. Bonferroni correction was applied to account for multiple comparisons. The formula for our GLMM is expressed through `glmer` syntax in Equation 10.2 as follows:

$$\begin{aligned}\Upsilon = RT \sim\ & Distance + Vision + Shadow + Order + Vision : Distance \\ & + Vision : Shadow + (1 + Order|Id)\end{aligned} \tag{10.2}$$

We, again, did not anticipate an interaction between vision loss and the experiment block order conditions (`Vision Loss x Order`). However, to ensure that order did not have a pronounced effect on our experimental results related to simulated vision loss, we conducted a model comparison between our hypothesized model and the same model with the `Vision Loss x Order` interaction included using the `performance` package in R [288]. Model comparison revealed the model with the interaction included performed worse (25%) than the original model without the interaction (75%). Because model fit was not improved by including this interaction and because the `Vision Loss x Order` interaction

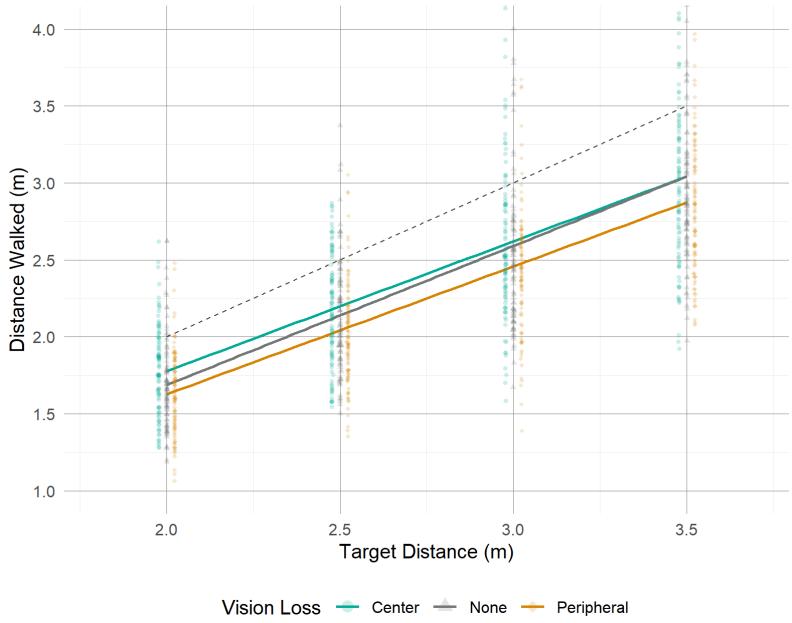

Figure 10.5: The solid lines show the average walked distance for each vision loss condition at the target distances. The dotted line represents veridical performance and individual points show the raw walked distance

reported in this model was not significant, we opted not to include this interaction term in our final model.

10.3.6 Results

Distance Judgments

Participants' distance judgments were recorded and statistically analyzed in meters. However, to facilitate comparisons between the current work and prior research, we also provide the error ratios of people's distance judgements in the discussion of our results. To create error ratios, participants' walked distances were divided by the actual distances to the target for each trial. A ratio less than 1.0 indicates distance underestimation, and a ratio greater than 1.0 indicates overestimation.

A summary of the results of our statistical analysis of walked distances data in meters are summarized in Table 10.1. Overall, we found that participants underestimated distances to targets in our study. The mean walked distance across all experimental conditions and participants was 2.34m ($SD = .66, Min = 1.06, Max = 5.23$). An average distance judgement of 2.34m in our experiment corresponds to a distance ratio of 0.851, or 14.9%, underestimation. This amount of underestimation is in line with what has been reported in prior distance estimation studies conducted in VR [63, 121, 224]

Distance As shown in the positive slopes of the regression lines in Figure 10.5, participants increased their egocentric distance judgements, as measured through blind walking, to virtual targets as the actual distance to the targets increased. This outcome is supported by a significant main effect of distance ($B = 0.86, t = 31.29, p < 0.001$) in our model. The coefficient for distance indicates that

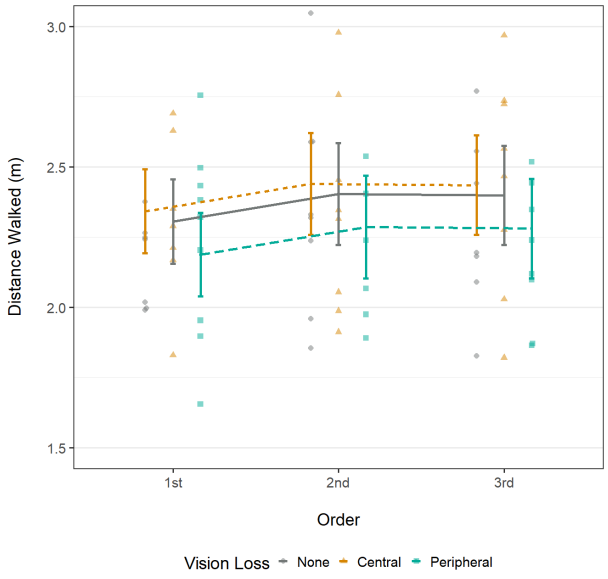

Figure 10.6: Estimated marginal means of walked distances for each Vision Loss condition by Order

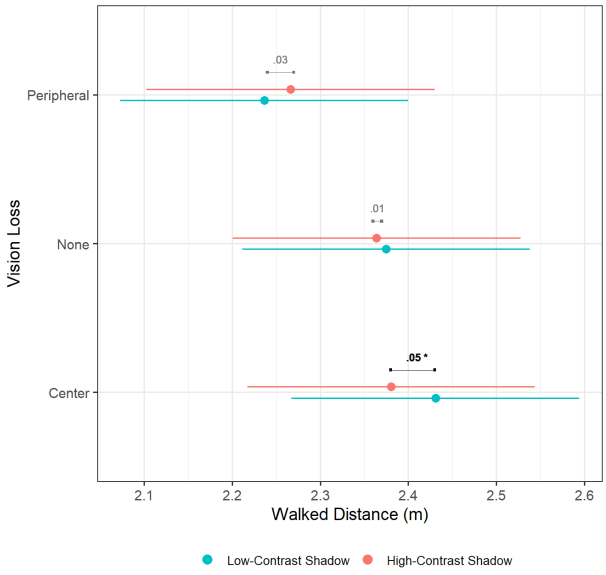

Figure 10.7: Results of Vision Loss × Shadow Contrast interactions for walked distance

participants estimated distances to be approximately 0.86m farther for every meter of increase in actual distance to the target, on average.

Order As for main effects of order in our planned analyses, there were no significant difference in distance judgments between experimental blocks after correction for multiple comparisons. However, there was a trend in which distance judgments were somewhat shorter in the first block of trials compared to the second ($B = -0.10, t = -2.02, p = 0.17$) and the third block of trials ($B = -0.09, t = -2.07, p = 0.15$). Coefficient estimates indicate that people's distance judgments were approximately 10cm shorter in the first block of trials. In comparison, the difference in distance judgements between the second and third blocks was quite small, only 1cm, on average ($-0.01, t = -.63, p = 1.00$). Distance judgements for each of the vision loss conditions by experimental order are visualized in Figure 10.6.

As discussed in Philbeck et al. [371], it is important to consider order effects in experimental analyses for blind walking, given that participants may grow in confidence when walking blindly over the course of an experiment. Accordingly, although the differences did not meet significance, it is worthwhile to note the trends in people's performance based on order in this work. A positive takeaway from our analysis of order is that there is little evidence of fatigue bias over the duration of the experiment.

Shadow Contrast There was no main effect of shadow contrast on people's distance judgements ($B = -.01, t = -.74, p = .33$). This outcome is unsurprising, given that we developed the high contrast shading condition only for the center vision loss condition. Further, based on the results of the prior research studies conducted in this dissertation, we did not anticipate that stylized graphics would perturb distance perception judgements outside of the center vision loss condition.

Table 10.1: Estimated coefficients for the linear mixed-effects model (LMM) predicting **walked distance** from fixed and random effects. *Estimates* represent the regression coefficients, *CI* is the confidence interval of the regression coefficient. The *Intercept* value is the grand mean of walked distances. Negative values for *Estimates* indicate that walked distances were farther for the first factor in the comparison, whereas positive values indicate that walked distances were farther for the second factor. Simple slopes were calculated for Vision Loss × Distance. Planned contrasts on the estimated marginal means were evaluated for Vision Loss × Shadow Contact. Significance tests adjusted for multiple comparisons with Bonferroni correction.

Predictors		*Estimates*	*CI*	*t-value*	*p-value*
(Intercept)		2.34 ***	[2.19 2.49]	31.29	<.001
Distance		.86 ***	[.83 .88]	68.98	<.001
Shadow Contrast	high - low	.01	[-.02 .04]	-.74	.33
Vision Loss	none - center	.04	[-.04 .11]	.95	.46
	none - peripheral	-.12 *	[-.19 -.04]	-3.06	<.05
	center - peripheral	-.15 **	[-.23 -.08]	-3.98	<.01
Order	1st - 2nd	.10	[.00 .19]	2.02	.17
	1st - 3rd	.09	[.00 .18]	2.07	.15
	2nd - 3rd	-.01	[-.07 .06]	-.63	1.00

Random Effects	*Variance*	*Corr*
ICC	.72	
σ^2	.06	
$\tau_{00\ Id}$	.13	
$\tau_{11\ Id.2nd}$	.05	.41
$\tau_{11\ Id.3rd}$	.04	.41

Interactions		*Estimates*	*CI*	*t-ratio*	*p-value*
Vision Loss × Distance					
	none - center | distance	.05	[-.03 .12]	1.55	.36
	none - peripheral | distance	.07	[-.01 .14]	2.18	.09
	center - peripheral | distance	.02	[-.05 .09]	.63	1.00
Vision Loss × Shadow Contrast					
	high - low | none	.01	[-.04 .06]	.44	.66
	high - low | center	.05 *	[.00 .10]	2.09	<.05
	high - low | peripheral	-.03	[-.08 .02]	-1.25	.21

Model Fit	
Marginal R^2	.55
Conditional R^2	.87

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Vision Loss Our hypothesis (**H1**), at least for the main effect of vision loss condition, was rejected. Instead, our results indicated that distance judgements were significantly different for the peripheral vision loss condition. Planned contrasts, when collapsed across shadow and order, showed that participants underestimated distances significantly more in the peripheral vision loss condition compared to either the no vision loss ($B = -.12, t = -3.06, p < .05$) or the central vision loss ($B = -.15, t = -3.98, p < .01$) conditions. We did not find a significant difference between no vision loss and center vision loss ($B = -.04, t = -.95, p = .46$). This relationship is demonstrated visually by the PVL condition having the lowest values of the regression lines featured in Figure 10.5.

On average, with no vision loss, participants walked 2.37m ($SD = .70, Min = 1.19, Max = 5.23, Ratio = .86$). With center vision loss participants walked 2.41m ($SD = .66, Min = 1.28, Max = 4.46, Ratio = .88$) and with peripheral vision loss participants walked 2.25m ($SD = .61, Min = 1.06, Max = 4.62, Ratio = .82$).

Vision Loss×Shadow Contrast We anticipated that the high contrast shadow connecting cues, specifically, would be beneficial to individuals with central vision loss. To be precise, we predicted that,

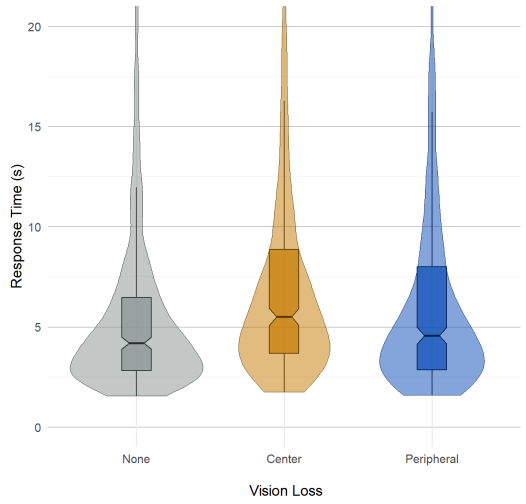

Figure 10.8: Violin plots of response times for each vision condition

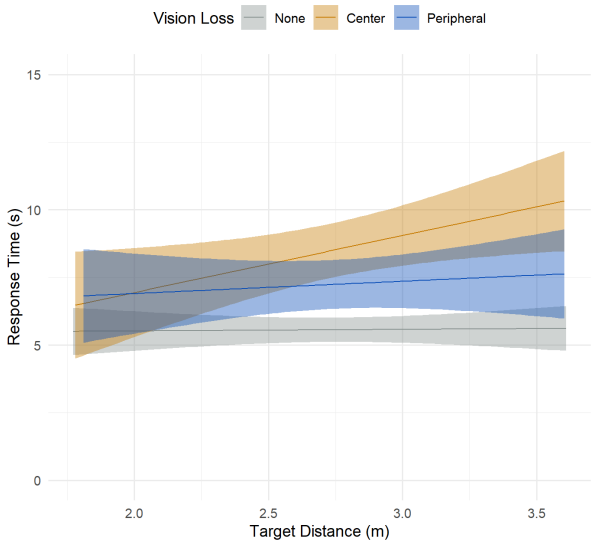

Figure 10.9: Simple slopes of Vision Loss × Distance with confidence intervals for response times

for people with central vision loss, depth judgements made to targets with low-contrast shadows would be overestimated relative to depth judgments made to targets with high-contrast shadows (**H2**). This hypothesis was confirmed.

Planned contrasts, collapsed across order, revealed a significant difference in people's distance judgments between the shadow conditions—but only for the central vision loss condition ($B = .05, t = -2.09, p < .05$). The relationship between vision loss condition and shadow contrast are visualized in Figure 10.7. On average, participants with central vision loss walked 2.43m ($SD = .67, Ratio = .884$) in the low-contrast condition and 2.38m ($SD = 0.64, Ratio = .866$) in the high contrast condition. This effect was small but significant, as indicated by the value of the fixed effect estimate ($B = .05$).

Vision Loss×Distance Analysis of the vision loss condition by distance simple slopes did not indicate significant differences between vision loss conditions (See Table 10.1. As such, we find no evidence in support of or against target distances having a differential impact across vision loss conditions for distance judgments in the range of distances evaluated for this study.

Response Time

To address our experimental hypotheses (**H3**, in particular) and to support our primary measure of interest, we analyzed participant response time (RT) data for each blind walking distance estimation trial. A summary of the results from the generalized linear mixed model (GLMM) that was fitted to response time data is shown in Table 10.2.

When we discuss RTs, we report both mean and median RT values in text, since the median is a more representative measure of central tendency given the strong right skew of the RT data. Overall, The median response time from participants was 4.73s ($Mean = 7.11, SD = 8.31, Min = 1.56, Max = 122.00$).

Table 10.2: Estimated coefficients and confidence intervals for the GLMM with a Gamma distribution (identity-link) predicting **response time** from fixed and random effects. Planned contrasts on the estimated marginal means were evaluated for Vision Loss ×Contact. Simple slopes were calculated for Vision Loss × Distance. Significance tests adjusted for multiple comparisons with Bonferroni correction.

Predictors		*Estimates*	*CI*	*t-value*	*p-value*	**Random Effects**	*Variance*	*Corr*
(Intercept)		7.23 ***	[5.79 8.68]	9.84	<.001	*ICC*	.95	
Distance		.54 **	[.22 .86]	3.30	<.01	σ^2	.61	
Shadow Contrast	high - low	.03	[-.31 .37]	.17	.864	$\tau_{00\ Id}$	8.32	
						$\tau_{11\ Id.2nd}$	9.22	-.57
Vision Loss	none - center	2.58 ***	[1.48 3.67]	4.63	<.001	$\tau_{11\ Id.3rd}$	6.86	-.45
	none - peripheral	.93	[-.19 2.04]	1.63	.309			
	center - peripheral	-1.65 **	[-2.71 -.59]	-3.06	<.01			
Order	1st - 2nd	-2.07 *	[-3.62 -.53]	-2.64	<.05			
	1st - 3rd	-2.16 **	[-3.51 -.81]	-3.15	<.01			
	2nd - 3rd	-.09	[-1.07 .89]	-.18	1.00			

Interactions		*Estimates*	*CI*	*z-ratio*	*p-value*	**Model Fit**	
Vision Loss × Distance							
	none - center | distance	1.44 **	[-.43 2.45]	3.41	<.01	Marginal R^2	.18
	none - peripheral | distance	.57	[-.29 1.42]	1.58	.345	Conditional R^2	.96
	center - peripheral | distance	-.87	[-1.88 .13]	-2.09	.110		
Vision Loss × Shadow Contrast							
	high - low | none	-.36	[-.90 .19]	-1.28	.201		
	high - low | center	.52	[-.19 1.22]	1.4	.150		
	high - low | peripheral	-.05	[-.59 .49]	-.18	.856		

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Distance Response times increased as the distance to targets increased, which is indicated by a main effect of distance in our GLMM ($B = .24, t = 2.72, p < .01$). This outcome is interesting, because our response time measure reflects the time it took for participants to determine where the target was positioned immediately before blind walking. We excluded actual walking time, since this would bias response times for the farther targets. Despite this, we still found a main effect of distance for response time. A significant, positive effect of distance for RTs implies that participants required more time to process visual stimuli when targets were positioned farther away.

When we evaluate the median response times at each difference, we can observe this positive trend, as well. The median RT was 4.13s ($Mean = 5.83, SD = 5.36, Min = 1.56, Max = 53.19$) at 2.0m, 4.81s ($Mean = 7.86, SD = 11.68, Min = 1.70, Max = 122.00$) at 2.5m, 5.02s ($Mean = 7.04, SD = 6.44, Min = 1.67, Max = 51.19$) at 3.0m, and 5.12s ($Mean = 7.70, SD = 8.26, Min = 1.60, Max = 80.66$) at 3.5m.

Order For order effects, we found that RTs were slower in the first block of trials. This outcome provides some evidence of familiarization with the experimental task during the first block of trials. The median response time was 5.27s ($Mean = 8.64, SD = 9.73, Min = 1.78, Max = 122.00$) in the first block versus 4.33s ($Mean = 6.35, SD = 6.30, Min = 1.56, Max = 73.27$) in the second and 4.28s ($Mean = 6.31, SD = 8.31, Min = 1.67, Max = 82.47$) in the third experimental blocks. Post-hoc pairwise comparisons, indicated that only differences between the first and second block ($B = 2.18, z =$

2.76, $p < .05$) and the first and third block ($B = 2.20, z = 3.16, p < .01$) of trials were significant. The difference between the second and third blocks ($B = .03, z = .05, p = 1.00$) was not.

Shadow Contrast There was no significant main effect of shadow contrast condition on people's response times ($B = .03, z = .17, p = .864$). The median RT for the low-contrast shadow condition was 4.73s ($Mean = 7.18, SD = 9.19, Min = 1.56, Max = 122.00$), and the median RT for the high-contrast shadow condition was 4.74s ($Mean = 7.04, SD = 7.34, Min = 1.60, Max = 82.47$).

Vision Loss People with vision impairments are often able to accomplish tasks as well as people without vision impairments. However, it can take them longer to do so. Accordingly, we predicted that RTs would be longer in the vision loss conditions than in the no vision loss condition (**H3**). This hypothesis was partially supported. Violin plots of people's RTs for each of the vision conditions are shown in Figure 10.8

Planned contrasts, collapsed across shadow and order, showed that response times were significantly longer for people in the CVL condition than both the no vision loss condition ($B = 2.51, z = 4.42, p < .001$) and the peripheral condition ($B = 1.60, z = 2.90, p < .05$). There was no significant difference between the peripheral vision loss and no vision loss conditions ($B = -.912, z = -1.55, p = .27$). The median RT for the central vision loss was 5.50s ($Mean = 8.49, SD = 10.01, Min = 1.76, Max = 82.47$). The median RTs for the no vision loss and peripheral vision loss conditions were 4.19s ($Mean = 5.63, SD = 4.40, Min = 1.56, Max = 30.70$) and 4.57s ($Mean = 7.24, SD = 9.01, Min = 1.60, Max = 122.00$).

Vision Loss×Shadow Contrast There was no significant interaction between vision loss condition and shadow contrast condition for response times (See Table 10.2).

Vision Loss×Distance We report an interaction between vision loss condition and distance in which, as distances increased, the RTs for the center vision loss condition increased at a faster rate than RTs increased for the no vision loss condition ($B = -1.44, z = -3.41, p < .01$). Interaction effects between the other vision groups did not reach significance (See Table 10.2). For the CVL condition in Figure 10.9, we can observe a positive slope in the regression lines for the plots of raw RT data across distances.

10.3.7 Discussion

The outcomes of our experiment have ramification for our understanding of the visual information that people attend to when making depth perception judgments. And they have implications for individuals with peripheral and central visual field loss.

Given that this experiment was conducted in the HTC Vive Pro Eye, we expected people to underestimate distances, regardless of any experimental manipulations. The amount of underestimation we found (24.2%) is less than the amount reported in our previous distance estimation experiments that were conducted with the Varjo XR-3 through verbal report (Exp 1: 20.2%, Exp2: 29%) [4]. It is also less than the amount of underestimation Buck et al. [63] reported for the HTC Vive Pro with blind walking estimation protocol (34%). However, more accurate distance estimation is not a surprising outcome, given that we evaluated notably closer distances.

For example, the amount of underestimation we find in the current work is quite similar to the 13.2% reported by Kelly et al. [224] in the HTC Vive for targets positioned at 1-5m. In contrast, in our prior work target distances ranged between 3-6m, and target distances were positioned even farther (5-10m) in Buck et al. [63]. In addition, the Varjo XR-3 display that we used in our previous experiment weighs more than the HTC Vive Pro, which may have further contributed to compressed distance judgments in that device [4, 501].

10.3.8 Peripheral vision loss exacerbates distance underestimation

In this experiment, we found that people with peripheral vision loss (PVL) underestimated distances more than any other vision condition. If we look at the results of prior VR display research, several studies have demonstrated that a restricted FOV in these displays (likely in conjunction with their weight [501]) contributes to distance compression [64, 303, 501]. These experiments were almost exclusively conducted with complete occlusion of peripheral vision, since they were modeled to simulate the properties of an HMD. However, there are a few notable exceptions that can provide additional context to the current work.

For example, Jones et al. [212] demonstrated that applying constant white light to the periphery, distance underestimation in VR could be partially mitigated. In followup investigations, Jones et al. [214], compared distance judgements made with peripheral light stimulation in a restricted FOV condition ($47° \times 40°$) to distance judgments made when a larger FOV was provided ($150° \times 85°$) in a virtual environment and reported no significant difference in people's distance judgements. Li and colleagues [272, 273] then probed into this phenomenon further by comparing the effect of different levels of luminance along the boarders of an HMD's screen on distance judgements. The center of the display ($47° \times 40°$) was rendered normally, and the remaining FOV ($110° \times 90°$) was rendered with a grayscale color value. They found that a certain threshold of brightness was required to evoke improvements in distance judgements. In the third experiment conducted by Li et al. [273], the authors compared distance judgements made to targets when the area in the remaining FOV was rendered as black to distance judgements made when this region was rendered with a downsampled, pixelated image. They once again found that peripheral stimulation, this time with a pixelated image, improved distance judgements.

The PVL condition in this work differs from those in the aforementioned experiments in a few key aspects. First, the loss of peripheral vision is applied through a radial gradient in our vision simulation. In contrast, the peripheral vision manipulations employed by the aforementioned works introduced abrupt changes in vision along a straight edge within or beyond the display screen. Even the pixelated treatment provided by Li et al. [273] consisted of sharp edges brought on by the large blocks of the pixelation method. Peripheral vision is quite sensitive to edge and motion detection. It is possible that the peripheral stimulation effects demonstrated by these prior studies were effective in part because they added sharp, high-contrast stimulation to the periphery—not just the presence of light, as argued. In contrast, our method provides light stimulation, but contrast is reduced due to Gaussian blurring.

It is possible that the gradual onset of vision loss coupled with the reduction of contrast at the periphery discouraged participants from visually scanning their environment to gather the spatial information that they needed to make more accurate distance judgements. Wu et al. [510] demonstrated that even when FOV is restricted, an observer can accurately judge absolute distances in the real world by scanning local patches of the ground surface in the environment. Similarly, other experiments in the real world have demonstrated that even when FOV is restricted, people can make accurate distance

judgements when head movements are not restricted [91, 233, 510]. If participants did not perform visual scanning to compensate for their vision loss, then this could also explain why the PVL condition in our study exhibited more underestimation than our other conditions. In the final publication for this work, it may be fruitful to look at gaze data collected throughout the experiment to better understand how differences in visual scanning behavior may have impacted people's behavior in this experiment.

A second major difference between this work and the previous studies is that peripheral vision loss in our study was eye-tracked. In all the aforementioned experiments, whether they were conducted in VR or in a real environment, FOV restrictions were not tied to a participant's gaze direction. They were tied to a participant's heading. It is highly likely that behaviors induced by FOV restrictions based on gaze direction will differ from those induced by FOV restrictions based on head direction, especially with regard to behaviors like visually scanning.

Another argument could be made that the difference in our result is due to the FOV restriction employed in this work being simply smaller than what was evaluated in the peripheral stimulation studies. Perhaps the peripheral vision restriction was too severe for positive effects of peripheral stimulation to work as expected. One might also consider that the egocentric distance estimation task, in which targets were always presented immediately in front of the viewer, naturally discouraged participants from visually scanning the environment since they did not need to look around the environment to locate the target. This theory is less compelling, given that the peripheral stimulation studies employed the same blind walking protocol [212, 214, 272, 273].

Regardless, the current work provides interesting new insights into the relationship between peripheral vision and spatial perception in HMDs. It also provides inspiration for a number of potential avenues for future investigations that may further elucidate their relationship.

As for the implications of our work for low vision patients with severe peripheral vision loss, this work in conjunction with prior research investigating peripheral vision restrictions indicate that peripheral vision is quite important for spatial perception. The importance of peripheral vision has been demonstrated in a number of domains related to spatial perception in prior research, as well. For example, peripheral vision is important for understanding the layout of an environment [259].

However, the results of research that has evaluated depth perception judgements in real patients with peripheral vision loss are less clear. In studies where PVL patients are asked to make depth judgments and to make fine motor controlled behaviors, patients with PVL often perform better than patients with CVL [355, 470]. However, some of this research also reports evidence that size and scale at near-distances may be adversely affected by PVL [238, 431]. In the current work, people were least accurate in the peripheral vision loss condition. The differences in experimental outcomes between simulated PVL studies and studies conducted with real patients with PVL are important to note. It may be that these differences arise due to the evaluated distances.While much of the simulated work has been conducted in action space (distances beyond 2m), research conducted with real PVL patients have been conducted in near-space (distances less than 2m). In future research, it may be beneficial to consider evaluating closer distances in order to draw better comparisons between simulated and real PVL studies.

10.3.9 Stylized shadows impact people with central vision loss

We found that distance judgments made to targets with realistic, low-contrast shadows were overestimated relative to judgments made to targets with stylized, high-contrast shadows in the center vision loss (CVL) condition.

Like much of the work in this dissertation, our investigation into the influence of shadow shading in this work was motivated by the importance of surface contact for spatial perception. In the case of center vision loss, the relationship between floating objects and the surface beneath them can become obscured. To compensate for reduced vision in the center of the visual field for participants in the CVL condition, we included high contrast cast shadows in our experiment.

Recall the 5th experiment of this dissertation, in which we evaluated distance perception judgements to floating targets with and without the presence of cast shadows. In this experiment, participants in the no shadow condition perceived floating targets as farther away, because in the absence of a surface contact cue people rely on optical contact to determine the position of objects in space [4]—where optical contact refers to the location at which the projected image of an object contacts the image of the ground beneath it [314, 332, 352]. In essence, we anticipated that a similar effect would occur when central vision was impaired by the vision simulation. This idea is motivated by prior research by Rand et al. [383, 384] in which a similar effect was found when surface contact cues (e.g., an object placed on a stand) when global reductions in visual acuity obscured the appearance of the stand.

We found our predicted effect in this study. The high-contrast shadow cue anchored people's distance judgements in the CVL condition. As a result, their distance judgements in this condition were similar to the distance judgements made without any vision loss. This can be seen visually in Figure 10.7. However, the effect of shadow contrast for CVL was small. This is likely explained by the small vertical offset used for targets in this study. We anticipate that a more pronounced effect of high-contrast graphics would be found in different contexts.

For example, prior studies with real patients have found that people with CVL often perform worse than people with PVL [356, 431] and age-matched controls [470, 482] at near-distance tasks that require fine motor control. In this context, it is likely that visual aids like the ones we present in the current work would be helpful. Moving forward, it may be worthwhile to consider how visual accessibility aids can be modified based on the size of a central scotoma, as well. For example, the size of the stylized cast shadows featured in the current work may not be as beneficial for a patient with a larger central scotoma. Given the heterogeneity of scotoma shapes and sizes in real patients, it may be beneficial to study the relationship between the size of a scotoma and the difficulty of interpreting the depth of objects in space based on the size of the retinal image of those objects.

The impact of shadow shading in this experiment is somewhat modest, but this small effect opens up the possibility to pursue more ambitious visualizations to enhance spatial perception. And this first study provides the groundwork for future investigations into the influence of central vision loss on spatial perception.

10.3.10 Central vision loss increases response times

Another interesting finding from this work was that people's response times (RTs) were longer in the central vision loss condition. Further, participants' RTs in this condition increased with target distance. This finding reinforces the importance of acknowledging the relationship between the size of a center scotoma and the size of an object on the retinal image of the human eye.

The size of an object on the retinal image of the human eye decreases as the distance of the object from the viewer increases [147]. However, the size of the simulated central scotoma in the CVL condition of our study remained constant. As a result, in the CVL condition, participants required more time to make distance judgements when the target was positioned farther away. Figure 10.9 shows the increased response times for the CVL condition as distances increased.

10.3.11 Limitations

Of course, there are a number of limitations to this study. First, the data quality of the eye tracker and the resolution of the HTC Vive Pro Eye are not ideal for true vision simulation. It is also possible that the presence of a tether for the HMD exacerbated distance underestimation in this experiment, since the HTC Vive Pro Eye was connected to a computer. Although prior research has found blind walking assessments of distance perception can be less susceptible to the influence of environmental context [248], it is also possible that properties of the environmental context provided by the environment could have influenced our results [303]. It may be worthwhile to conduct assessments across varied virtual environments in future work to ensure results are generalizable.

The adaptation phase for acclimating to each vision loss condition was only two minutes long. The compensatory strategies developed by participants to make depth judgements are likely different from those used by patients with similar forms of vision loss who have developed compensatory strategies over time, often years. Our research has interesting implications for the impact of visual field loss on depth judgements, but we cannot assume that this work provides a perfect parallel for understanding perception and behavior of real patients.

In future research, it may be worthwhile to investigate the relationship between target size, in degrees of field of view, and the extent of visual field loss on depth perception judgments. One may expect that as the central scotoma occludes more visual information and environmental context, that the impact of central vision loss on depth perception may become more pronounced.

Adaptations to Simulated Vision Loss

Vision simulations become particularly useful as a design tool if they can introduce functional impairments similar to the ones experienced by people with low vision. Prior research on desktop simulation of low vision has shown the potential of vision simulation for studying the behaviors of people with low vision. This research provides evidence that, at least in simple 2D tasks, visual impairment simulations can introduce functional impairments in normally-sighted individuals that mimic the impairments of people with visual field loss [85, 387, 476, 488]. When given a simulated central scotoma, people even develop similar compensatory vision behaviors to those seen in real low vision patients [85, 488]. Specifically, people with simulated central scotomas learn to rely on areas of vision outside of a central scotoma, using regions termed preferred retinal loci (PRL), to improve performance on visual tasks like object identification [85, 488] and reading [387, 476].

However, such adaptations do not occur spontaneously and, indeed, the adaptations that occurred in the aforementioned studies developed after hours of exposure. It is likely that functional impairments may be exaggerated in low vision simulations, since users will only temporarily experience vision impairment. Patients with low vision adapt their behaviors over time, often months, to compensate for gradual vision loss [67, 411]. In contrast, someone with a simulated scotoma experiences a rapid onset of vision loss. Because of this, the compensatory strategies between real and simulated low vision people may differ [17, 476, 488]. Unfortunately, the nature of these differences is unclear due to the lack of prior research investigating differences between real and simulated low vision. Such investigations will be worthwhile to pursue in future research to better determine the efficacy of vision simulation's ability to reproduce behaviors of real patients with low vision.

10.4 Summary

In this chapter, we finally extended our research findings on participants with normal vision to those with vision impairments via a simulation study. The findings of our investigation suggest that people with peripheral vision loss (PVL) underestimated distances more than people with central or no vision loss in action space, possibly due to limitations in scanning the environment or using the visual horizon. Contrary to expectations, central vision loss (CVL) did not significantly impact distance judgments. This outcome may be related to the large size (20cm diameter) of the target object. Longer response times in the CVL condition, suggests that participants employed different viewing strategies to determine the position of a target in space compared to the other vision loss conditions. It is likely that this reflects the use of visual scanning to make use of intact peripheral vision to view the target. Analyzing gaze data could shed light on these compensatory strategies.

In a similar manner to what we found in our previous experiments, in general we did not observe significant differences in depth perception judgements between the realistic, low-contrast shadow connecting cue and the stylized high-contrast connecting cue. However, we did find the predicted interaction between vision condition and shadow contrast in which participants with CVL overestimated depth judgments to targets with low-contrast shadows compared to depth judgments to targets with high-contrast connecting cues.

To our knowledge, the current work is the first to look at direct measures of distance perception in action space for people without central vision. We believe our vision simulation and our approach to better understanding the impact of visual field loss on spatial perception provided the necessary groundwork for future research that seeks to understand how visual information impacts people's spatial perception.

Part VII

Closing

Finish what you started, human
Kamaji, Spirited Away (2001)

11 Conclusions

This dissertation investigated how visual design choices in mixed reality (MR) displays influence depth perception. Our investigations are particularly impactful for MR, given the large amount of research that has demonstrated that spatial perception is distorted in head-mounted displays (HMDs). In addition, our research allows for the expansion of MR depth perception research to begin expanding to a broader, more inclusive audience through the development of a vision simulation that can be used for the controlled study of depth perception in people with vision loss.

Our work was initially inspired by theories on depth cue integration in augmented reality (AR). Previously, researchers have suggested that graphical realism was crucial for mitigating these distortions in AR. However, our investigations in this research on realistic vs. non-photorealistic shading for cast shadows (a vital depth cue) revealed that realistic graphics may not always be necessary for good spatial perception in MR (Aim 1).

Building on this initial finding, subsequent studies (Aim 2) explored the influence of realistic and non-photorealistic rendering approaches on depth perception across various MR displays with different optical and graphical features. This broader exploration aimed to understand how display properties and shadow shading methods impact depth perception. Aim 2 specifically focused on absolute egocentric distance judgments in AR, extending the investigation to another common depth perception measure for broader understanding and to better draw connections between our investigations and existing research. Accessible AR interfaces often employ object highlights and edge enhancements to improve contrast, and we hypothesized that non-photorealistic graphics similar to those evaluated in Aims 1 & 2, by increasing the shadow's contrast against the background, might benefit depth perception in AR for individuals with low vision.

This research also examined how different shadow rendering approaches could enhance distance perception for objects positioned above the ground, a particularly relevant aspect for AR user interfaces often located mid-air. Understanding perception of mid-air virtual objects is crucial for user interaction in AR, especially considering the prevalence of AR interfaces that position content above the ground for easy viewing.

However, prior research has shown that individuals with severe visual impairments face substantial challenges in determining the position of such objects [383, 384]. To test this hypothesis and to provide other people a valuable tool for future research and development in the absence of real test users, Aim 3 proposed the development of a data-driven low vision simulation with eye-tracking. We then applied this vision simulation to better understand how central vision loss and peripheral vision loss impact depth perception and to determine if the non-photorealistic shading approaches for cast shadows would be beneficial for individuals with vision loss.

Overall, this dissertation contributes valuable insights into how both normally sighted and visually

impaired users perceive virtual objects in mixed reality. Furthermore, it lays important groundwork for future research on depth perception and accessibility in AR. The main contributions of this dissertation are as follows:

Challenging the Need for Photorealism

The first two aims of this dissertation (Aim 1 and Aim 2) challenged the assumption that photorealistic rendering techniques are necessary for accurate depth perception in MR. These studies demonstrated that people's ability to perceive both surface contact and egocentric distances was actually improved when virtual objects were rendered with non-photorealistic shadows compared to photorealistic versions. This finding holds significant promise for two audiences: developers working within the limitations of contemporary MR displays and those working in the field of low vision accessibility, where applications often employ non-photorealistic graphical elements to enhance object visibility.

The Benefits of Shadows, Regardless of Realism

Aim 2 also established that the presence of shadows, regardless of whether they are photorealistic or not, improves depth judgments for virtual objects regardless of their position (on or above the ground). However, the study did not determine if this relationship remains consistent when objects are placed at greater heights, such as eye-level. This finding offers developers a simple and empirically proven approach for improving user perception of objects in space within MR environments. Additionally, it helps to fill existing gaps in the literature regarding the roles of virtual shadows and object height in AR depth perception.

The Development of Vision Simulation Accessibility Tool

The final aim of this dissertation (Aim 3) focused on the development of a data-driven low vision simulation testbed integrated with eye-tracking. This tool aims to empower MR developers to design for individuals with low vision without requiring access to real-world test participants. We provided valuable insight into the status of modern eye-tracked HMDs through a data quality assessment, and we demonstrated how understanding data quality is imperative for the design of applications and experiments that employ eye tracking.

Extending our Findings to People with Vision Loss

Our final experiment bridged the gap in research between the fields of spatial perception and accessibility, and it lays a strong foundation for future research endeavors at the intersections of human perception and behavior and accessibility. In addition, we revealed differential impacts of central and peripheral vision loss on depth perception in action space. And we demonstrated that high contrast surface contact cues could be used to improve depth judgments for people with central vision loss.

Future Directions and Significance

This dissertation contributes significantly to our understanding of how visual design choices in MR displays impact depth perception for both normally sighted and visually impaired users. Additionally,

the proposed low vision simulation has the potential to accelerate accessibility advancements by allowing developers to design and test applications without requiring real-world test participants with low vision. Overall, this research paves the way for more inclusive and accessible MR experiences for a broader audience.

References

[1] H. Adams. Resolving cue conflicts in augmented reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 547–548. IEEE, 2020.

[2] H. Adams, H. Gagnon, S. Creem-Regehr, J. Stefanucci, and B. Bodenheimer. Stay in touch! shape and shadow influence surface contact in xr displays. *IEEE TVCG [In Preparation]*, 2022.

[3] H. Adams, G. Narasimham, J. Rieser, S. Creem-Regehr, J. Stefanucci, and B. Bodenheimer. Locomotive recalibration and prism adaptation of children and teens in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1408–1417, Apr. 2018.

[4] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Depth perception in augmented reality: The effects of display, shadow, and position. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 792–801. IEEE, 2022.

[5] H. Adams, J. Stefanucci, S. H. Creem-Regehr, G. Pointon, W. B. Thompson, and B. Bodenheimer. Shedding light on cast shadows: An investigation of perceived ground contact in ar and vr. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[6] J.-g. Ahn, E. Ahn, S. Min, H. Choi, H. Kim, and G. J. Kim. Size perception of augmented objects by different ar displays. In C. Stephanidis, editor, *HCI International 2019 - Posters*, pages 337–344, Cham, 2019. Springer International Publishing.

[7] Z. Ai, B. K. Gupta, M. Rasmussen, Y. J. Lin, F. Dech, W. Panko, and J. C. Silverstein. Simulation of eye diseases in a virtual environment. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, page 5. IEEE Computer Society, 2000.

[8] M. Aitkin. Posterior bayes factors. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):111–128, 1991.

[9] B. P. Allen. Shadows as sources of cues for distance of shadow-casting objects. *Perceptual and Motor Skills*, 89(2):571–584, 1999. PMID: 10597594.

[10] U. Altangerel, G. L. Spaeth, and W. C. Steinmann. Assessment of function related to vision (afrev). *Ophthalmic epidemiology*, 13(1):67–80, 2006.

[11] D. G. Altman and J. M. Bland. Statistics notes: the normal distribution. *Bmj*, 310(6975):298, 1995.

[12] D. G. Altman and J. M. Bland. Statistics notes variables and parameters. *Bmj*, 318(7199):1667, 1999.

[13] J. Andre and S. Rogers. Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics*, 68(3):353–361, 2006.

[14] M. Antonov, N. Mitchell, A. Reisse, L. Cooper, S. LaValle, and M. Katsev. Oculus software development kit. oculus vr. *Inc., CA, USA*, 2013.

[15] C. Armbrüster, M. Wolter, T. Kuhlen, W. Spijkers, and B. Fimm. Depth perception in virtual reality: distance estimations in peri-and extrapersonal space. *Cyberpsychology & Behavior*, 11(1):9–15, 2008.

[16] R. Asaoka, D. P. Crabb, T. Yamashita, R. A. Russell, Y. X. Wang, and D. F. Garway-Heath. Patients have two eyes!: binocular versus better eye visual field indices. *Investigative ophthalmology & visual science*, 52(9):7007–7011, 2011.

[17] D. S. Asfaw, P. R. Jones, V. M. Mönter, N. D. Smith, and D. P. Crabb. Does glaucoma alter eye movements when viewing images of natural scenes? a between-eye study. *Investigative ophthalmology & visual science*, 59(8):3189–3198, 2018.

[18] F. G. Ashby and J. T. Townsend. Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, 21(2):93–123, 1980.

[19] D. H. Ashmead, D. Guth, R. S. Wall, R. G. Long, and P. E. Ponchillia. Street crossing by sighted and blind pedestrians at a modern roundabout. *Journal of Transportation Engineering*, 131(11):812–821, 2005.

[20] P. Aspinall, A. Hill, P. Nelson, C. O'Brien, E. O'Connell, L. McCloughan, A. Azuara-Blanco, R. Brice, S. Green, and C. Steeds. Quality of life in patients with glaucoma: a conjoint analysis approach. *Visual impairment research*, 7(1):13–26, 2005.

[21] H. C. Ates, A. Fiannaca, and E. Folmer. Immersive simulation of visual impairments using a wearable see-through display. In *Proceedings of the ninth international conference on tangible, embedded, and embodied interaction*, pages 225–228, 2015.

[22] S. Aziz and O. Komogortsev. An assessment of the eye tracking signal quality captured in the hololens 2. pages 1–6, 06 2022.

[23] L. Azoulay, P. Chaumet-Riffaud, S. Jaron, C. Roux, S. Sancho, N. Berdugo, I. Audo, J.-A. Sahel, and S. Mohand-Saïd. Threshold levels of visual field and acuity loss related to significant decreases in the quality of life and emotional states of patients with retinitis pigmentosa. *Ophthalmic research*, 54(2):78–84, 2015.

[24] R. T. Azuma. A survey of augmented reality. *Presence: Teleoper. Virtual Environ.*, 6(4):355–385, Aug. 1997.

[25] I. B. Adhanom, S. C. Lee, E. Folmer, and P. MacNeilage. Gazemetrics: An open-source tool for measuring the data quality of hmd-based eye trackers. In *ACM symposium on eye tracking research and applications*, pages 1–5, 2020.

[26] R. H. Baayen, D. J. Davidson, and D. M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):390–412, 2008.

[27] R. H. Baayen and P. Milin. Analyzing reaction times. *International journal of psychological research*, 3(2):12–28, 2010.

[28] R. J. Bailey, C. M. Grimm, and C. Davoli. The real effect of warm-cool colors. *Report Number: WUCSE-2006-17. All Computer Science and Engineering Research*, 2006.

[29] G. Ballestin, F. Solari, and M. Chessa. Perception and action in peripersonal space: A comparison between video and optical see-through augmented reality devices. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 184–189, 2018.

[30] D. Banks and R. J. McCrindle. Visual eye disease simulator. *Proceedings 7th ICDVRAT with ArtAbilitation*, 2008.

[31] M. S. Banks, J. C. A. Read, R. S. Allison, and S. J. Watt. Stereoscopy and the human visual system. *SMPTE Motion Imaging Journal*, 121(4):24–43, 2012.

[32] E. Barba, B. MacIntyre, and E. D. Mynatt. Here we are! where are we? locating mixed reality in the age of the smartphone. *Proceedings of the IEEE*, 100(4):929–936, 2012.

[33] E. M. Barhorst-Cates, K. M. Rand, and S. H. Creem-Regehr. The effects of restricted peripheral field-of-view on spatial learning while navigating. *PloS one*, 11(10):e0163785, 2016.

[34] E. M. Barhorst-Cates, K. M. Rand, and S. H. Creem-Regehr. Navigating with peripheral field loss in a museum: learning impairments due to environmental complexity. *Cognitive research: principles and implications*, 4(1):1–10, 2019.

[35] K. Bark, C. Tran, K. Fujimura, and V. Ng-Thow-Hing. Personal navi: Benefits of an augmented reality navigational aid using a see-thru 3d volumetric hud. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 1–8, 2014.

[36] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278, 2013.

[37] J. Barreira, M. Bessa, L. Barbosa, and L. Magalhães. A context-aware method for authentically simulating outdoors shadows for mobile augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 24(3):1223–1231, 2018.

[38] P. Barrouillet, S. Bernardin, S. Portrat, E. Vergauwe, and V. Camos. Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3):570, 2007.

[39] J. J. Barton, M. Benatar, and M. A. Samuels. *Field of vision: A manual and atlas of perimetry*. Humana Press, 2003.

[40] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[41] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 585–592, 2019.

[42] W. Becker and A. F. Fuchs. Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points. *Vision research*, 9(10):1247–1258, 1969.

[43] F. Behrens, M. MacKeben, and W. Schröder-Preikschat. An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters. *Behavior research methods*, 42(3):701–708, 2010.

[44] R. Bender and S. Lange. Adjusting for multiple testing—when and how? *Journal of clinical epidemiology*, 54(4):343–349, 2001.

[45] O. J. Berger. *Encyclopedia of Statistical Sciences*, chapter Bayes Factor, pages 378–386. Wiley, Hoboken, NJ, 2006.

[46] M. Bernhard, K. Grosse, and M. Wimmer. Bimodal task-facilitation in a virtual traffic scenario through spatialized sound rendering. *ACM Trans. Appl. Percept.*, 8(4):24:1–24:22, 2008.

[47] M. Berning, D. Kleinert, T. Riedel, and M. Beigl. A study of depth perception in hand-held augmented reality using autostereoscopic displays. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 93–98, 2014.

[48] Z. Bian, M. L. Braunstein, and G. J. Andersen. The ground dominance effect in the perception of 3–D layout. *Perception & Psychophysics*, 67(5):801–815, 2005.

[49] M. Billinghurst, A. Clark, and G. Lee. A survey of augmented reality. *Found. Trends Hum.-Comput. Interact.*, 8(2–3):73–272, Mar. 2015.

[50] J. M. Bland and D. G. Altman. Statistics notes: Transforming data. *Bmj*, 312(7033):770, 1996.

[51] P. Blignaut and T. Beelders. The precision of eye-trackers: a case for a new measure. In *Proceedings of the symposium on eye tracking research and applications*, pages 289–292, 2012.

[52] P. Blignaut and D. Wium. Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior research methods*, 46(1):67–80, 2014.

[53] T. M. Bochsler, G. E. Legge, R. Gage, and C. S. Kallie. Recognition of ramps and steps by people with low vision. *Investigative Ophthalmology & Visual Science*, 54(1):288–294, 2013.

[54] B. Bodenheimer, H. Adams, M. Whitaker, J. Stefanucci, and S. Creem-Regehr. Perceiving absolute distance in augmented reality displays with realistic and non-realistic shadows. In *ACM Symposium on Applied Perception 2023*, pages 1–9, 2023.

[55] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113(4):700, 2006.

[56] R. Bogdanova, P. Boulanger, and B. Zheng. Depth perception of surgeons in minimally invasive surgery. *Surgical innovation*, 23(5):515–524, 2016.

[57] Y. Boumenir, A. Kadri, N. Suire, C. Mury, and E. Klinger. Impact of simulated low vision on perception and action. *International Journal of Child Health and Human Development (IJCHD)*, 7(4):441–450, 2014.

[58] L. Brehm and P. M. Alday. Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125:104334, 2022.

[59] R. S. Brenton and C. D. Phelps. The normal visual field on the humphrey field analyzer. *Ophthalmologica*, 193(1-2):56–74, 1986.

[60] M. E. Brooks, K. Kristensen, K. J. Van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Machler, and B. M. Bolker. glmmtmb balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, 9(2):378–400, 2017.

[61] M. M. Brown, G. C. Brown, S. Sharma, A. F. Smith, and J. Landy. A utility analysis correlation with visual acuity: methodologies and vision in the better and poorer eyes. *International ophthalmology*, 24(3):123–127, 2001.

[62] S. Bruckner and E. Gröller. Enhancing depth-perception with flexible volumetric halos. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1344–1351, Nov 2007.

[63] L. Buck, R. Paris, and B. Bodenheimer. Distance compression in the htc vive pro: A quick revisitation of resolution. *Frontiers in Virtual Reality*, 2:728667, 2021.

[64] L. E. Buck, M. K. Young, and B. Bodenheimer. A comparison of distance estimation in hmd-based virtual environments with different hmd-based conditions. *ACM Trans. Appl. Percept.*, 15(3):21:1–21:15, July 2018.

[65] H. H. Bülthoff and H. A. van Veen. Vision and action in virtual environments: Modern psychophysics in spatial cognition research. In *Vision and attention*, pages 233–252. Springer, 2001.

[66] P.-C. Bürkner. brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, 80:1–28, 2017.

[67] D. Burmedi, S. Becker, V. Heyl, H.-W. Wahl, and I. Himmelsbach. Behavioral consequences of age-related low vision. *Visual Impairment Research*, 4(1):15–45, 2002.

[68] P. J. Burt. The pyramid as a structure for efficient computation. In *Multiresolution image processing and analysis*, pages 6–35. Springer, 1984.

[69] T. Butt, M. D. Crossland, P. West, S. W. Orr, and G. S. Rubin. Simulation contact lenses for amd health state utility values in nice appraisals: A different reality. *British Journal of Ophthalmology*, 99(4):540–544, 2015.

[70] B. Caldwell, M. Cooper, L. G. Reid, G. Vanderheiden, W. Chisholm, J. Slatin, and J. White. Web content accessibility guidelines (wcag) 2.0. *WWW Consortium (W3C)*, 290, 2008.

[71] K. Y. Cao and S. N. Markowitz. Residual stereopsis in age-related macular degeneration patients and its impact on vision-related abilities: a pilot study. *Journal of Optometry*, 7(2):100–105, 2014.

[72] Carl Zeiss Meditec, Inc. Humphrey field analyzer (HFA) II.

[73] Carl Zeiss Meditec, Inc. *Humphrey Field Analyzer II-i series System Software Version 5.1*, 2010. This document applies to the HFA II-i series instrument, System Software Version 5.1 or higher, unless superseded.

[74] B. T. Carter and S. G. Luke. Best practices in eye tracking research. *International Journal of Psychophysiology*, 155:49–62, 2020.

[75] R. Casati. Size from shadow: Some informational paths less traveled. *Ecological Psychology*, 28(1):56–63, 2016.

[76] R. Casati and P. Cavanagh. *The visual world of shadows*. MIT Press, 2019.

[77] N. Cattari, F. Cutolo, R. D'amato, U. Fontana, and V. Ferrari. Toed-in vs parallel displays in video see-through head-mounted displays for close-up view. *IEEE Access*, 7:159698–159711, 2019.

[78] P. Cavanagh and Y. G. Leclerc. Shape from shadows. *Journal of Experimental Psychology: Human Perception and Performance*, 15(1):3, 1989.

[79] N. T. Choplin and R. P. Edwards. *Visual field testing with the Humphrey Field Analyzer*. Slack Incorporated, 1995.

[80] R. Chou, T. Dana, C. Bougatsos, S. Grusing, and I. Blazina. Screening for impaired visual acuity in older adults: updated evidence report and systematic review for the us preventive services task force. *Jama*, 315(9):915–933, 2016.

[81] M. A. Cidota, R. M. S. Clifford, S. G. Lukosch, and M. Billinghurst. Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 172–177, 2016.

[82] A. Clemotte., M. Velasco., D. Torricelli., R. Raya., and R. Ceres. Accuracy and precision of the tobii x2-30 eye-tracking under non ideal conditions. In *Proceedings of the 2nd International Congress on Neurotechnology, Electronics and Informatics - NEUROTECHNIX,*, pages 111–116. INSTICC, SciTePress, 2014.

[83] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013.

[84] A. L. Corn and J. N. Erin. *Foundations of low vision: Clinical and functional perspectives*. American Foundation for the Blind, 2010.

[85] F. W. Cornelissen, K. J. Bruin, and A. C. Kooijman. The influence of artificial scotomas on eye movements during visual search. *Optometry & Vision Science*, 82(1):27–35, 2005.

[86] D. Crabb. A view on glaucoma—are we seeing it clearly? *Eye*, 30(2):304–313, 2016.

[87] D. P. Crabb, N. D. Smith, F. C. Glen, R. Burton, and D. F. Garway-Heath. How does glaucoma look?: patient perception of visual field loss. *Ophthalmology*, 120(6):1120–1126, 2013.

[88] D. P. Crabb, A. C. Viswanathan, A. I. McNaught, D. Poinoosawmy, F. W. Fitzke, and R. A. Hitchings. Simulating binocular visual field status in glaucoma. *British Journal of Ophthalmology*, 82(11):1236–1241, 1998.

[89] S. H. Creem and D. R. Proffitt. Defining the cortical visual systems:"what","where", and "how". *Acta psychologica*, 107(1-3):43–68, 2001.

[90] S. Creem-Regehr, J. Stefanucci, and W. Thompson. Perceiving absolute scale in virtual environments: How theory and application have mutually informed the role of body-based perception. *Psychology of Learning and Motivation - Advances in Research and Theory*, 62:195–224, 2015.

[91] S. H. Creem-Regehr, A. A. Gooch, C. S. Sahm, and W. B. Thompson. Perceiving virtual geographical slant: Action influences perception. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5):811–821, 2004.

[92] S. H. Creem-Regehr and B. R. Kunz. Perception and action. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6):800–810, 2010.

[93] S. H. Creem-Regehr, J. K. Stefanucci, and B. Bodenheimer. Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philosophical Transactions of the Royal Society B*, 378(1869):20210456, 2023.

[94] S. H. Creem-Regehr, J. K. Stefanucci, W. B. Thompson, N. Nash, and M. McCardell. Egocentric distance perception in the oculus rift (dk2). In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, SAP '15, pages 47–50, New York, NY, USA, 2015. ACM.

[95] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception*, 34(2):191–204, 2005. PMID: 15832569.

[96] J. E. Cutting and P. M. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of space and motion*, pages 69–117. Elsevier, 1995.

[97] L. Dandona and R. Dandona. Revision of visual impairment definitions in the international statistical classification of diseases. *BMC medicine*, 4(1):1–7, 2006.

[98] E. David, J. Beitner, and M. L.-H. Võ. Effects of transient loss of vision on head and eye movements during visual search in a virtual environment. *Brain sciences*, 10(11):841, 2020.

[99] A. D. Deemer, C. K. Bradley, N. C. Ross, D. M. Natale, R. Itthipanichpong, F. S. Werblin, and R. W. Massof. Low vision enhancement with head-mounted video display systems: are we there yet? *Optometry and vision science: official publication of the American Academy of Optometry*, 95(9):694, 2018.

11.0 REFERENCES

[100] A. Dey, A. Cunningham, and C. Sandor. Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, VRST '10, page 211–218, New York, NY, USA, 2010. Association for Computing Machinery.

[101] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr. Tablet versus phone: Depth perception in handheld augmented reality. In *2012 IEEE international symposium on mixed and augmented reality (ISMAR)*, pages 187–196. IEEE, 2012.

[102] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir. Designing for depth perceptions in augmented reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 111–122, Oct 2017.

[103] N. P. Diwakar. Empatheyes: Immersive virtual reality simulation of visual impairments using smartphones. Master's thesis, Indian Institute of Science, India, 2017.

[104] T. Do, J. J. LaViola Jr., and R. McMahan. The effects of object shape, fidelity, color, and luminance on depth perception in handheld mobile augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality(ISMAR)*, 2020.

[105] N. A. Dodgson. Variation and extrema of human interpupillary distance. In *Stereoscopic displays and virtual reality systems XI*, volume 5291, pages 36–46. International Society for Optics and Photonics, 2004.

[106] T. A. Doty and J. W. Kelly. Distance perception in virtual reality: Comparing modern headsets to the real world. *Technology, Mind, and Behavior*, 2021.

[107] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In M. T. Bolas, S. S. Fisher, M. T. Bolas, S. S. Fisher, and J. O. Merritt, editors, *Stereoscopic Displays and Virtual Reality Systems III*, volume 2653, pages 123 – 134. International Society for Optics and Photonics, SPIE, 1996.

[108] E. Eisemann, U. Assarsson, M. Schwarz, and M. Wimmer. Casting shadows in real time. In *ACM SIGGRAPH ASIA 2009 Courses*, SIGGRAPH ASIA '09, New York, NY, USA, 2009. Association for Computing Machinery.

[109] F. El Jamiy, A. N. R. Chandra, and R. Marsh. Distance accuracy of real environments in virtual reality head-mounted displays. In *2020 IEEE International Conference on Electro Information Technology (EIT)*, pages 281–287. IEEE, 2020.

[110] J. H. Elder, S. Trithart, G. Pintilie, and D. MacLean. Rapid processing of cast and attached shadows. *Perception*, 33(11):1319–1338, 2004.

[111] T. Endo, H. Kanda, M. Hirota, T. Morimoto, K. Nishida, and T. Fujikado. False reaching movements in localization test and effect of auditory feedback in simulated ultra-low vision subjects and patients with retinitis pigmentosa. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 254(5):947–956, 2016.

[112] S. Epstein. The stability of behavior: Ii. implications for psychological research. *American Psychologist*, 35(9):790–806, 09 1980.

[113] E. Ergun, N. Maar, W. Radner, I. Barbazetto, U. Schmidt-Erfurth, and M. Stur. Scotoma size and reading speed in patients with subfoveal occult choroidal neovascularization in age-related macular degeneration. *Ophthalmology*, 110(1):65–69, 2003.

[114] M. O. Ernst and H. H. Bülthoff. Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4):162–169, 2004.

[115] M. A. Eskenazi. Best practices for cleaning eye movement data in reading research. *Behavior Research Methods*, pages 1–11, 2023.

[116] K. Evans, S. K. Law, J. Walt, P. Buchholz, and J. Hansen. The quality of life impact of peripheral versus central vision loss with a focus on glaucoma versus age-related macular degeneration. *Clinical Ophthalmology (Auckland, NZ)*, 3:433, 2009.

[117] J. P. Ewins, M. D. Waller, M. White, and P. F. Lister. Mip-map level selection for texture mapping. *IEEE Transactions on Visualization and Computer Graphics*, 4(4):317–329, 1998.

[118] B. Farell and D. G. Pelli. Psychophysical methods, or how to measure a threshold and why. *Vision research: A practical guide to laboratory methods*, 5:129–136, 1999.

[119] G. T. Fechner. *Elemente der psychophysik*, volume 2. Breitkopf u. Härtel, 1860.

[120] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 Chi conference on human factors in computing systems*, pages 1118–1130, 2017.

[121] I. T. Feldstein, F. M. Kölsch, and R. Konrad. Egocentric distance perception: A comparative study investigating differences between real and virtual environments. *Perception*, 49(9):940–967, 2020.

[122] J. A. Ferwerda. Elements of early vision for computer graphics. *IEEE computer graphics and applications*, 21(5):22–33, 2001.

[123] S. Fiedler, M. Schulte-Mecklenbeck, F. Renkewitz, and J. L. Orquin. Guideline for reporting standards of eye-tracking research in decision sciences, Sep 2020.

[124] D. C. Fletcher, R. A. Schuchard, and L. W. Renninger. Patient awareness of binocular central scotoma in age-related macular degeneration. *Optometry and Vision Science*, 89(9):1395–1398, 2012.

[125] R. F. for Statistical Computing. R: A language and environment for statistical computing, 2010.

[126] F. C. Fortenbaugh, J. C. Hicks, L. Hao, and K. A. Turano. Losing sight of the bigger picture: Peripheral field loss compresses representations of space. *Vision research*, 47(19):2506–2520, 2007.

[127] F. C. Fortenbaugh, J. C. Hicks, L. Hao, and K. A. Turano. A technique for simulating visual field losses in virtual environments to study human navigation. *Behavior Research Methods*, 39(3):552–560, 2007.

[128] F. C. Fortenbaugh, J. C. Hicks, and K. A. Turano. The effect of peripheral visual field loss on representations of space: evidence for distortion and adaptation. *Investigative ophthalmology & visual science*, 49(6):2765–2772, 2008.

[129] J. Frankenstein, F. Kessler, and C. Rothkopf. Applying psychophysics to applied spatial cognition research. In J. Šķilters, N. S. Newcombe, and D. Uttal, editors, *Spatial Cognition XII*, pages 196–216, Cham, 2020. Springer International Publishing.

[130] W. Fuhl and E. Kasneci. Eye movement velocity and gaze data generator for evaluation, robustness testing and assess of eye tracking software and visualization tools. *arXiv preprint arXiv:1808.09296*, 2018.

[131] H. C. Gagnon, L. Buck, T. N. Smith, G. Narasimham, J. Stefanucci, S. H. Creem-Regehr, and B. Bodenheimer. Far distance estimation in mixed reality. In *ACM Symposium on Applied Perception 2020*, pages 1–8, 2020.

[132] H. C. Gagnon, C. S. Rosales, R. Mileris, J. K. Stefanucci, S. H. Creem-Regehr, and R. E. Bodenheimer. Estimating distances in action space in augmented reality. *ACM Transactions on Applied Perception (TAP)*, 18(2):1–16, 2021.

[133] H. C. Gagnon, Y. Zhao, M. Richardson, G. D. Pointon, J. K. Stefanucci, S. H. Creem-Regehr, and B. Bodenheimer. Gap affordance judgments in mixed reality: testing the role of display weight and field of view. *Frontiers in Virtual Reality*, 2:654656, 2021.

[134] M. Gagrani, J. Ndulue, D. Anderson, S. Kedar, V. Gulati, J. Shepherd, R. High, L. Smith, Z. Fowler, D. Khazanchi, et al. What do patients with glaucoma see: a novel ipad app to improve glaucoma patient awareness of visual field loss. *British Journal of Ophthalmology*, 106(2):218–222, 2022.

[135] Y. Gao, E. Peillard, J.-M. Normand, G. Moreau, Y. Liu, and Y. Wang. Influence of virtual objects' shadows and lighting coherence on distance perception in optical see-through augmented reality. *Journal of the Society for Information Display*, 2019.

[136] M. A. Garcıa-Pérez and E. Peli. Intrasaccadic perception. *Journal of Neuroscience*, 21(18):7313–7322, 2001.

[137] W. S. Geisler and J. S. Perry. Real-time simulation of arbitrary visual fields. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 83–87, 2002.

[138] A. Gelman and E. Loken. The statistical crisis in science. *The best writing on mathematics (Pitici M, ed)*, pages 305–318, 2016.

[139] M. A. Georgeson, K. A. May, T. C. Freeman, and G. S. Hesse. From filters to features: Scale–space analysis of edge and blur coding in human vision. *Journal of vision*, 7(13):7–7, 2007.

[140] F. Geringswald, E. Porracin, and S. Pollmann. Impairment of visual memory for objects in natural scenes by simulated central scotomata. *Journal of vision*, 16(2):6–6, 2016.

[141] D. R. Geruschat, K. Fujiwara, and R. S. W. Emerson. Traffic gap detection for pedestrians with low vision. *Optometry and vision science: official publication of the American Academy of Optometry*, 88(2):208, 2011.

[142] D. R. Geruschat, S. E. Hassan, K. A. Turano, H. A. Quigley, and N. G. Congdon. Gaze behavior of the visually impaired during street crossing. *Optometry and Vision Science*, 83(8):550–558, 2006.

[143] A. Gibaldi, M. Vanegas, P. J. Bex, and G. Maiello. Evaluation of the tobii eyex eye tracking controller and matlab toolkit for research. *Behavior research methods*, 49(3):923–946, 2017.

[144] J. J. Gibson. *The perception of the visual world*. Houghton Mifflin, 1950.

[145] J. J. Gibson. *The implications of experiments on the perception of space and motion*. US Department of Commerce, National Technical Information Service, 1975.

[146] J. J. Gibson. *The ecological approach to visual perception*. Houghton Mifflin, Boston, MA, 1979.

[147] A. S. Gilinsky. Perceived size and distance in visual space. *Psychological review*, 58(6):460, 1951.

[148] F. C. Glen, N. D. Smith, and D. P. Crabb. Impact of superior and inferior visual field loss on hazard detection in a computer-based driving test. *British Journal of Ophthalmology*, 99(5):613–617, 2015.

[149] A. Godfroid and B. Hui. Five common pitfalls in eye-tracking research. *Second Language Research*, 36(3):277–305, 2020.

[150] H. J. Godwin, M. C. Hout, K. J. Alexdóttir, S. C. Walenchok, and A. S. Barnhart. Avoiding potential pitfalls in visual search and eye-movement experiments: A tutorial review. *Attention, Perception, & Psychophysics*, 83(7):2753–2783, 2021.

[151] M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25, 1992.

[152] J. Goodman-Deane, S. Waller, and P. J. Clarkson. Simulating impairment. In *Proceedings of (Re) Actor3, the Third International Conference on Digital Live Art*, pages 21–22, 2008.

[153] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Trans. Appl. Percept.*, 7:26:1–26:18, July 2010.

[154] D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*, volume 1. Wiley New York, 1966.

[155] L. Gritz and E. d'Eon. The importance of being linear. *GPU Gems*, 3:529–542, 2007.

[156] L. Gruber, J. Ventura, and D. Schmalstieg. Image-space illumination for augmented reality in dynamic environments. In *2015 IEEE Virtual Reality (VR)*, pages 127–134, 2015.

[157] U. Gruenefeld, Y. Brück, and S. Boll. Behind the scenes: Comparing x-ray visualization techniques in head-mounted optical see-through augmented reality. In *19th International Conference on Mobile and Ubiquitous Multimedia*, pages 179–185, 2020.

[158] P. Grünwald. Model selection based on minimum description length. *Journal of mathematical psychology*, 44(1):133–152, 2000.

[159] R. Gueorguieva and J. H. Krystal. Move over anova: progress in analyzing repeated-measures data and its reflection in papers published in the archives of general psychiatry. *Archives of general psychiatry*, 61(3):310–317, 2004.

[160] J. Guez, J. Le Gargasson, F. Rigaudiere, and J. O'Regan. Is there a systematic location for the pseudo-fovea in patients with central scotoma? *Vision Research*, 33:1271–1279, 1993.

[161] C. R. Guibal and B. Dresp. Interaction of color and geometric cues in depth perception: When does "red" mean "near"? *Psychological Research*, 69(1-2):30–40, 2004.

[162] D. Guth, D. Ashmead, R. Long, R. Wall, and P. Ponchillia. Blind and sighted pedestrians' judgments of gaps in traffic at roundabouts. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(2):314–331, 2005.

[163] D. Guth, R. Long, R. Wall, P. Ponchillia, and D. Ashmead. Blind and sighted pedestrians' road-crossing judgments at a single-lane roundabout. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55:632, 2013.

[164] M. Haller, C. Hanl, and J. Diephuis. Non-photorealistic rendering techniques for motion in computer games. *Comput. Entertain.*, 2(4):11, Oct. 2004.

[165] F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the american statistical association*, 69(346):383–393, 1974.

[166] C. R. Harrison and K. M. Robinette. Caesar: Summary statistics for the adult population (ages 18-65) of the united states of america. Technical report, Air Force Research Lab Wright-Patterson AFB OH Human Effectiveness Directorate, 06 2002.

[167] S. E. Hassan. Are normally sighted, visually impaired, and blind pedestrians accurate and reliable at making street crossing decisions? *Investigative Ophthalmology and Visual Science*, 53:2593–2600, 2012.

[168] S. E. Hassan, J. E. Lovie-Kitchin, R. L. Woods, et al. Vision and mobility performance of subjects with age-related macular degeneration. *Optometry and vision science*, 79(11):697–707, 2002.

[169] S. E. Hassan and B. D. Snyder. Street-crossing decision-making: a comparison between patients with age-related macular degeneration and normal vision. *Investigative ophthalmology & visual science*, 53(10):6137–6144, 2012.

[170] A.-K. Hauperich, L. K. Young, and H. E. Smithson. What makes a microsaccade? a review of 70 years of research prompts a new detection method. *Journal of Eye Movement Research*, 12(6), 2019.

[171] Z. J. He, B. Wu, T. L. Ooi, G. Yarbrough, and J. Wu. Judging egocentric distance on the ground: Occlusion and surface integration. *Perception*, 33(7):789–806, 2004.

[172] A. Heijl, V. M. Patella, and B. Bengtsson. *The field analyzer primer: effective perimetry*. Carl Zeiss Meditec Incorporated, 4 edition, 2012.

[173] K. Heikkilä, N. Pudas, and S. Ukkola. Comparing eye tracking technologies. B.S. thesis, K. Heikkilä; N. Pudas; S. Ukkola, 2023.

[174] R. T. Held and M. S. Banks. Misperceptions in stereoscopic displays: a vision science perspective. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 23–32, 2008.

[175] J. Hertel and F. Steinicke. Augmented reality for maritime navigation assistance-egocentric depth perception in large distance outdoor environments. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 122–130. IEEE, 2021.

[176] R. S. Hessels, R. Andersson, I. T. Hooge, M. Nyström, and C. Kemner. Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, 20(6):601–633, 2015.

[177] R. S. Hessels and I. T. Hooge. Eye tracking in developmental cognitive neuroscience–the good, the bad and the ugly. *Developmental cognitive neuroscience*, 40:100710, 2019.

[178] R. S. Hessels, C. Kemner, C. van den Boomen, and I. T. Hooge. The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior research methods*, 48:1694–1712, 2016.

[179] J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy. Combining sensory information: mandatory fusion within, but not between, senses. *Science*, 298(5598):1627–1630, 2002.

[180] J. C. Ho. Real-world and virtual-world practices for virtual reality games: Effects on spatial perception and game performance. *Multimodal Technologies and Interaction*, 4(1):1, 2020.

[181] C. Hochberg, E. Maul, E. S. Chan, S. Van Landingham, L. Ferrucci, D. S. Friedman, and P. Y. Ramulu. Association of vision loss in glaucoma and age-related macular degeneration with iadl disability. *Investigative ophthalmology & visual science*, 53(6):3201–3206, 2012.

[182] M. Hogervorst and W. Van Damme. Visualizing visual impairments. *Gerontechnology*, 5(4):208–221, 2006.

[183] M. A. Hogervorst and W. J. van Damme. Visualizing the limits of low vision in detecting natural image features. *Optometry and Vision Science*, 85(10):E951–E962, 2008.

[184] K. Holmqvist. Common predictors of accuracy, precision and data loss in 12 eye-trackers, 2017.

[185] K. Holmqvist and R. Andersson. Eye tracking: A comprehensive guide to methods. *paradigms and measures*, 2017.

[186] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[187] K. Holmqvist, M. Nyström, and F. Mulvey. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*, pages 45–52, 2012.

[188] K. Holmqvist, S. L. Örbom, I. T. Hooge, D. C. Niehorster, R. G. Alexander, R. Andersson, J. S. Benjamins, P. Blignaut, A.-M. Brouwer, L. L. Chuang, et al. Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior research methods*, pages 1–53, 2022.

[189] I. T. Hooge, R. S. Hessels, and M. Nyström. Do pupil-based binocular video eye trackers reliably measure vergence? *Vision Research*, 156:1–9, 2019.

[190] A. J. Hornof and T. Halverson. Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers*, 34(4):592–604, 2002.

[191] I. P. Howard. The perception of posture, self motion, and the visual vertical. *Handbook of perception and human performance*, pages 18–1, 1986.

[192] J. Hox and D. McNeish. Small samples in multilevel modeling. *Small sample size solutions*, pages 215–225, 2020.

[193] B. Hu and C. Brown. *Cast Shadows in Augmented Reality Systems*. PhD thesis, University of Rochester, USA, 2005.

[194] C. X. Hu, C. Zangalli, M. Hsieh, L. Gupta, A. L. Williams, J. Richman, and G. L. Spaeth. What do patients with glaucoma see? visual symptoms reported by patients with glaucoma. *The American journal of the medical sciences*, 348(5):403–409, 2014.

[195] H. H. Hu, A. A. Gooch, W. B. Thompson, B. E. Smits, J. J. Rieser, and P. Shirley. Visual cues for imminent object contact in realistic virtual environment. In *IEEE Visualization*, pages 179–185, 2000.

[196] S. Hu, N. D. Smith, L. J. Saunders, and D. P. Crabb. Patterns of binocular visual field loss derived from large-scale patient data from glaucoma clinics. *Ophthalmology*, 122(12):2399–2406, 2015.

[197] H. Hua, P. Krishnaswamy, and J. P. Rolland. Video-based eyetracking methods and algorithms in head-mounted displays. *Optics Express*, 14(10):4328–4350, 2006.

[198] G. S. Hubona, P. N. Wheeler, G. W. Shirah, and M. Brandt. The relative contributions of stereo, lighting, and background scenes in promoting 3d depth visualization. *ACM Trans. Comput.-Hum. Interact.*, 6(3):214–242, Sept. 1999.

[199] A. D. Hwang and E. Peli. Instability of the perceived world while watching 3d stereoscopic imagery: a likely source of motion sickness symptoms. *i-Perception*, 5(6):515–535, 2014.

[200] J. E. Hyde. Some characteristics of voluntary human ocular movements in the horizontal plane. *American journal of ophthalmology*, 48(1):85–94, 1959.

[201] International classification of diseases (icd), Jan 2022.

[202] B. Iglewicz and D. C. Hoaglin. *Volume 16: how to detect and handle outliers*. Quality Press, 1993.

[203] S. Ikeda, Y. Kimura, S. Manabe, A. Kimura, and F. Shibata. Shadow induction on optical see-through head-mounted displays. *Computers & Graphics*, 91:141 – 152, 2020.

[204] T. Imura, N. Shirai, M. Tomonaga, M. K. Yamaguchi, and A. Yagi. Asymmetry in the perception of motion in depth induced by moving cast shadows. *Journal of Vision*, 8(13):10–10, 2008.

[205] V. Interrante, B. Ries, and L. Anderson. Distance perception in immersive virtual environments, revisited. In *IEEE virtual reality conference (VR 2006)*, pages 3–10. IEEE, 2006.

[206] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano. Light attenuation display: Subtractive see-through near-eye display via spatial color filtering. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1951–1960, 2019.

[207] A. J. Jackson and J. S. Wolffsohn. *Low vision manual*. Elsevier Health Sciences, 2007.

[208] R. A. Jacobs. What determines visual cue reliability? *Trends in Cognitive Sciences*, 6(8):345 – 350, 2002.

[209] C. Jacquemin, G. Gagneré, and B. Lahoz. Shedding light on shadow: Real-time interactive artworks based on cast shadows or silhouettes. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, page 173–182, New York, NY, USA, 2011. Association for Computing Machinery.

[210] B. Jin, Z. Ai, and M. Rasmussen. Simulation of eye disease in virtual reality. In *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5128–5131. IEEE Computer Society, 2005.

[211] J. A. Jones, E. A. Suma, D. M. Krum, and M. Bolas. Comparability of narrow and wide field-of-view head-mounted displays for medium-field distance judgments. In *Proceedings of the ACM Symposium on Applied Perception*, pages 119–119, 2012.

[212] J. A. Jones, J. E. Swan, G. Singh, and S. R. Ellis. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, APGV '11, page 29–36, New York, NY, USA, 2011. Association for Computing Machinery.

[213] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, APGV '08, page 9–14, New York, NY, USA, 2008. Association for Computing Machinery.

[214] J. A. Jones, J. E. Swan II, and M. Bolas. Peripheral stimulation and its effect on perceived spatial scale in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):701–710, 2013.

[215] P. R. Jones and G. Ometto. Degraded reality: using vr/ar to simulate visual impairments. In *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)*, pages 1–4. IEEE, 2018.

[216] P. R. Jones, T. Somoskeöy, H. Chow-Wing-Bom, and D. P. Crabb. Seeing other perspectives: evaluating the use of virtual and augmented reality to simulate visual impairments (openvissim). *NPJ digital medicine*, 3(1):1–9, 2020.

[217] V. Jurgens, A. Cockburn, and M. Billinghurst. Depth cues for augmented reality stakeout. In *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI*, pages 117–124, 2006.

[218] M. Kahook and R. J. Noecker. How do you interpret a 24-2 humphrey visual field printout. *Glaucoma Today*, 1:57–62, 2007.

[219] T. Kaminokado, Y. Hiroi, and Y. Itoh. Stainedview: Variable-intensity light-attenuation display with cascaded spatial color filtering for improved color fidelity. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3576–3586, 2020.

[220] S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn. Arett: Augmented reality eye tracking toolkit for head mounted displays. *Sensors*, 21(6):2234, 2021.

[221] S. Kasahara, S. Higa, and A. Komori. Fragment shadow: Generating fragmented shadows with multi-projectors geometry and color calibration. In *ACM SIGGRAPH 2019 Studio*, SIGGRAPH '19, New York, NY, USA, 2019. Association for Computing Machinery.

[222] J. Keil, A. Korte, A. Ratmer, D. Edler, and F. Dickmann. Augmented reality (ar) and spatial cognition: effects of holographic grids on distance estimation and location memory in a 3d indoor scenario. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(2):165–172, 2020.

[223] J. W. Kelly. Distance perception in virtual reality: A meta-analysis of the effect of head-mounted display characteristics. *IEEE transactions on visualization and computer graphics*, 2022.

[224] J. W. Kelly, L. A. Cherep, and Z. D. Siegel. Perceived space in the htc vive. *ACM Trans. Appl. Percept.*, 15(1), July 2017.

[225] J. W. Kelly, L. S. Donaldson, L. A. Sjolund, and J. B. Freiberg. More than just perception–action recalibration: Walking through a virtual environment causes rescaling of perceived space. *Attention, Perception, & Psychophysics*, 75(7):1473–1485, 2013.

[226] J. W. Kelly, W. W. Hammel, Z. D. Siegel, and L. A. Sjolund. Recalibration of perceived distance in virtual environments occurs rapidly and transfers asymmetrically across scale. *IEEE transactions on visualization and computer graphics*, 20(4):588–595, 2014.

[227] M. Keo. Graphical style in video games. *Programme in Information Technology Riihimäki*, 2017.

[228] D. Kersten, P. Mamassian, and D. C. Knill. Moving cast shadows induce apparent motion in depth. *Perception*, 26(2):171–192, 1997.

[229] A. Keshava, F. N. Nezami, N. Maleki, L. Tiemann, and P. König. Stress testing vr eye-tracking system performance. In *Neuroergonomics Conference*, volume 11, 2021.

[230] F. Kilpatrick and W. Ittelson. The size-distance invariance hypothesis. *Psychological review*, 60(4):223, 1953.

[231] M. Kinateder and E. A. Cooper. Assessing effects of reduced vision on spatial orientation ability using virtual reality. *Baltic Journal of Modern Computing*, 9(3):243–259, 2021.

[232] E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt. Measurement protocols for medium-field distance perception in large-screen immersive displays. In *2009 IEEE virtual reality conference*, pages 107–113. IEEE, 2009.

[233] J. M. Knapp and J. M. Loomis. Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators and Virtual Environments*, 13:572–577, 2004.

[234] D. C. Knill, P. Mamassian, and D. Kersten. Geometry of shadows. *JOSA A*, 14(12):3216–3232, 1997.

[235] H. Kolivand, A. H. Zakaria, and M. S. Sunar. Shadow generation in mixed reality: A comprehensive survey. *IETE Technical Review*, 32(1):3–15, 2015.

[236] M. Koller. robustlmm: an r package for robust estimation of linear mixed-effects models. *Journal of statistical software*, 75:1–24, 2016.

[237] H. Komatsu. The neural mechanisms of perceptual filling-in. *Nature reviews neuroscience*, 7(3):220–231, 2006.

[238] A. Kotecha, N. O'Leary, D. Melmoth, S. Grant, and D. P. Crabb. The functional consequences of glaucoma for eye–hand coordination. *Investigative ophthalmology & visual science*, 50(1):203–213, 2009.

[239] E. Kowler. Eye movements: The past 25 years. *Vision research*, 51(13):1457–1483, 2011.

[240] K. Krösl. *Simulating Vision Impairments in Virtual and Augmented Reality*. PhD thesis, Wien, 2020.

[241] K. Krösl, D. Bauer, M. Schwärzler, H. Fuchs, G. Suter, and M. Wimmer. A vr-based user study on the effects of vision impairments on recognition distances of escape-route signs in buildings. *Vis. Comput.*, 34(6–8):911–923, June 2018.

[242] K. Krösl, C. Elvezio, M. Hürbe, S. Karst, S. Feiner, and M. Wimmer. Xreye: Simulating visual impairments in eye-tracked xr. In *2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*, pages 830–831. IEEE, 2020.

[243] K. Krösl, C. Elvezio, M. Hürbe, S. Karst, M. Wimmer, and S. Feiner. Icthroughvr: Illuminating cataracts through virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 655–663. IEEE, 2019.

[244] K. Krösl, C. Elvezio, L. R. Luidolt, M. Hürbe, S. Karst, S. Feiner, and M. Wimmer. Cataract: Simulating cataracts in augmented reality. In *2020 IEEE ISMAR*, pages 682–693, 2020.

[245] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 3–12, 2010.

[246] S. A. Kuhl, W. B. Thompson, and S. H. Creem-Regehr. Hmd calibration and its effects on distance judgments. *ACM Transactions on Applied Perception (TAP)*, 6(3):1–20, 2009.

[247] B. R. Kunz, S. H. Creem-Regehr, and W. B.Thompson. Evidence for motor simulation in imagined locomotion. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5):1458–1471, 2009.

[248] B. R. Kunz, L. Wouters, D. Smith, W. B. Thompson, and S. H. Creem-Regehr. Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception, & Psychophysics*, 71(6):1284–1293, 2009.

[249] T. Kuyk, J. L. Elliott, and P. S. Fuhr. Visual correlates of obstacle avoidance in adults with low vision. *Optometry and vision science: official publication of the American Academy of Optometry*, 75(3):174–182, 1998.

[250] A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, et al. Package 'lmertest'. *R package version*, 2(0):734, 2015.

[251] S. G. Kwak and J. H. Kim. Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology*, 70(2):144, 2017.

[252] M. Kwon, A. S. Nandy, and B. S. Tjan. Rapid and persistent adaptability of human oculomotor control in response to simulated central vision loss. *Current Biology*, 23(17):1663–1669, 2013.

[253] M. Kwon, C. Ramachandra, P. Satgunam, B. W. Mel, E. Peli, and B. S. Tjan. Contour enhancement benefits older adults with simulated central field loss. *Optometry and vision science: official publication of the American Academy of Optometry*, 89(9):1374, 2012.

[254] M. Kytö, A. Mäkinen, J. Häkkinen, and P. Oittinen. Improving relative depth judgments in augmented reality with auxiliary augmentations. *ACM Trans. Appl. Percept.*, 10(1), Mar. 2013.

[255] M. Kytö, A. Mäkinen, T. Tossavainen, and P. T. Oittinen. Stereoscopic depth perception in video see-through augmented reality within action space. *Journal of Electronic Imaging*, 23(1):1 – 11, 2014.

[256] M. S. Landy, M. S. Banks, and D. C. Knill. Ideal-observer models of cue integration. *Sensory cue integration*, pages 5–29, 2011.

[257] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision research*, 35(3):389–412, 1995.

[258] J. S. Lappin, A. L. Shelton, and J. J. Rieser. Environmental context influences visually perceived distance. *Perception & psychophysics*, 68(4):571–581, 2006.

[259] A. M. Larson and L. C. Loschky. The contributions of central versus peripheral vision to scene gist recognition. *Journal of vision*, 9(10):6–6, 2009.

[260] S. LaValle. *Virtual reality*. National Programme on Technology Enhanced Learning (NPTEL), 2016.

[261] R. Lawson and H. H. Bülthoff. Using morphs of familiar objects to examine how shape discriminability influences view sensitivity. *Perception & psychophysics*, 70(5):853–877, 2008.

[262] S. J. Leat, G. E. Legge, and M. Bullimore. What is low vision? a re-evaluation of definitions. *Optometry and Vision Science*, 76:198–211, 1999.

[263] S. Lee, E. Eisemann, and H.-P. Seidel. Real-time lens blur effects and focus control. *ACM Transactions on Graphics (TOG)*, 29(4):1–7, 2010.

[264] S. S.-Y. Lee, A. A. Black, and J. M. Wood. Effect of glaucoma on eye movement patterns and laboratory-based hazard detection ability. *PLoS One*, 12(6):e0178876, 2017.

[265] W. J. Lee, J. H. Kim, Y. U. Shin, S. Hwang, and H. W. Lim. Differences in eye movement range based on age and gaze direction. *Eye*, 33(7):1145–1151, 2019.

[266] G. E. Legge, R. Gage, Y. Baek, and T. M. Bochsler. Indoor spatial updating with reduced visual information. *PLoS One*, 11(3):e0150708, 2016.

[267] G. E. Legge, C. Granquist, Y. Baek, and R. Gage. Indoor spatial updating with impaired vision. *Investigative Ophthalmology & Visual Science*, 57(15):6757–6765, 2016.

[268] G. E. Legge, D. Yu, C. S. Kallie, T. M. Bochsler, and R. Gage. Visual accessibility of ramps and steps. *Journal of vision*, 10(11):8–8, 2010.

[269] J. Lewis, D. Brown, W. Cranton, and R. Mason. Simulating visual impairments using the unreal engine 3 game engine. In *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–8. IEEE, 2011.

[270] J. Lewis, L. Shires, and D. Brown. Development of a visual impairment simulator using the microsoft xna framework. In *Proc. 9th Intl Conf. Disability, Virtual Reality & Associated Technologies, Laval, France*, 2012.

[271] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, 49(4):764–766, 2013.

[272] B. Li, A. Nordman, J. Walker, and S. A. Kuhl. The effects of artificially reduced field of view and peripheral frame stimulation on distance judgments in hmds. In *Proceedings of the ACM Symposium on Applied Perception*, pages 53–56, 2016.

[273] B. Li, J. Walker, and S. A. Kuhl. The effects of peripheral vision and light stimulation on distance judgments through hmds. *ACM Transactions on Applied Perception (TAP)*, 15(2):1–14, 2018.

[274] B. Li, R. Zhang, A. Nordman, and S. Kuhl. The effects of minification and display field of view on distance judgments in real and hmd-based environments. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '15, New York, NY, USA, 2015. ACM.

[275] Z. Li, J. Phillips, and F. H. Durgin. The underestimation of egocentric distance: Evidence from frontal matching tasks. *Attention, Perception, & Psychophysics*, 73(7):2205–2217, 2011.

[276] R. W. Lindeman and H. Noma. A classification scheme for multi-sensory augmented reality. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, pages 175–178, 2007.

[277] C. C. Liu and M. Aitkin. Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, 52(6):362–375, 2008.

[278] L. Liu, Y.-C. T. Shih, R. L. Strawderman, D. Zhang, B. A. Johnson, and H. Chai. Statistical analysis of zero-inflated nonnegative continuous data. *Statistical Science*, 34(2):253–279, 2019.

[279] N. Liu and M. Pang. A survey of shadow rendering algorithms: Projection shadows and shadow volumes. In *2009 Second International Workshop on Computer Science and Engineering*, volume 1, pages 488–492, Oct 2009.

[280] D. J. Lohr, L. Friedman, and O. V. Komogortsev. Evaluating the data quality of eye tracking signals from a virtual reality system: Case study using smi's eye-tracking htc vive. *arXiv preprint arXiv:1912.02083*, 2019.

[281] J. M. Loomis, J. A. Da Silva, N. Fujita, and S. S. Fukusima. Visual space perception and visually directed action. *Journal of experimental psychology: Human Perception and Performance*, 18(4):906, 1992.

[282] J. M. Loomis and J. M. Knapp. *Virtual and Adaptive Environments*, chapter Visual perception of egocentric distance in real and virtual environments, pages 21–46. Erlbaum, Mahwah, NJ, 2003.

[283] J. M. Loomis and J. W. Philbeck. Measuring spatial perception with spatial updating and action. In R. L. Klatzky, B. MacWihnney, and M. Behrmann, editors, *Embodiment, Ego-Space, and Action*. Taylor & Francis, New York, 2008.

[284] J. Lopez-Moreno, J. Jimenez, S. Hadap, E. Reinhard, K. Anjyo, and D. Gutierrez. Stylized depiction of images based on depth perception. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering*, page 109–118, New York, NY, USA, 2010. Association for Computing Machinery.

[285] S. R. Lord and J. Dayhew. Visual risk factors for falls in older people. *Journal of the American Geriatrics Society*, 49(5):508–515, 2001.

[286] L. C. Loschky and G. W. McConkie. User performance with gaze contingent multiresolutional displays. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 97–103, 2000.

[287] L. C. Loschky and G. S. Wolverton. How late can you update gaze-contingent multiresolutional displays without detection? *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 3(4):1–10, 2007.

[288] D. Lüdecke, D. Makowski, P. Waggoner, and I. Patil. Package 'performance', 2019.

[289] R. Ludt and G. L. Goodrich. Change in visual perceptual detection distances for low vision travelers as a result of dynamic visual assessment and training. *Journal of Visual Impairment & Blindness*, 96(1):7–21, 2002.

[290] L. R. Luidolt, M. Wimmer, and K. Krösl. Gaze-dependent simulation of light perception in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3557–3567, 2020.

[291] S. G. Luke. Evaluating significance in linear mixed-effects models in r. *Behavior research methods*, 49:1494–1502, 2017.

[292] M. H. Lynn, G. Luo, M. Tomasi, S. Pundlik, and K. E. Houston. Measuring virtual reality headset resolution and field of view: implications for vision care applications. *Optometry and Vision Science*, 97(8):573–582, 2020.

[293] R. MacAlpine and D. R. Flatla. Real-time mobile personalized simulations of impaired colour vision. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 181–189, 2016.

[294] D. J. Mack, S. Belfanti, and U. Schwarz. The effect of sampling rate and lowpass filters on saccades–a modeling approach. *Behavior Research Methods*, 49:2146–2162, 2017.

[295] C. Madison, W. Thompson, D. Kersten, P. Shirley, and B. Smits. Use of interreflection and shadow for surface contact. *Perception & Psychophysics*, 63(2):187–194, 2001.

[296] P. Mamassian. Impossible shadows and the shadow correspondence problem. *Perception*, 33(11):1279–1290, 2004.

[297] P. Mamassian, D. C. Knill, and D. Kersten. The perception of cast shadows. *Trends in cognitive sciences*, 2(8):288–295, 1998.

[298] S. Manabe, S. Ikeda, A. Kimura, and F. Shibata. Casting virtual shadows based on brightness induction for optical see-through displays. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 627–628, 2018.

[299] S. Manabe, S. Ikeda, A. Kimura, and F. Shibata. Shadow inducers: Inconspicuous highlights for casting virtual shadows on ost-hmds. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1331–1332, March 2019.

[300] J. Mankoff, H. Fait, and R. Juang. Evaluating accessibility by simulating the experiences of users with vision or motor impairments. *IBM Systems Journal*, 44(3):505–517, 2005.

[301] S. Mann and S. M. Nnlf. *Mediated reality*. Citeseer, 1994.

[302] C. Mao, K. Go, Y. Kinoshita, K. Kashiwagi, M. Toyoura, I. Fujishiro, J. Li, and X. Mao. Different eye movement behaviors related to artificial visual field defects–a pilot study of video-based perimetry. *IEEE Access*, 9:77649–77660, 2021.

[303] S. Masnadi, K. Pfeil, J.-V. T. Sera-Josef, and J. LaViola. Effects of field of view on egocentric distance perception in virtual reality. In *CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2022.

[304] S. Masnadi, K. P. Pfeil, J.-V. T. Sera-Josef, and J. J. LaViola. Field of view effect on distance perception in virtual reality. In *2021 ieee conference on virtual reality and 3d user interfaces abstracts and workshops (vrw)*, pages 542–543. IEEE, 2021.

[305] E. H. Matsushima, A. M. Vaz, R. A. Cazuza, and N. P. Ribeiro Filho. Independence of egocentric and exocentric direction processing in visual space. *Psychology & Neuroscience*, 7(3):277–284, 2014.

[306] H. Matuschek, R. Kliegl, S. Vasishth, H. Baayen, and D. Bates. Balancing type i error and power in linear mixed models. *Journal of memory and language*, 94:305–315, 2017.

[307] J. W. McCandless, S. R. Ellis, and B. D. Adelstein. Localization of a time-delayed, monocular virtual object superimposed on a real environment. *Presence*, 9(1):15–24, 2000.

[308] G. W. McConkie and L. C. Loschky. Perception onset time during fixations in free viewing. *Behavior Research Methods, Instruments, & Computers*, 34(4):481–490, 2002.

[309] W. J. McGill and J. Gibbon. The general-gamma distribution and reaction times. *Journal of mathematical Psychology*, 2(1):1–18, 1965.

[310] L. McIlreavy, J. Fiser, and P. J. Bex. Impact of simulated central scotomas on visual search in natural scenes. *Optometry and vision science: official publication of the American Academy of Optometry*, 89(9):1385, 2012.

[311] D. Medeiros, M. Sousa, D. Mendes, A. Raposo, and J. Jorge. Perceiving depth: optical versus video see-through. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 237–240, 2016.

[312] A. Meir, T. Oron-Gilad, and Y. Parmet. Are child-pedestrians able to identify hazardous traffic situations? measuring their abilities in a virtual reality environment. *Safety Science*, 80:33 – 40, 2015.

[313] D. R. Melmoth, A. L. Finlay, M. J. Morgan, and S. Grant. Grasping deficits and adaptations in adults with stereo vision losses. *Investigative ophthalmology & visual science*, 50(8):3711–3720, 2009.

[314] J. Meng and H. Sedgwick. Distance perception across spatial discontinuities. *Perception & Psychophysics*, 64(1):1–14, 2002.

[315] J. C. Meng and H. Sedgwick. Distance perception mediated through nested contact relations among surfaces. *Perception & Psychophysics*, 63(1):1–15, 2001.

[316] R. Michael and A. Bron. The ageing lens and cataract: a model of normal and pathological ageing. *Philosophical transactions of the Royal Society B: Biological sciences*, 366(1568):1278–1292, 2011.

[317] Microsoft Soundscape. https://www.microsoft.com/en-us/research/product/soundscape, 2019.

[318] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[319] B. J. Mohler, S. H. Creem-Regehr, and W. B. Thompson. The influence of feedback on egocenteric distance judgments in real and virtual environments. In *Proc. Symposium on Applied Perception in Graphics and Visualization*, pages 9–14, 2006.

[320] G. Montesano, A. Chen, R. Lu, C. S. Lee, and A. Y. Lee. Uwhvf: A real-world, open source dataset of perimetry tests from the humphrey field analyzer at the university of washington. *Translational vision science & technology*, 11(1):2–2, 2022.

[321] B. A. Morrongiello, M. Corbett, J. Switzer, and T. Hall. Using a virtual environment to study pedestrian behaviors: How does time pressure affect children's and adults' street crossing behaviors? *Journal of Pediatric Psychology*, 40(7):697–703, 2015.

[322] M. Mott, E. Cutrell, M. G. Franco, C. Holz, E. Ofek, R. Stoakley, and M. R. Morris. Accessible by design: An opportunity for virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 451–454. IEEE, 2019.

[323] D. J. Mundfrom, J. J. Perrett, J. Schaffer, A. Piccone, and M. Roozeboom. Bonferroni adjustments in tests for regression coefficients. *Multiple Linear Regression Viewpoints*, 32(1):1–6, 2006.

[324] R. Musil. Hmd geometry database | collected geometry data from some commercially avaiable vr headsets, 2021. https://doi.org/10.3390/brainsci11020223.

[325] S. Nagata. How to reinforce perception of depth in single two-dimensional pictures. In *Pictorial Communication In Real And Virtual Environments*, pages 553–573. CRC Press, 1991.

[326] B.-G. Nam, M.-w. Lee, and H.-J. Yoo. Development of a 3-d graphics rendering engine with lighting acceleration for handheld multimedia systems. *IEEE Transactions on Consumer Electronics*, 51(3):1020–1027, 2005.

[327] P. E. Napieralski, B. M. Altenhoff, J. W. Bertrand, L. O. Long, S. V. Babu, C. C. Pagano, J. Kern, and T. A. Davis. Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Transactions on Applied Perception (TAP)*, 8(3):1–19, 2011.

[328] National eye institute | at a glance: Low vision, 2022. accessed 2 July 2022.

[329] National eye institute media library, 2022. accessed 2 July 2022.

[330] P. Nelson, P. Aspinall, O. Papasouliotis, B. Worton, and C. O'Brien. Quality of life in glaucoma and its relationship with visual function. *Journal of glaucoma*, 12(2):139–150, 2003.

[331] W. T. Nelson, M. M. Roe, R. S. Bolia, and R. M. Morley. Assessing simulator sickness in a see-through hmd: Effects of time delay, time on task, and task complexity. Technical report, Air Force Research Lab Wright-Patterson AFB OH, 2000.

[332] R. Ni, M. Braunstein, and G. Andersen. Distance perception from motion parallax and ground contact. *Visual Cognition*, 12(6):1235–1254, 2005.

[333] R. Ni, M. L. Braunstein, and G. J. Andersen. Perception of scene layout from optical contact, shadows, and motion. *Perception*, 33(11):1305–1318, 2004. PMID: 15693673.

[334] D. C. Niehorster, T. Santini, R. S. Hessels, I. T. Hooge, E. Kasneci, and M. Nyström. The impact of slippage on the data quality of head-worn eye trackers. *Behavior research methods*, 52:1140–1160, 2020.

[335] D. C. Niehorster, R. Zemblys, T. Beelders, and K. Holmqvist. Characterizing gaze position signals and synthesizing noise during fixations in eye-tracking data. *Behavior Research Methods*, 52(6):2515–2534, 2020.

[336] D. C. Niehorster, R. Zemblys, and K. Holmqvist. Is apparent fixational drift in eye-tracking data due to filters or eyeball rotation? *Behavior Research Methods*, 53(1):311–324, 2021.

[337] Z. Noh and M. S. Sunar. A review of shadow techniques in augmented reality. In *2009 Second International Conference on Machine Vision*, pages 320–324, 2009.

[338] K. Nouri-Mahdavi. Selecting visual field tests and assessing visual field deterioration in glaucoma. *Canadian Journal of Ophthalmology*, 49(6):497–505, 2014.

[339] M. Nürnberger, C. Klingner, O. W. Witte, and S. Brodoehl. Mismatch of visual-vestibular information in virtual reality: Is motion sickness part of the brains attempt to reduce the prediction error? *Frontiers in human neuroscience*, page 648, 2021.

[340] A. Nuthmann and T. Canas-Bajo. Visual search in naturalistic scenes from foveal to peripheral vision: A comparison between dynamic and static displays. *Journal of vision*, 22(1):10–10, 2022.

[341] M. Nyström, R. Andersson, K. Holmqvist, and J. Van De Weijer. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods*, 45(1):272–288, 2013.

[342] N. V. Odell, D. A. Leske, S. R. Hatt, W. E. Adams, and J. M. Holmes. The effect of bangerter filters on optotype acuity, vernier acuity, and contrast sensitivity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 12(6):555–559, 2008.

[343] C. A. Okoro, N. D. Hollis, A. C. Cyrus, and S. Griffin-Blake. Prevalence of disabilities and health care access by disability status and type among adults—united states, 2016. *Morbidity and Mortality Weekly Report*, 67(32):882, 2018.

[344] C. M. Oman. Sensory conflict in motion sickness: an observer theory approach. *Pictorial communication in virtual and real environments*, pages 362–376, 1991.

[345] C. W. Ooi and J. Dingliana. Colored cast shadows for improved visibility in optical see-through ar. In *SIGGRAPH Asia 2020 Posters*, SA '20, New York, NY, USA, 2020. Association for Computing Machinery.

[346] T. L. Ooi, B. Wu, and Z. J. He. Distance determined by the angular declination below the horizon. *Nature*, 414(6860):197–200, 2001.

[347] T. L. Ooi, B. Wu, and Z. J. He. Perceptual space in the dark affected by the intrinsic bias of the visual system. *Perception*, 35(5):605–624, 2006.

[348] J. L. Orquin, N. J. Ashby, and A. D. Clarke. Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, 29(2-3):103–115, 2016.

[349] J. L. Orquin and K. Holmqvist. Threats to the validity of eye-movement research in psychology. *Behavior research methods*, 50:1645–1656, 2018.

[350] A. Ö. Orta, Z. K. Öztürker, S. Ö. Erkul, Ş. Bayraktar, and O. F. Yilmaz. The correlation between glaucomatous visual field loss and vision-related quality of life. *Journal of glaucoma*, 24(5):e121–e127, 2015.

[351] J. W. Osborne. Advantages of hierarchical linear modeling. *Practical Assessment, Research, and Evaluation*, 7(1):1, 2019.

[352] K. Ozkan and M. L. Braunstein. Background surface and horizon effects in the perception of relative size and distance. *Visual cognition*, 18(2):229–254, 2010.

[353] C. C. Pagano and G. P. Bingham. Comparing measures of monocular distance perception: Verbal and reaching errors are not correlated. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1037, 1998.

[354] V. F. Pamplona, E. B. Passos, J. Zizka, M. M. Oliveira, E. Lawson, E. Clua, and R. Raskar. Catra: Interactive measuring and modeling of cataracts. In *ACM Trans. Graph.*, volume 30, New York, NY, USA, July 2011. Association for Computing Machinery.

[355] S. Pardhan, C. Gonzalez-Alvarez, and A. Subramanian. How does the presence and duration of central visual impairment affect reaching and grasping movements? *Ophthalmic and Physiological Optics*, 31(3):233–239, 2011.

[356] S. Pardhan, A. Scarfe, R. Bourne, and M. Timmis. A comparison of reach-to-grasp and transport-to-place performance in participants with age-related macular degeneration and glaucoma. *Investigative Ophthalmology & Visual Science*, 58(3):1560–1569, 2017.

[357] D. Pascolini and S. P. Mariotti. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, 96(5):614–618, 2012.

[358] S. Pastel, C.-H. Chen, L. Martin, M. Naujoks, K. Petri, and K. Witte. Comparison of gaze accuracy and precision in real-world and virtual reality. *Virtual Reality*, 25:175–189, 2021.

[359] A. Peer and K. Ponto. Evaluating perceived distance measures in room-scale spaces using consumer-grade head mounted displays. In *2017 IEEE symposium on 3d user interfaces (3dui)*, pages 83–86. IEEE, 2017.

[360] E. Peillard, F. Argelaguet, J. Normand, A. Lécuyer, and G. Moreau. Studying exocentric distance perception in optical see-through augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 115–122, 2019.

[361] E. Peillard, Y. Itoh, G. Moreau, J. M. Normand, A. Lécuyer, and F. Argelaguet. Can retinal projection displays improve spatial perception in augmented reality? In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 80–89, 2020.

[362] E. Peli. Test of a model of foveal vision by using simulations. *JOSA A*, 13(6):1131–1138, 1996.

[363] E. Peli and G. A. Geri. Validation and verification of a visual model for central and peripheral vision. Technical report, SCHEPENS EYE RESEARCH INST BOSTON MA, 2000.

[364] E. Peli, J.-H. Jung, and R. Goldstein. Better simulation of vision with central and paracentral scotomas. *Investigative Ophthalmology & Visual Science*, 61(7):3370–3370, 2020.

[365] E. Peli and M. Moharrer. Driving with low vision. In *Albert and Jakobiec's Principles and Practice of Ophthalmology*, pages 1–18. Springer, 2021.

[366] G. M. Pérez, S. M. Archer, and P. Artal. Optical characterization of bangerter foils. *Investigative ophthalmology & visual science*, 51(1):609–613, 2010.

[367] J. S. Perry and W. S. Geisler. Gaze-contingent real-time simulation of arbitrary visual fields. In *Human vision and electronic imaging VII*, volume 4662, pages 57–69. International Society for Optics and Photonics, 2002.

[368] K. Pfeil, S. Masnadi, J. Belga, J.-V. T. Sera-Josef, and J. LaViola. Distance perception with a video see-through head-mounted display. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2021.

[369] J. W. Philbeck, D. A. Gajewski, S. M. Jaidzeka, and C. P. Wallin. The role of top-down knowledge about environmental context in egocentric distance judgments. *Attention, Perception, & Psychophysics*, 80(2):586–599, 2018.

[370] J. W. Philbeck and J. M. Loomis. Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1):72, 1997.

[371] J. W. Philbeck, A. J. Woods, J. Arthur, and J. Todd. Progressive locomotor recalibration during blind walking. *Perception & psychophysics*, 70(8):1459–1470, 2008.

[372] J. Ping, B. H. Thomas, J. Baumeister, J. Guo, D. Weng, and Y. Liu. Effects of shading model and opacity on depth perception in optical see-through augmented reality. *Journal of the Society for Information Display*, 2020.

[373] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker. Distance perception in real and virtual environments. *ACM Transactions on Applied Perception*, 2:216–233, 2005.

[374] G. Pointon, C. Thompson, S. Creem-Regehr, J. Stefanucci, M. Joshi, R. Paris, and B. Bodenheimer. Judging action capabilities in augmented reality. In *Proceedings of the 15th ACM Symposium on Applied Perception*, SAP '18, pages 6:1–6:8. ACM, 2018.

[375] S. L. Polyak. *The retina.* Univ. Chicago Press, 1941.

[376] V. Powell. *Visual properties of virtual target objects: implications for reaching and grasping tasks in a virtual reality rehabilitation context*. PhD thesis, University of Portsmouth, 2012.

[377] N. Prins et al. *Psychophysics: a practical introduction*. Academic Press, 2016.

[378] M. Qiu, S. Y. Wang, K. Singh, and S. C. Lin. Association between visual field defects and quality of life in the united states. *Ophthalmology*, 121(3):733–740, 2014.

[379] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[380] V. S. Ramachandran. Filling in gaps in perception: Part i. *Current Directions in Psychological Science*, 1(6):199–205, 1992.

[381] V. S. Ramachandran and R. L. Gregory. Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350(6320):699–702, 1991.

[382] P. Y. Ramulu, A. Mihailovic, S. K. West, D. S. Friedman, and L. N. Gitlin. What is a falls risk factor? factors associated with falls per time or per step in individuals with glaucoma. *Journal of the American Geriatrics Society*, 67(1):87–92, 2019.

[383] K. M. Rand, M. R. Tarampi, S. H. Creem-Regehr, and W. B. Thompson. The influence of ground contact and visible horizon on perception of distance and size under severely degraded vision. *Seeing and Perceiving*, 25(5):425 – 447, 07 Jul. 2012.

[384] K. M. Rand, M. R. Tarampi, S. H. Creem-Regehr, and W. B. Thompson. The importance of a visual horizon for distance judgments under severely degraded vision. *Perception*, 40(2):143–154, 2011.

[385] R. Ratcliff. Group reaction time distributions and an analysis of distribution statistics. *Psychological bulletin*, 86(3):446, 1979.

[386] S. W. Raudenbush and A. S. Bryk. *Hierarchical linear models: Applications and data analysis methods*, volume 1. Sage, 2002.

[387] K. Rayner, B. R. Foorman, C. A. Perfetti, D. Pesetsky, and M. S. Seidenberg. How psychological science informs the teaching of reading. *Psychological science in the public interest*, 2(2):31–74, 2001.

[388] L. G. Reid and A. Snow-Weaver. Wcag 2.0: a web accessibility standard for the evolving web. In *Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, pages 109–115, 2008.

[389] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments — A review. *ACM Computer Surveys*, 46(2):23:1–23:40, 2013.

[390] R. A. Rensink and P. Cavanagh. The influence of cast shadows on visual search. *Perception*, 33(11):1339–1358, 2004.

[391] C. D. Riemann, S. Hanson, and J. A. Foster. A comparison of manual kinetic and automated static perimetry in obtaining ptosis fields. *Archives of ophthalmology*, 118(1):65–69, 2000.

[392] J. J. Rieser, D. H. Ashmead, C. R. Talor, and G. A. Youngquist. Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception*, 19(5):675–689, 1990.

[393] J. R. Rocke, R. F. J. O'Day, and T. C. Roydhouse. Angle closure glaucoma secondary to psychotropic medications. *Australasian Medical Journal (Online)*, 10(7):581, 2017.

[394] M. Roh, A. Selivanova, H. J. Shin, J. W. Miller, and M. L. Jackson. Visual acuity and contrast sensitivity are two important factors affecting vision-related quality of life in advanced age-related macular degeneration. *PloS one*, 13(5):e0196481, 2018.

[395] J. P. Rolland, W. Gibson, and D. Ariely. Towards quantifying depth and size perception in virtual environments. *Presence: Teleoperators & Virtual Environments*, 4(1):24–49, 1995.

[396] J. P. Rolland, R. L. Holloway, and H. Fuchs. Comparison of optical and video see-through, head-mounted displays. In H. Das, editor, *Telemanipulator and Telepresence Technologies*, volume 2351, pages 293 – 307. International Society for Optics and Photonics, SPIE, 1995.

[397] C. S. Rosales, G. Pointon, H. Adams, J. Stefanucci, S. Creem-Regehr, W. B. Thompson, and B. Bodenheimer. Distance judgments to on- and off-ground objects in augmented reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 237–243, 2019.

[398] A. S. Ross, X. Zhang, J. Fogarty, and J. O. Wobbrock. An epidemiology-inspired large-scale analysis of android app accessibility. *ACM Trans. Access. Comput.*, 13(1), Apr. 2020.

[399] J. Ross, M. C. Morrone, M. E. Goldberg, and D. C. Burr. Changes in visual perception at the time of saccades. *Trends in neurosciences*, 24(2):113–121, 2001.

[400] P. J. Rousseeuw and M. Hubert. Robust statistics for outlier detection. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, 1(1):73–79, 2011.

[401] R. A. Ruddle. Navigation: Am i really lost or virtually there? *Engineering Psychology and Cognitive Ergonomics*, 6:135–142, 2001.

[402] F. Ruotolo, I. J. van Der Ham, T. Iachini, and A. Postma. The relationship between allocentric and egocentric frames of reference and categorical and coordinate spatial information processing. *Quarterly Journal of Experimental Psychology*, 64(6):1138–1156, 2011.

[403] E. E. Sabelman and R. Lam. The real-life dangers of augmented reality. *IEEE Spectrum*, 52(7):48–53, 2015.

[404] L. N. Saftari and O.-S. Kwon. Ageing vision and falls: a review. *Journal of physiological anthropology*, 37(1):1–14, 2018.

[405] Y. Sauer, A. Sipatchin, S. Wahl, and M. García García. Assessment of consumer vr-headsets' objective and subjective field of view (fov) and its feasibility for visual field testing. *Virtual Reality*, pages 1–13, 2022.

[406] D. R. Saunders and R. L. Woods. Direct measurement of the system latency of gaze-contingent displays. *Behavior research methods*, 46(2):439–447, 2014.

[407] G. B. Schaalje, J. B. McBride, and G. W. Fellingham. Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics*, 7:512–524, 2002.

[408] H. Schielzeth, N. J. Dingemanse, S. Nakagawa, D. F. Westneat, H. Allegue, C. Teplitsky, D. Réale, N. A. Dochtermann, L. Z. Garamszegi, and Y. G. Araya-Ajoy. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9):1141–1152, 2020.

[409] V. R. Schinazi. Psychosocial implications of blindness and low-vision. *UCL (University College London), Centre for Advanced Spatial Analysis (UCL)*, 2007.

[410] E. F. Schubert. Human eye sensitivity and photometric quantities. *Light-emitting diodes*, pages 275–291, 2006.

[411] R. A. Schuchard. Adaptation to macular scotomas in persons with low vision. *The American Journal of Occupational Therapy*, 49(9):870–876, 1995.

[412] R. A. Schuchard, S. Naseer, K. de Castro, and J. R. R. Dev. Characteristics of amd patients with low vision receiving visual rehabilitation. *Journal of Rehabilitation Research & Development*, 36(4), 1999.

[413] I. Schuetz and K. Fiehler. Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research*, 15(3), 2022.

[414] I. Schuetz, T. S. Murdison, K. J. MacKenzie, and M. Zannoli. An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[415] I. Schuetz, T. S. Murdison, and M. Zannoli. A psychophysics-inspired model of gaze selection performance. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–5, 2020.

[416] L. Schwetlick, L. O. M. Rothkegel, H. A. Trukenbrod, and R. Engbert. Modeling the effects of perisaccadic attention on gaze statistics during scene viewing. *Communications biology*, 3(1):727, 2020.

[417] H. Sedgwick. Environment-centered representation of spatial layout: Available visual information from texture and perspective. *Human and machine vision*, pages 425–458, 1983.

[418] H. Sedgwick. Visual space perception. *Blackwell handbook of sensation and perception*, pages 128–167, 2005.

[419] H. A. Sedgwick. *The visible horizon: A potential source of visual information for the perception of size and distance*. Cornell University, 1973.

[420] S. Seo. *A review and comparison of methods for detecting outliers in univariate data sets*. PhD thesis, University of Pittsburgh, 2006.

[421] A. E. Seward, D. H. Ashmead, and B. Bodenheimer. Using virtual environments to assess time-to-contact judgments from pedestrian viewpoints. *ACM Transactions on Applied Perception*, 4(3), Nov. 2007.

[422] F. Shic, K. Chawarska, and B. Scassellati. The amorphous fixation measure revisited: With applications to autism. In *30th Annual Meeting of the Cognitive Science Society*, pages 1–6, 2008.

[423] S. Shioiri. Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics*, 53(3):305–314, 1993.

[424] L. Sidenmark, M. N. Lystbæk, and H. Gellersen. Ge-simulator: An open-source tool for simulating real-time errors for hmd-based eye trackers. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, pages 1–6, 2023.

[425] Z. D. Siegel and J. W. Kelly. Walking through a virtual environment improves perceived size within and beyond the walked space. *Attention, Perception, & Psychophysics*, 79(1):39–44, 2017.

[426] Z. D. Siegel, J. W. Kelly, and L. A. Cherep. Rescaling of perceived space transfers across virtual environments. *Journal of Experimental Psychology: Human Perception and Performance*, 43(10):1805, 2017.

[427] M. J. Sinai, T. L. Ooi, and Z. J. He. Terrain influences the accurate judgement of distance. *Nature*, 395(6701):497–500, 1998.

[428] G. Singh, S. R. Ellis, and J. E. Swan. The effect of focal distance, age, and brightness on near-field augmented reality depth matching. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1385–1398, 2020.

[429] G. Singh, J. E. Swan, J. A. Jones, and S. R. Ellis. Depth judgments by reaching and matching in near-field augmented reality. In *2012 IEEE Virtual Reality Workshops (VRW)*, pages 165–166. IEEE, 2012.

[430] A. Sipatchin, S. Wahl, and K. Rifai. Eye-tracking for clinical ophthalmology with virtual reality (vr): A case study of the htc vive pro eye's usability. In *Healthcare*, volume 9, page 180. Mdpi, 2021.

[431] B. Sivak and C. L. MacKenzie. Integration of visual information and motor output in reaching and grasping: the contributions of peripheral and central vision. *Neuropsychologia*, 28(10):1095–1116, 1990.

[432] R. Skarbez, M. Smith, and M. C. Whitton. Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2:27, 2021.

[433] M. Slater, M. Usoh, and Y. Chrysanthou. The influence of dynamic shadows on presence in immersive virtual environments. In *Selected Papers of the Eurographics Workshops on Virtual Environments '95*, VE '95, page 8–21, Berlin, Heidelberg, 1995. Springer-Verlag.

[434] M. Smith, N. Doutcheva, J. L. Gabbard, and G. Burnett. Optical see-through head up displays' effect on depth judgments of real world objects. In *2015 IEEE Virtual Reality (VR)*, pages 401–405. IEEE, 2015.

[435] P. L. Smith and D. R. Little. Small is beautiful: In defense of the small-n design. *Psychonomic bulletin & review*, 25(6):2083–2101, 2018.

[436] O. Špakov. Comparison of eye movement filters used in hci. In *Proceedings of the symposium on eye tracking research and applications*, pages 281–284, 2012.

[437] T. Stan Development. *RStan: the R interface to Stan*, 2018.

[438] J. K. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Comparing distance judgments in real and augmented reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 82–86. IEEE, 2021.

[439] N. Stein, D. C. Niehorster, T. Watson, F. Steinicke, K. Rifai, S. Wahl, and M. Lappe. A comparison of eye tracking latencies among several commercial head-mounted displays. *i-Perception*, 12(1):2041669520983338, 2021.

[440] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE Transactions on Visualization and Computer Graphics*, 16:17–27, 2010.

[441] W. Steptoe, S. Julier, and A. Steed. Presence and discernability in conventional and non-photorealistic immersive augmented reality. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 213–218, 2014.

[442] Y. Sterken, A. Postma, E. H. F. D. Haan, and A. Dingemans. Egocentric and exocentric spatial judgements of visual displacement. *The Quarterly Journal of Experimental Psychology Section A*, 52(4):1047–1055, 1999.

[443] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93, 1992.

[444] G. Stewart and R. McCrindle. Visual impairment simulator for auditing and design. *Journal of Alternative Medical Research*, 9(4), 2017.

[445] S. Stock, C. Erler, and W. Stork. Realistic simulation of progressive vision diseases in virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–2, 2018.

[446] S. C. Stock, C. Erler, W. Stork, G. Labuz, H. S. Son, R. Khoramnia, and G. U. Auffarth. Suitability of virtual reality for vision simulation–a case study using glaucomatous visual fields. *Investigative Ophthalmology & Visual Science*, 60(9):2441–2441, 2019.

[447] A. E. Stoper and M. M. Cohen. Judgments of eye level in light and in darkness. *Perception & Psychophysics*, 40(5):311–316, 1986.

[448] H. Strasburger, I. Rentschler, and M. Jüttner. Peripheral vision and pattern recognition: A review. *Journal of vision*, 11(5):13–13, 2011.

[449] N. Sugano, H. Kato, and K. Tachibana. The effects of shadow representation of virtual objects in augmented reality. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '03, page 76, USA, 2003. IEEE Computer Society.

[450] J. S. Sunness, J. Gonzalez-Baron, C. A. Applegate, N. M. Bressler, Y. Tian, B. Hawkins, Y. Barron, and A. Bergman. Enlargement of atrophy and visual acuity loss in the geographic atrophy form of age-related macular degeneration. *Ophthalmology*, 106(9):1768–1779, 1999.

[451] R. T. Surdick, E. T. Davis, R. A. King, G. M. Corso, A. Shapiro, L. Hodges, and K. Elliot. Relevant cues for the visual perception of depth: is where you see it where it is? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 38, pages 1305–1309. SAGE Publications Sage CA: Los Angeles, CA, 1994.

[452] R. T. Surdick, E. T. Davis, R. A. King, and L. F. Hodges. The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence: Teleoperators & Virtual Environments*, 6(5):513–531, 1997.

[453] I. E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764, 1968.

[454] J. Swan, M. A. Livingston, H. S. Smallman, D. Brown, Y. Baillot, J. L. Gabbard, and D. Hix. A perceptual matching technique for depth judgments in optical, see-through augmented reality. *Proceedings of the IEEE Virtual Reality Conference, 2006*, pages 19–26, 2006.

[455] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE transactions on visualization and computer graphics*, 13(3):429–442, 2007.

[456] J. E. Swan, G. Singh, and S. R. Ellis. Matching and reaching depth judgments with real and augmented reality targets. *IEEE Transactions on Visualization and Computer Graphics*, 21(11):1289–1298, 2015.

[457] N.-C. Tai and M. Inanici. Luminance contrast as depth cue: Investigation and design applications. *Computer-aided Design and Applications*, 9:691–705, 2012.

[458] A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi. Development of a stereo video see-through hmd for ar systems. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pages 68–77. IEEE, 2000.

[459] S. Tanimoto and T. Pavlidis. A hierarchical data structure for picture processing. *Computer graphics and image processing*, 4(2):104–119, 1975.

[460] M. R. Tarampi. *Complex spatial updating in simulated low vision*. PhD thesis, Department of Psychology, University of Utah, 2010.

[461] M. R. Tarampi, S. H. Creem-Regehr, and W. B. Thompson. Visually directed walking to targets viewed with severely degraded vision is surprisingly accurate. *Journal of Vision*, 8(6):754–754, 2008.

[462] D. J. Taylor, L. A. Edwards, A. M. Binns, and D. P. Crabb. Seeing it differently: self-reported description of vision loss in dry age-related macular degeneration. *Ophthalmic and Physiological Optics*, 38(1):98–105, 2018.

[463] T. Technology. Accuracy and precision test method for remote eye trackers, 2011.

[464] M. Teghtsoonian and R. Teghtsoonian. Scaling apparent distance in natural indoor settings. *Psychonomic Science*, 16(6):281–283, 1969.

[465] R. Teghtsoonian and M. Teghtsoonian. Scaling apparent distance in a natural outdoor setting. *Psychonomic Science*, 21(4):215–216, 1970.

[466] W. B. Thompson, V. Dilda, and S. H. Creem-Regehr. Absolute distance perception to locations off the ground plane. *Perception*, 36(11):1559–1571, 2007.

[467] W. B. Thompson, P. Shirley, B. Smits, D. J. Kersten, and C. Madison. Visual glue. *University of Utah Technical Report UUCS-98-007*, 1998.

[468] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments. *Presence: Teleoperators and Virtual Environments*, 13:560–571, 2004.

[469] G. W. Tigwell. Nuanced perspectives toward disability simulations from digital designers, blind, low vision, and color blind people. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[470] G. T. Timberlake, E. Omoscharka, B. M. Quaney, S. A. Grose, and J. H. Maino. Effect of bilateral macular scotomas from age-related macular degeneration on reach-to-grasp hand movement. *Investigative Ophthalmology & Visual Science*, 52(5):2540–2550, 2011.

[471] C. Tochel, J. Morton, J. Jay, and J. Morrison. Relationship between visual field loss and contrast threshold elevation in glaucoma. *BMC ophthalmology*, 5(1):1–12, 2005.

[472] M. Treisman. Noise and weber's law: The discrimination of brightness and other dimensions. *Psychological review*, 71(4):314, 1964.

[473] K. A. Turano, G. S. Rubin, and H. A. Quigley. Mobility performance in glaucoma. *Investigative Ophthalmology & Visual Science*, 40(12):2803–2809, 1999.

[474] Valve. Openvr api.

[475] R. Varma, T. S. Vajaranant, B. Burkemper, S. Wu, M. Torres, C. Hsu, F. Choudhury, and R. McKean-Cowdin. Visual impairment and blindness in adults in the united states: demographic and geographic variations from 2015 to 2050. *JAMA ophthalmology*, 134(7):802–809, 2016.

[476] M. Varsori, A. Perez-Fornos, A. B. Safran, and A. R. Whatham. Development of a viewing strategy during adaptation to an artificial central scotoma. *Vision Research*, 44(23):2691–2705, 2004.

[477] J. Väyrynen, A. Colley, and J. Häkkilä. Head mounted display design tool for simulating visual disabilities. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, pages 69–73, 2016.

[478] K. Vaziri, M. Bondy, A. Bui, and V. Interrante. Egocentric distance judgments in full-cue video-see-through vr conditions are no better than distance judgments to targets in a void. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 1–9. IEEE, 2021.

[479] K. Vaziri, P. Liu, S. Aseeri, and V. Interrante. Impact of visual and experiential realism on distance perception in vr using a custom video see-through system. In *Proceedings of the ACM Symposium on Applied Perception*, pages 1–8, 2017.

[480] A. Vehlen, W. Standard, and G. Domes. How to choose the size of facial areas of interest in interactive eye tracking. *Plos one*, 17(2):e0263594, 2022.

[481] R. Velázquez, J. Varona, and P. Rodrigo. Computer-based system for simulating visual impairments. *IETE Journal of Research*, 62(6):833–841, 2016.

[482] P. Verghese, T. L. Tyson, S. Ghahghaei, and D. C. Fletcher. Depth perception and grasp in central field loss. *Investigative ophthalmology & visual science*, 57(3):1476–1487, 2016.

[483] W. B. Verwey and H. A. Veltman. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of experimental psychology: Applied*, 2(3):270, 1996.

[484] M. Vinnikov, R. S. Allison, and D. Swierad. Real-time simulation of visual defects with gaze-contingent display. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 127–130, 2008.

[485] V. Šoltészová, D. Patel, and I. Viola. Chromatic shadows for improved perception. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, NPAR '11, page 105–116, New York, NY, USA, 2011. Association for Computing Machinery.

[486] M. Wagner. The metric of visual space. *Perception & psychophysics*, 38(6):483–495, 1985.

[487] H. Wallach and A. O'Leary. Slope of regard as a distance cue. *Perception & Psychophysics*, 31(2):145–148, 1982.

[488] D. V. Walsh and L. Liu. Adaptation to a simulated central scotoma during visual search training. *Vision Research*, 96:75–86, 2014.

[489] C.-W. Wang, C. L. Chan, and I. Chi. Overview of quality of life research in older people with visual impairment. *Advances in Aging Research*, 2014, 2014.

[490] D. Wang, F. B. Mulvey, J. B. Pelz, and K. Holmqvist. A study of artificial eyes for the measurement of precision in eye-trackers. *Behavior Research Methods*, 49(3):947–959, 2017.

[491] L. Wanger. The effect of shadow quality on the perception of spatial relationships in computer generated imagery. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 39–42, 1992.

[492] L. R. Wanger, J. A. Ferwerda, D. P. Greenberg, et al. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications*, 12(3):44–58, 1992.

[493] S. V. Wass, L. Forssman, and J. Leppänen. Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy*, 19(5):427–460, 2014.

[494] S. V. Wass, T. J. Smith, and M. H. Johnson. Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior research methods*, 45(1):229–250, 2013.

[495] B. Wentink. Eye-hand coordination in laparoscopy-an overview of experiments and supporting aids. *Minimally Invasive Therapy & Allied Technologies*, 10(3):155–162, 2001.

[496] F. Werfel, R. Wiche, J. Feitsch, and C. Geiger. Empathizing audiovisual sense impairments: Interactive real-time illustration of diminished sense perception. In *Proceedings of the 7th Augmented Human International Conference 2016*, pages 1–8, 2016.

[497] C. Wheatstone. Xviii. contributions to the physiology of vision.—part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London*, pages 371–394, 1838.

[498] B. Wheelwright, Y. Sulai, Y. Geng, S. Luanava, S. Choi, W. Gao, and J. Gollier. Field of view: not just a number. In *Digital Optics for Immersive Displays*, volume 10676, page 1067604. International Society for Optics and Photonics, 2018.

[499] R. Whelan. Effective analysis of reaction time data. *The psychological record*, 58:475–482, 2008.

[500] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanics on distance judgements in virtual environments. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, pages 36–48, 2004.

[501] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Transactions on Applied Perception (TAP)*, 6(2):1–14, 2009.

[502] B. G. Witmer and W. J. J. Sadowski. Nonvisually guided locomotion to a previously viewed target in real and virutal environments. *Human Factors*, 40:478–488, 1998.

[503] W. Wittich, D. H. Watanabe, M. A. Kapusta, and O. Overbury. Subjective perception of visual distortions or scotomas in individuals with retinitis pigmentosa. *Journal of Visual Impairment & Blindness*, 105(1):50–55, 2011.

[504] H. Woltman, A. Feldstain, J. C. MacKay, and M. Rocchi. An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology*, 8(1):52–69, 2012.

[505] S. H. Wong and G. T. Plant. How to interpret visual fields. *Practical neurology*, 15(5):374–381, 2015.

[506] W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C.-Y. Cheng, and T. Y. Wong. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health*, 2(2):106–116, 2014.

[507] A. J. Woods, T. Docherty, and R. Koch. Image distortions in stereoscopic video systems. In *Stereoscopic displays and applications IV*, volume 1915, pages 36–48. SPIE, 1993.

[508] A. J. Woods, J. W. Philbeck, and J. V. Danoff. The various perceptions of distance: An alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4):1104, 2009.

[509] B. Wu, Z. J. He, and T. L. Ooi. Inaccurate representation of the ground surface beyond a texture boundary. *Perception*, 36(5):703–721, 2007.

[510] B. Wu, T. L. Ooi, and Z. J. He. Perceiving distance accurately by a directional process of integrating ground information. *Nature*, 428(6978):73–77, 2004.

[511] H. Wu, D. H. Ashmead, H. Adams, and B. Bodenheimer. 3d sound rendering in a virtual environment to evaluate pedestrian street crossing decisions at a roundabout. In *2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pages 1–6. IEEE, 2018.

[512] H. Wu, D. H. Ashmead, H. Adams, and B. Bodenheimer. Using virtual reality to assess the street crossing behavior of pedestrians with simulated macular degeneration at a roundabout. *Frontiers in ICT*, 5:27, 2018.

[513] XR Access. `https://xraccess.org`, 2020.

[514] A. Yonas, L. T. Goldsmith, and J. L. Hallstrom. Development of sensitivity to information provided by cast shadows in pictures. *Perception*, 7(3):333–341, 1978.

11.0 REFERENCES

[515] R. Young. The gaussian derivative model for spatial vision. i- retinal mechanisms. *Spatial vision*, 2(4):273–293, 1987.

[516] T. Zhan, K. Yin, J. Xiong, Z. He, and S.-T. Wu. Augmented reality and virtual reality displays: Perspectives and challenges. *Iscience*, page 101397, 2020.

[517] Y. Zhao, E. Cutrell, C. Holz, M. R. Morris, E. Ofek, and A. D. Wilson. Seeingvr: A set of tools to make virtual reality more accessible to people with low vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[518] C. J. Ziemer, J. M. Plumert, J. F. Cremer, and J. K. Kearney. Estimating distance in real and virtual environments: Does order make a difference? *Attention, Perception, & Psychophysics*, 71(5):1095–1106, 2009.

[519] J. L. Zuckerman, D. Miller, W. Dyes, and M. Keller. Degradation of vision through a simulated cataract. *Investigative Ophthalmology & Visual Science*, 12(3):213–224, 1973.

[520] D. Zur and S. Ullman. Filling-in of retinal scotomas. *Vision research*, 43(9):971–982, 2003.