PERCEIVED DEPTH IN VIRTUAL AND PHYSICAL ENVIRONMENTS

BRITTNEY HARTLE

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

AUGUST 2022

ABSTRACT

Theoretically, stereopsis provides accurate depth information if information regarding absolute distance is accurate and reliable. However, assessments of stereopsis often report depth distortions, particularly for virtual stimuli. These distortions are often attributed to misestimates of viewing distance caused by limited distance cues and/or the presence of conflicts between ocular distance cues in virtual displays. To understand how these factors contribute to depth distortions, I conducted a series of experiments in which depth was estimated under a range of viewing conditions and cue combinations.

In the first series (Chapter 2), I evaluated if conflicts between oculomotor distance cues drive depth underconstancy observed in virtual environments by comparing judgments of virtual and physical objects. The results showed that depth judgments of physical stimuli were accurate and exhibited depth constancy, but judgments of virtual stimuli failed to achieve depth constancy. This failure was due in part to the presence of the vergence-accommodation conflict. Further, prior experience with each environment had a profound effect on depth judgments, e.g., performance in virtual environments was enhanced by limited exposure to a similar task using physical objects.

In Chapter 3, I assessed if limitations of virtual environments contributed to previous failures of linear combination models to account for the integration of stereopsis and motion cues. I measured the perceived depth of virtual and physical objects defined by motion parallax, binocular disparity, or their combination. Accuracy was remarkedly similar for both environments, but estimates were more precise when depth was defined by binocular disparity than motion parallax. A linear combination model did not adequately describe performance in either physical or virtual conditions.

In Chapter 4, I evaluated if reaching to virtual objects provides distance information that can be used to scale stereopsis using an interactive ring game. Brief experience reaching to virtual objects improved the accuracy and scaling of subsequent depth judgements. Overall, experience with physical objects or reaching-in-depth enhanced performance on tasks dependent on distance perception. To fully understand how binocular depth perception is used to interact with objects in the real world, it is important to assess these cues in a rich, full-cue natural scenes.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my supervisor, Laurie Wilcox, for all her continued support, understanding, and expertise over these past few years of my graduate studies. I have learned an incredible amount from you in this time. Thank you for always encouraging me to be confident in my studies.

Thank you to my supervisory members, Robert Allison and Laurence Harris for guiding me throughout this process, and my examination committee, Erez Freud, Taylor Cleworth, and Peter Scarfe for their insightful feedback. I would also like to thank my colleagues and lab-mates over the years for listening to my rants about current issues and providing their support. A big thank you to all the observers who contributed their time participating in my experiments.

Thank you to my close friends and family for putting up with my slow descent into madness while completing this dissertation over the course of the covid-19 pandemic. Especially my sister, Chantelle, who lived with me throughout my graduate studies and listened to me practice too many talks about stereoscopic depth perception. Lastly, I would like to thank my mom, Karen for all her support and encouragement, despite my insistence that she does not need to read every paper I publish.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The ability to accurately estimate the depth and distance of objects is critical to our interpretation of, and interaction with the world around us. Here, depth refers to the extent of an object along the z-dimension, while distance refers to the amount of space from the eye to a point on the object's surface. To maintain a stable three-dimensional (3D) percept of the world, the perceived depth between relative positions on an object should remain constant over a range of viewing distances. One of the primary sources of depth information within near space is stereopsis[1].

1.1 Geometry of Stereopsis

Stereopsis uses the positional disparity between each eye's retinal image, the observer's interpupillary distance, and the knowledge of the observer's absolute viewing distance to the object to compute metric depth. The geometry of this relationship is illustrated in Figure 1.1. This binocular geometry defines two types of signals for the stereoscopic system: absolute and relative disparity.

[1] We use the term stereopsis as a short-hand for stereoscopic depth perception as suggested by Duane in 1917 (see Wade, 2021 for review) based on the terminology of Helmholtz who used the term 'stereoscopic parallax' when referring to the depth percept that results when viewing stereoscopic imagery (Helmholtz, 1925, page 299).

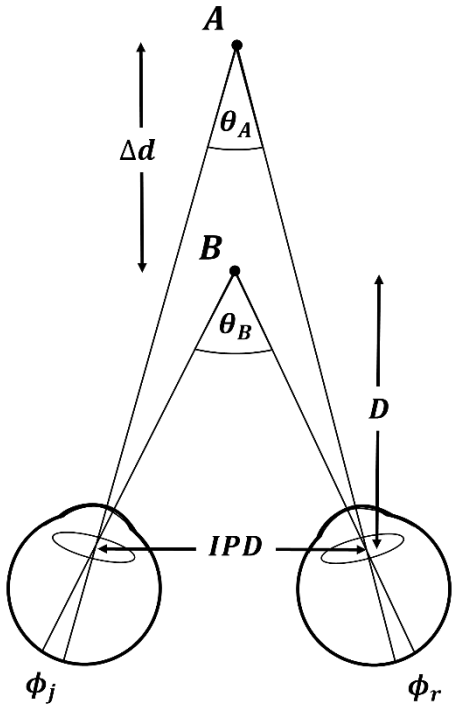


Figure 1.1. *An illustration of the geometrical definition of binocular disparity for two points located on the median plane. The eyes converge on point B at distance D, while a second point A is placed at Δd distance beyond point B. The visual angle between point A and B for the left eye is ϕ_j, and the same angle for the right eye is ϕ_r. The binocular disparity of point A is determined from the fixation point B at distance D, the distance between point A and B (Δd), and the interpupillary distance (IPD), where the binocular substance of A and B are θ_A and θ_B, respectively. This illustration is based on Figure 14.5 in Howard & Rogers (2012).*

Absolute disparity refers to the difference between the angles of rotation of the two eyes when they converge on a single object. In Figure 1.1, the point of fixation B has zero disparity, while point A has an uncrossed absolute disparity of θ_A. To accurately estimate the Δd of point A, the visual system must derive an estimate of fixation distance (D) based on the angle of convergence between the two eyes. Absolute depth estimation is notoriously erroneous and biased, causing large variability in depth estimates and poor stereoacuity (Blakemore, 1970; Westheimer, 1979). Depth estimation accuracy is improved by providing an additional point of reference and comparing the relative disparity (i.e., the difference between absolute disparities) between two points (Poggio & Poggio, 1984). In the case of Figure 1.1, the relative binocular disparity (δ) between point A and B is defined as the difference between their absolute disparities ($\theta_B - \theta_A$), which in this case is also equivalent to $\phi_j - \phi_r$. Note that because

the eyes are converged on point B, Figure 1.1 represents a special case where the relative disparity between point A and B is equivalent to the absolute disparity of point A.

If we assume symmetrical convergence and the small angle approximation (i.e., the tangent of an angle is approximately equal to the angle in radians), then the predicted depth from binocular disparity is approximated by the following formula (see Howard & Rogers, 2012, pp. 154):

$$\Delta d = \frac{D^2 * \delta}{IPD - \delta * D}$$

That is, given binocular disparity (δ) in radians, interpupillary distance (IPD), and fixation distance (D), the relative depth between two points (Δd) is approximated assuming that the observer has access to information concerning the absolute viewing distance to the object. The accuracy of depth judgements from both types of disparity signal (absolute and relative) depend on this knowledge of absolute viewing distance.

1.2 Accuracy of Stereoscopic Depth Perception

The need for a veridical estimate of absolute viewing distance to scale binocular disparities is a fundamental limitation of stereopsis. While stereopsis does provide a profound sense of depth even in the absence of distinct object features (Julesz, 1971), supports depth scaling at larger viewing distances (Allison et al., 2009; Palmisano et al., 2010), and has been shown to support accurate depth perception for objects presented at near viewing distances (less than 2m) along the midline (for review see Foley, 1980; Ono & Comerford, 1977), perceived depth from binocular disparity is often distorted. Many psychophysical studies of stereopsis report these distortions in depth magnitude estimation, particularly for virtual stimuli over a wide variety of stimuli, tasks, and viewing distances (Bradshaw et al., 1996; Brenner & Landy, 1999; Brenner & Van Damme, 1999; Glennerster et al., 1996; Johnston, 1991; Johnston et al., 1994; Todd & Norman, 2003; Willemsen et al., 2008; Witmer & Kline, 1998).

These systematic distortions typically consist of two primary types of distortion: (1) distance compression, where objects at different distances appear closer to each other than their true distances (Foley, 1980), and (2) depth underconstancy, where the depth of an object appears increasingly shallow as its viewing distance increases (Johnston, 1991). The former results in a failure to estimate an object's absolute distance, whereby the perceived distance of near objects is overestimated and the perceived distance of far objects is underestimated (Foley, 1980). As a result, the overall range of perceived

distances is compressed. This compression of visual space affects not only the perceived distances within the scene, but also the scaling of depth from binocular disparities. For instance, for a given binocular disparity, if the absolute distance to the object is overestimated, then depth is overestimated. Conversely, if absolute distance is underestimated, perceived depth is underestimated. This relationship between binocular disparity, viewing distance, and relative depth is illustrated in Figure 1.2. This pattern of depth distortions is referred to as depth underconstancy, where perceived depth estimates are overestimated at small, and underestimated at large viewing distances (Foley, 1980; Gogel, 1977; Norman et al., 1996). Thus, the compression of perceived distance gives rise to the depth underconstancy.

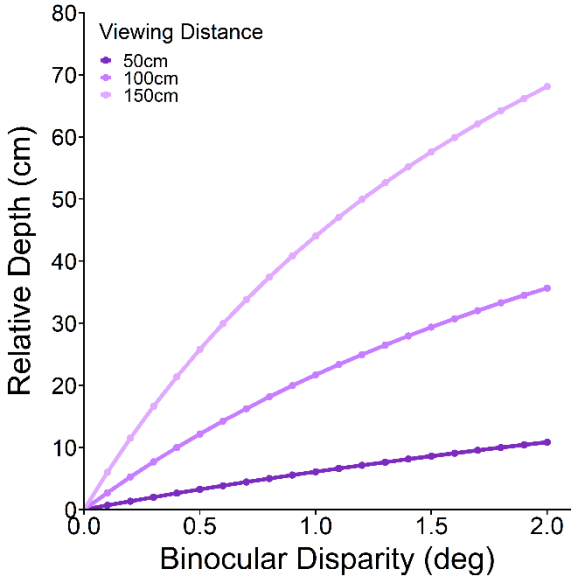


Figure 1.2. *An illustration of the relationship between binocular disparity and relative depth with changes in viewing distance. Each line represents a difference viewing distance (50cm, 100cm, or 150cm). For a given binocular disparity, the predicted relative depth increases as the viewing distance from the observer increases.*

1.2.1 Sparse Environments with Limited Cues

The first, and the most cited cause of misestimation of depth from binocular disparity is the absence or degradation of absolute distance information (Foley, 1980; Johnston, 1991; Rogers & Bradshaw, 1993). To isolate depth from binocular disparities to achieve better experimental control, other cues to depth and distance are often removed from the environment. Unfortunately, this also limits the information available to support distance information, resulting in distance information from only the pattern of vertical disparities and the vergence angle of the eyes (Foley & Richards, 1972; Foley, 1985; Rogers & Bradshaw, 1993; Wallach & Zuckerman, 1963). Unless the content fills a wide area of the visual

field (e.g., 70 deg) at viewing distances below 50cm, vertical disparity signals are very weak (Backus et al., 1999; Rogers & Bradshaw, 1995). Similarly, vergence is known to be highly unreliable and on its own provides an insufficient signal to support accurate depth estimation (Foley & Held, 1972; Gogel, 1961; 1977; Johnston, 1991; Komoda & Ono, 1974), except at distances less than 30cm (Mon-Williams et al., 2000). Thus, the removal or unreliability of these cues in sparse test environments contribute to the systematic biases in distance perception which consequently distorts stereoscopic depth perception. This is supported by other studies of stereoscopic depth constancy have shown that the addition of other monocular cues to depth and size improve the scaling of stereoscopic depth (Brenner & Van Damme, 1999; Collett et al., 1991; Foley, 1968; Mon-Williams et al., 2000).

1.2.2 Distance Information in Virtual Environments

In addition to issues caused by reduced-cue test environments, virtual environments have other potential sources of errors in distance perception. For instance, computerized display systems with a single focal plane have an inherent conflict between the accommodative distance that specifies the distance to the screen plane and vergence distance that specifies the distance to the virtual object. The resultant discrepancy between vergence and accommodative distance increases as objects are rendered further in depth from the screen plane (i.e., vergence changes substantially while accommodation remains fixed at the focal plane); particularly when the screen is placed at near viewing distances (Fry, 1939). Assessments of perceived depth using virtual stimuli with conflicting distance cues consistently show distortions characteristic of an unreliable estimate of absolute distance (Johnston, 1991; Scarfe & Hibbard, 2006), in addition to disrupting relative depth judgements and increasing discomfort (Hoffman et al., 2008). Other studies with virtual stimuli where these cue conflicts are reduced show improved depth judgments, but depth is still underestimated (Watt et al., 2005).

In contrast, when viewing physical stimuli, accommodation and vergence responses are coupled irrespective of the distance of the object. Early direct evidence of the impact of decoupling vergence and accommodation was provided by Wallach and Zuckerman (1963). Using physical wireframe targets they were able to systematically bias depth estimates using trial lenses to vary vergence and accommodation relative to the true distance of the physical object. They showed that estimates achieve near constancy at close viewing distances even when vergence and accommodation are the only distance information available (Wallach & Zuckerman, 1963). In later studies, physical stimuli without vergence accommodation conflict demonstrated near-accurate depth constancy at close viewing distances in natural viewing environments (Durgin et al., 1995; Ritter, 1977). Other studies have specifically examined

the role of focus cues and their systematic effect on perceived distance and depth (Hoffman et al., 2008; Watt et al., 2005). However, it is important to note that depth distortions can still be present in physical environments, but are often less extreme than those observed with virtual stimuli (Bradshaw et al., 2000; Cuijpers et al., 2000; Loomis & Philbeck, 1999; Tittle et al., 1995).

The lack of conflicts between oculomotor distance cues in physical test environments provides a natural viewing environment more consistent with everyday scenarios. The comparison of performance between carefully controlled virtual and physical test environments provides insight into whether hypotheses generalize to natural settings. For instance, Frisby and Buckley conducted a series of studies that evaluated the integration of texture, binocular disparity, and blur cues in virtual and physical textured surfaces. They showed that the integration of binocular disparity and texture cues depended on the orientation of the surface for virtual ridges, but no such relationship was found for physical ridges (Buckley & Frisby, 1993). They later determined that this lack of anisotropy in physical stimuli was likely due to the presence of accommodative blur (Frisby et al., 1995). These results highlight both the importance of avoiding generalizations based only on virtual stereograms and the value of using ecologically valid natural viewing environments for such experiments. One way to resolve the confounds and to identify which factors are critical to stereoscopic depth constancy, is to replicate the physical viewing environment in a virtual counterpart.

1.3 Depth Cue Integration

There are sources of ambiguity between cues and related environmental properties because of (1) noise in the measurements of the visual system (e.g., internal errors in estimating disparity, see Cormack et al., 1997), and (2) many-to-one relationships between properties of the external environment and retinal images (e.g., binocular correspondence problem, see Parker et al., 1996). Given all these potential sources of ambiguity and the fact that measured cue values will vary unpredictably across different environments, it makes sense that the visual system would combine information from several depth cues in order to estimate depth with greater precision than relying on any single cue alone (Ernst & Banks, 2002; Knill & Saunders, 2003; Landy et al., 1995).

1.3.1 Bayesian Cue Combination

To understand how the visual system integrates depth information from multiple sources with varying reliability, researchers have applied models to distal world properties to represent the combination of depth cues using Bayesian decision theory (Landy et al., 1995; Maloney & Landy, 1989). In

general, Bayesian cue combination models estimate the probability of different depth values, D being 'true' given a set of observed data, d, based on likelihoods and prior knowledge of depth cues in a given environment,

$$P(D|d) \propto P(d|D) * P(D)$$

This formula describes the probability of the depth value, D being 'true' given the observed data, $P(D|d)$, as proportional to the product of the likelihood of perceiving a depth value given the observed data, $P(d|D)$, and prior knowledge, $P(D)$, about the properties of the true depth, D, in the scene[2]. Thus, a Bayesian observer interprets imprecise depth information relying on prior experience. For each possible stimulus, the Bayesian observer considers the probability of a hypothesized depth value given the depth of the stimulus (i.e., the likelihood) and the prevalence of the stimulus from experience (i.e., the prior).

1.3.2 Linear Cue Integration

The most common approach to depth cue integration uses weighted linear cue integration (Hillis et al., 2004). Dosher et al. (1986) conducted one of the first empirical studies that modelled the interaction between multiple depth cues as a weighted linear combination. Maloney & Landy (1989) further developed this idea into a simple statistical framework for modelling the "fusion" of depth estimates from multiple sources. This method simplifies the more general Bayesian approach by making a series of assumptions. Weighted linear combination methods assume that each single depth cue is processed separately and integrated into a combined estimate with greater weight placed on the more reliable cue (Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003). In this case, it is assumed that an observer has access to unbiased estimates of a world property from each depth cue, the sensory noise associated with each depth cue is independent and their reliabilities are normally distributed, uncorrelated, and Bayesian priors are uniform and non-informative (Landy et al., 1995). If these strict conditions are met, then the combined cue estimate is calculated as a simple average of the single cue estimates (Maloney & Landy, 1989). If we have an unbiased depth cue Q_a with variance σ_a^2 and a second, independent and unbiased depth cue Q_b with variance σ_b^2, then the combined depth estimate $\hat{Q}_c$ based on these two cues is,

[2] The marginal probability, $P(d)$ (i.e., model evidence) was intentionally excluded from the denominator in this equation, as in cue combination models this parameter is commonly used as a constant, normalizing term that ensures the equation integrates to 1.

$$\hat{Q}_c = w_a Q_a + w_b Q_b$$

The weights for each cue are proportional to the inverse of the variance (i.e., reliability) of each distribution (Cochran, 1937), such that greater weight is placed on the more reliable cue (Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003). Therefore, the reliability of each cue is represented as,

$$r_a = \frac{1}{\sigma_a^2} \text{ and } r_b = \frac{1}{\sigma_b^2}$$

and the weights are represented as,

$$w_a = \frac{r_a}{r_a + r_b} \text{ and } w_b = \frac{r_b}{r_a + r_b}, \text{ respectively.}$$

In this approach, the cues are integrated linearly and optimal cue integration maximizes reliability (Ernst & Banks, 2002; Landy et al., 1995). Several studies have found that the visual system does combine depth from multiple cues in an optimal fashion by taking into account the reliability of individual cues (Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003).

1.3.2.1 Correlated Error Model

While it is often safe to assume that the variability of estimates from different cues are uncorrelated when the combination of cues is between sensory modalities (Ernst & Banks, 2002), when combining multiple visual cues that all depend on properties of the retinal image, cues could share at least one source of noise. In this case, there could be correlation or non-independence in the errors of the cue distributions (Oruç et al., 2003). The correlated error model proposed by Oruç, Maloney, and Landy (2003) adjusts the optimal reliability according to the estimated correlation (ρ) between two depth cues. This model is still a weighted linear combination, but the weights account for the covariance of the cues. The reliability of each single cue condition is corrected by $-\rho\sqrt{r_a r_b}$. Thus, this model captures if observers are using a linear combination method, but with suboptimal weights.

1.3.3 Non-linear Cue Integration

In scenarios where the visual information is noisy, incomplete, or in conflict, alternative non-linear combination methods may be used (Maloney & Landy, 1989). When two cues are highly discrepant

it suggests that something maybe wrong with one or more of the estimates, thus it makes less sense to combine them using a simple weighted average. For instance, if one cue is highly unreliable or biased, a viable strategy for the visual system may be to adopt a robustness mechanism by weighting the discrepant cue less heavily or completely vetoing it, akin to removing outliers in statistics (Landy et al., 1995; Norman & Todd, 1995). Further, if there is a large conflict between two depth cues, then averaging (as in linear integration) would inappropriate, since the combined average would be an intermediate value, inconsistent with both cues. In the veto method, the single more reliable cue is used, while the other is ignored. This behaviour is often observed in scenarios where multiple cues are presented with large conflicts, such as when disparity and perspective cues define different surfaces, resulting in a rivalrous stimulus (Girshick & Banks, 2009; Knill, 2007; Norman & Todd, 1995).

Another scenario in which the visual system may adopt a non-linear combination method is when information provided by two cues interact. One such situation is when information from one cue disambiguates information in the scene that another cue requires. One such example is the combination of binocular disparity and relative motion cues. Like binocular disparity, motion parallax can be used to determine the relative depth of objects (Rogers & Graham, 1979; Rogers & Graham, 1983). To provide an estimate of relative depth both cues must be scaled by a measure of viewing distance, where binocular disparity depends on the square of viewing distance and relative motion depends on viewing distance directly. Landy et al. (1995) referred to this conversion of cues into common units as 'cue promotion'. It has been proposed that because each cue depends differently on viewing distance, their combination could disambiguate both relative depth and viewing distance (Richards, 1985). However, empirical studies of this phenomenon are inconsistent (Brenner & Landy, 1999; Johnston et al., 1994; Landy & Brenner, 2001). Studies typically show that depth is misperceived even when both binocular disparity and motion cues are available (Scarfe & Hibbard, 2011; Todd, 1985; Todd & Norman, 2003). In other models, viewing distance is modelled as a hidden variable (i.e., a nuisance parameter in statistics) that is taken into account in the final likelihood functions (for details see Knill & Saunders, 2003).

1.3.4 Integration of Biased Depth Estimates

Given well-established empirical studies documenting the systematic distortions of perceived depth from binocular disparity, there has been much debate as to how the visual system combines depth from multiple sources of varying reliability and whether the visual system is even capable of generating unbiased estimates of 3D world properties (Todd & Norman, 2003). The assumption that the visual system has access to unbiased estimates is a fundamental assumption of linear cue integration. However,

a weighted averaging framework can successfully model the effects of perceptual biases if the model takes into account the interaction between the variances and biases of the cues (Scarfe & Hibbard, 2011).

Another common way models of depth cue integration account for biases in stereoscopic depth perception is to use a Bayesian prior that represent flatness cues. Cues to flatness were introduced by Young, Landy, and Maloney (1993) as other cues, such as accommodation or motion parallax that indicate the computerized display is flat. These residual flatness cues are commonly associated with a prior for front-parallel surfaces in limited cue situations and/or flatness cues inherent to a flat monitor (Watt et al., 2003). This prior for flatness is often modelled as a Gaussian distribution centered on zero disparity, where the standard deviation of this distribution represents the strength of the prior and the reliability of the residual flatness cues. As mentioned in Section 1.2.2, Buckley and Frisby demonstrated the important of considering flatness cues, such as accommodative focus, by comparing the combination of binocular disparity and texture cues in virtual and physical objects (Buckley & Frisby, 1993; Frisby et al., 1995). Removing the focus information via pinhole viewing produced the same perceptual distortions in physical stimuli as their virtual counterparts.

1.3.4.1 Alternative Models of Cue Integration

Some alternative models of cue integration propose that the visual system does not have access to an unbiased estimate of a world property from a single depth cue. These models suggest that instead of maximizing the reliability of a biased estimate, the goal of the visual system is to instead minimize the bias of the final combined estimate. For instance, while observers are often unaware of the bias in their sensory estimates, perceptual biases can be modelled by taking into account the interaction between the variances and biases of the cues (Scarfe & Hibbard, 2011). In theory, if observers were aware of the bias of individual cues through sensory feedback, then they could minimize the bias of their combined estimate.

1.3.4.2 Intrinsic Constraint Model

The intrinsic constraint (IC) model proposed by Domini, Caudek, and Tassinari (2006) proposes that the visual system does not estimate a metric depth map, but a scaled depth map that is related to the metric depth map through an unknown constant. The IC model proposes two successive stages of processing: (1) first stage combines image signals to estimate a local affine depth map, and (2) a second stage where the local estimates are scaled by a constant to create a global representation of metric depth. The IC model differs from a linear cue integration in several ways. First, the IC model assumes that

instead of a weighted combination of depth estimates from single cues, the visual system instead performs a weighted combination of image signals *before* a metric interpretation of depth is applied. This bypasses the need to scale cues into a common meaningful unit before cue combination. Second, instead of maximizing the reliability of an estimate, akin to linear cue integration, the IC model assumes that the visual system combines these image signals with the goal of maximizing the signal-to-noise ratio (SNR) of the combined image signal (Tassinari et al., 2008). The result is the most precise estimate of an affine depth map that produces a family a depth maps via the linear scaling of a constant (Koenderink & van Doorn, 1991; Todd et al., 2001). Domini et al. (2006) proposed that the weights are estimated by principal component analysis on each image signal scaled by the standard deviation of their measurement noise (assuming measurement noise is constant). The SNR of the combined signal is always larger than the SNR of the single image signals. If two image signals have the same SNR, then the constants for those signals have the same value and they should be perceived as matched in depth. Empirical studies of the IC model have shown that it successfully predicts the biases in the combination of disparity and velocity signals in scenarios where a linear cue integration fails (Domini et al., 2006; Domini & Caudek, 2009).

Like linear cue integration, there are some limitations to the IC model. While the IC model assumes measurement noise is constant, there are some studies that show noise varies with signal intensity. MacKenzie et al. (2008) related the IC model to Fechnerian theories of sensory scaling, where perceived depth is measured in terms of just-noticeable differences (JNDs). Their results were consistent with other studies that show a simple sum of JNDs do not predict differences in sensory magnitudes (Newman, 1933; Stevens, 1961). They conclude that while linear cue integration could account for these findings by allowing perceived depth and depth discriminability to vary independently, this is an issue for the IC model. However, Domini and Caudek (2010) showed that JNDs could predict differences in sensory magnitude using a different methodology. Thus, the IC model works well for a local analysis of image signals that vary by a small amount, such that noise variation is negligible, but is less appropriate for the global analysis of a visual scene (with large variations in signal intensity).

1.4 Current Objectives

Given the complexity of integration and interaction between multiple sources of depth information, it is difficult to determine what information is necessary to support an accurate and reliable sense of depth, especially in virtual environments. To breakdown this complex problem into manageable components it is necessary to evaluate depth perception in many diverse environments with various combinations of depth cues using wide variety of stimuli, tasks, and viewing distances. Especially given

the known distance conflicts and limitations in computerized display systems, it is critical to compare performance between natural viewing environments with rich, consistent depth cues. This is particularly important given estimation strategies may be specific to the test environment and/or task demands (Glennerster et al., 1996; Scarfe & Hibbard, 2006; Todd, 2004). The visual system may rely on alternative strategies or sources of information in constrained environments.

This dissertation consists of a series of studies that assessed the impact of monocular, binocular, and extraretinal information on perceived depth distortions for virtual objects relative to complex physical stimuli. In Chapter 2, depth judgements in a highly controlled physical test environment were compared to virtual depth judgements using two common computerized display systems used to assess stereopsis (stereoscope and head-mounted display system) at multiple viewing distances to evaluate the impact of display-based cue conflicts on depth constancy. In Chapter 3, the assessment of the accuracy and precision of depth judgements between virtual and physical objects was used to evaluate if the previous failures of linear cue combination models to fully describe the combination of binocular disparity and motion parallax cues was driven by the limitations of computerized displays systems. Given well-established depth distortions from visual stimuli, Chapter 4 investigated the possibility that non-visual cues, such as distance information from reaching in depth, can be used to aid the scaling of stereoscopic depth. This was accomplished using a ring placement task with error-based feedback that depended on the accuracy of distance judgements. The overall aim of these studies was to determine how depth information is integrated in complex viewing environments. Part of this objective focused on the impact of computerized display systems on prominent depth distortions from stereopsis by careful comparison of virtual objects to their matched physical counterparts.

CHAPTER 2: STEREOSCOPIC DEPTH CONSTANCY FOR PHYSICAL OBJECTS AND THEIR VIRTUAL COUNTERPARTS

Preface

This chapter takes advantage of a highly controlled physical test environment to determine if systematic depth distortions commonly observed in virtual test environments persist in carefully matched physical stimuli. Part of this investigation focused on the impact of display-based cue conflicts on depth constancy. Further, the use of multiple virtual test environments with different viewing geometry determined if the magnitude and direction of these conflicts systematically affect depth distortions. To accomplish these objectives virtual depth judgements were assessed using two common computerized display systems (stereoscope and head-mounted display system) and compared to judgments of full-cue physical stimuli at multiple viewing distances.

The contents of this chapter was published as a manuscript titled, *Stereoscopic depth constancy for physical objects and their virtual counterparts,* in volume 22, issue 4 of the Journal of Vision (Hartle & Wilcox, 2022). Both authors contributed the conceptualization, design of methodology, review, and editing of the manuscript. Brittney Hartle completed all programming, performing of experiments, data collection, statistical analysis, and prepared the original draft of the published work. Laurie M. Wilcox supervised, managed, and coordinated responsibility for the research.

2.1 Introduction

The ability to accurately estimate the depth and distance of objects is critical to our interpretation of, and interaction with the world around us. Not only must we assess the relative location of objects in space, but to maintain a stable 3D percept the perceived depth between relative positions should remain constant over a range of viewing distances. Such depth constancy is often reported for objects presented at near viewing distances (less than 2m) along the midline (for review see Foley, 1980; Ono & Comerford, 1977). One of the primary sources of relative depth information within near space is binocular disparity; using the positional disparity between each eye's retinal image, the observer's interpupillary distance, and the knowledge of the observer's absolute viewing distance to the object it is theoretically possible to compute metric depth. However, the results of psychophysical studies of perceived depth from binocular disparity are mixed and often report distortions in depth magnitude estimation. This is particularly true for virtual stimuli over a wide variety of stimuli, tasks, and viewing distances (Todd & Norman, 2003; Willemsen et al., 2008; Witmer & Kline, 1998).

These errors are perhaps not that surprising given that there are several potential sources of error in virtual environments. The first, and the most cited cause of misestimation is the absence or degradation of absolute distance information (Foley, 1980; Johnston, 1991; Rogers & Bradshaw, 1993). In simple test environments there is little information available to support reliable estimates of absolute distance apart from the pattern of vertical disparities and the vergence angle of the eyes (Foley & Richards, 1972; Foley, 1985; Rogers & Bradshaw, 1993; Wallach & Zuckerman, 1963). Unless the content fills a wide area of the visual field (e.g., 70 deg) at viewing distances below 50cm, vertical disparity signals are very weak (Backus et al., 1999; Rogers & Bradshaw, 1995). Similarly, vergence is known to be highly variable and on its own provides an insufficient signal to support accurate depth estimation (Foley & Held, 1972; Gogel, 1961; 1977; Johnston, 1991; Komoda & Ono, 1974; Linton, 2020), except at distances less than 30cm (Mon-Williams et al., 2000). Another potential source of error that is endemic to computerized display systems is the conflict between the accommodative distance that specifies the distance to the screen plane and vergence distance that specifies the distance to the virtual object. The resultant discrepancy between vergence and accommodative distance increases as objects are rendered further in depth from the screen plane (i.e., vergence changes substantially while accommodation remains fixed at the focal plane). Assessments of perceived depth using virtual stimuli with conflicting distance cues consistently show distortions characteristic of an unreliable estimate of absolute distance (Johnston, 1991; Scarfe & Hibbard, 2006), in addition to disrupting relative depth judgements and increasing discomfort (Hoffman et al., 2008).

In contrast, when viewing physical stimuli, accommodation and vergence responses are coupled irrespective of the distance of the object. Early direct evidence of the impact of decoupling vergence and accommodation was provided by Wallach and Zuckerman (1963). Using physical wireframe targets they were able to systematically bias depth estimates using trial lenses to vary vergence and accommodation relative to the true distance of the physical object. They showed that estimates achieve near constancy at close viewing distances even when vergence and accommodation are the only distance information available (Wallach & Zuckerman, 1963). In later studies, physical stimuli without vergence accommodation conflict demonstrated near-accurate depth constancy at close viewing distances in natural viewing environments (Durgin et al., 1995; Ritter, 1977). While the use of physical stimuli eliminates the conflict between vergence and accommodative distance, it is unclear which factors are critical for stereoscopic depth constancy. For instance, Frisby and Buckley conducted a series of studies that evaluated the integration of texture, binocular disparity, and blur cues in virtual and physical textured surfaces. They showed that the integration of binocular disparity and texture cues depended on

the orientation of the surface for virtual ridges, but no such relationship was found for physical ridges (Buckley & Frisby, 1993). They later determined that this lack of anisotropy in physical stimuli was likely due to the presence of accommodative blur (Frisby et al., 1995). These results highlight both the importance of avoiding generalizations based only on virtual stereograms and the value of using ecologically valid natural viewing environments for such experiments. One way to resolve the confounds and to identify which factors are critical to stereoscopic depth constancy, is to replicate the physical viewing environment in a virtual counterpart.

Another important consideration for such experiments is *how* depth is estimated. For instance, in Frisby and Buckley's series of studies observers were trained to use a response scale using physical stimuli to perform depth judgements. As a result, observers always made depth estimates relative to the richer and more reliable physical test environment. This then limited their comparisons to the relative differences between virtual and physical judgements. In other experiments, simultaneous matching-tasks have shown that the relative depth from binocular disparity is accurate at viewing distances under 2m (Glennerster et al., 1996), while other matching-tasks have shown consistent depth distortions at near viewing distances (Scarfe & Hibbard, 2006). However, matching tasks of this type allow observers to minimize disparity differences between the target and reference stimuli and do not assess or reflect perceived depth. As a result, matching tasks do not convey the perceived magnitude of a percept, only the given perceptual magnitude that is equivalent to another (Foley et al., 1975). One way to avoid pitfalls associated with disparity-matching is to require that observers generate depth magnitude estimates, that is, for a given egocentric distance indicate 'how far' or 'how much' depth they perceive. Several methods can be used for this purpose, such as manual-pointing tasks (Foley et al., 1975), depth interval bisection tasks (Ogle, 1952b, 1953), ruler adjustment (Tsirlin et al., 2012), or haptic matching tasks (Brenner & Van Damme, 1999; Hornsey et al., 2020). We have developed a generative haptic method using a custom-built sensor strip that allowed us to assess the accuracy of depth estimation without introducing additional visual stimuli or relying on disparity matching (Hartle & Wilcox, 2016).

2.1.1 Current Study

The aim of this series of experiments was to unify the literature on stereoscopic depth constancy by evaluating the impact of other depth cues on suprathreshold percepts from stereopsis. We assessed depth constancy for virtual and physical stimuli in the presence of monocular and binocular depth cues. The comparison of virtual and physical stimuli allowed us to evaluate the impact of display-based cue conflicts (between accommodative and vergence distance) inherent to computerized displays on depth

judgements[3]. Three display environments were used to measure distortions (or lack thereof) of perceived depth; (1) mirror stereoscope, (2) HMD, and (3) a full-cue physical viewing environment using a purpose-built physical test environment (PTE). The comparison between the mirror stereoscope and HMD showed the extent to which the geometric distortions caused by the HMD optics impact the scaling of perceived depth. Stimuli consisted of virtual textured half-cylinders and geometrically identical physical stimuli at two viewing distances (83cm and 130cm) under monocular and binocular viewing conditions. The use of well-matched virtual and physical stimuli, an intuitive response method, and within-subject comparison allows us to reduce or eliminate the impact of differences in depth information and cue conflicts between environments, biases in measurement methods, and interobserver differences seen in past studies.

2.1.2 Rationale

To assess the relative impact of display-based cue conflicts we controlled the information present in each environment by replicating the physical viewing environment in the virtual counterparts. If the absence of conflict between ocular distance cues in a physical environment improves the accuracy of absolute distance, then depth scaling should be more accurate when physical objects are viewed binocularly. Further, comparison of monocular and binocular viewing conditions provides insight into the utility of monocular cues on their own and in combination with binocular depth information. The depth scaling in the monocular viewing condition should be significantly more shallow relative to binocular viewing due to a lack of binocular distance cues (e.g., vergence and vertical disparities). To analyze the changes in depth scaling in each binocular condition, a measure of inferred viewing distance (see Section 2.4) based on the slope of depth judgements was used to represent the observer's assumed viewing distance in each condition. To evaluate the absolute accuracy of depth scaling, the slopes of the functions fit to each observers' depth judgements were compared to the slope of theoretical predictions using an ideal observer model. Considered together these two analyses describes changes in depth scaling by comparing a measure of inferred viewing distance to the actual viewing distance in each condition. Given the distance of objects at eye level tend to be overestimated at relatively near distances within peripersonal space and underestimated at larger distances (Foley, 1985; Gogel & Tietz, 1979), observers should exhibit underestimates at both viewing distances in the current study (given they are further than peripersonal space). In addition, given the scaling of binocular disparities is further reduced at larger

[3] Depth constancy refers to the ability to maintain the consistency of depth judgements (along the z-dimension) when viewing distance is varied. Depth scaling refers to the incremental change in depth judgements as a function of binocular disparity (i.e., slope).

viewing distances, observers should be more accurate at the near relative to the far viewing condition. If observers achieve depth constancy in any viewing condition, then the magnitude of depth judgements should remain constant as viewing distance varies. Thus, if the intercepts and slopes were equivalent in the near and far viewing conditions, then the two linear functions (and the magnitude of depth judgements) would be considered equivalent.

Comparison of results obtained in the two virtual viewing environments (i.e., stereoscope and HMD) under monocular viewing will provide insight into the potential impact of the optics of the HMD system or cognitive factors (e.g., knowledge of the display on your head). The distortion correction applied to commercial HMD systems assume a fixed forward gaze angle to limit the potential influence of prism distortions from looking off-axis (Mon-Williams et al., 1993; Ogle, 1952a). While this distortion correction eliminates lens distortions for forward fixation, at large eccentric gaze angles optical distortions would be apparent in the periphery. If depth estimation accuracy is similar in the high-resolution virtual environment of the mirror stereoscope and the HMD, then we can conclude that the HMD does not distort depth in these stimuli. The latter issue is an important concern as HMDs are increasingly being used as 3D display systems for vision science (Scarfe & Glennerster, 2019). In addition, as outlined in the Apparatus section, our virtual displays had focal distances of 200cm (HMD) vs. 74cm (stereoscope). We capitalized on this difference to assess the impact of vergence accommodation conflict over this commonly used range of distances. That is, if increasing conflict affects the magnitude of distance estimates then depth scaling should be more accurate at near distances in the stereoscope and at far distances in the HMD conditions in binocular viewing conditions.

2.2 Methods

2.2.1 Observers

Sixteen observers were recruited from York University. To ensure observers could detect depth from binocular disparities of at least 40 arcseconds we used a Randot^{TM} stereoacuity test to screen participants prior to testing. All observers had normal to corrected-to-normal vision, and if necessary, wore their corrected lenses during testing. The research protocol was approved by York University's Research Ethics Board.

2.2.2 Stimuli

The dimensions of the half-cylinders were equated in the three viewing conditions, (1) mirror stereoscope, (2) HMD, and (3) the PTE. To match the 3D structure across conditions, the size and

binocular disparity of virtual cylinders were scaled to match the changes in visual angle of the physical cylinders at each viewing distance. All cylinders had a fixed height and width of 14cm, which subtended 9.6deg and 6.2deg at the near and far viewing distances, respectively. The distance along the z-dimension from the reference frame to the peak of the half-cylinder (i.e., the depth of the surface) were 1, 3, 5, 7, and 9cm. However, the 1cm condition was excluded from the physical viewing condition due to the presence of shadows that could not be eliminated with our lighting configuration. For virtual viewing conditions, the disparity of the cylinder's peak was calculated using each observer's interpupillary distance and the conventional formula (see Howard & Rogers, 2012, pp.152-154).

Each cylinder was textured with a random array of non-overlapping circular elements (Figure 2.1). The textured planar surface was deformed when placed on the curved surface of the cylinder. The aspect ratio and density of the circular elements provided observers with additional monocular cues to surface curvature to help them localize the position of the edges and peak of the cylinder (Blake et al., 1993; Cumming et al., 1993). The textures were generated in MATLAB. The radius of the circular elements ranged from 0.38 to 0.95deg at the near viewing distance, and 0.24deg to 0.60deg at the far viewing distance. The luminance of the circular elements had a positive or negative polarity relative to the background luminance with Michelson contrasts that ranged from 0.03 to 0.34. For the physical cylinders, a set of textures were printed on matte heavyweight paper with a flat finish and zero glare. The luminances of the physical and virtual cylinders were 52.2 and 50.3 cd/m^2, the small (perceptually indistinguishable) difference was due to a slight change in the setup between testing. The luminance of the background in the virtual viewing conditions was adjusted to match the contrast between the edge of the cylinder and the background in the PTE. To randomize the textures in the PTE, the textures were changed between observers and the cylinders were randomly rotated by 180deg between blocks to make the fixed position of the texture elements an uninformative reference for depth judgements.

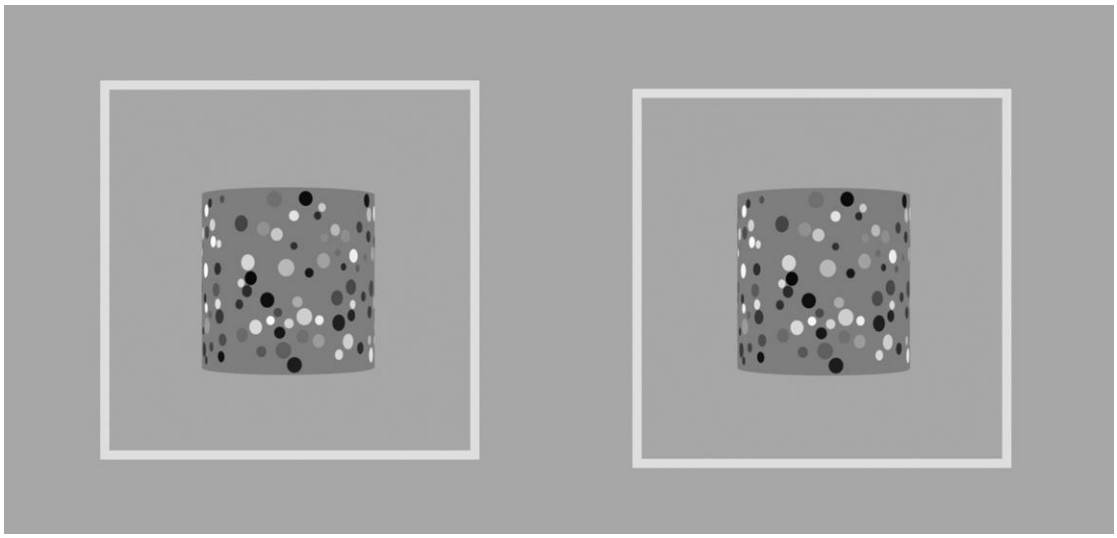


Figure 2.1. *A stereopair of a textured half-cylinder stimulus. The stereopair is arranged for crossed fusion. The cylinder and reference frame are not to scale.*

All cylinders were rendered at eye level in the center of the observer's field of view on a grey background surrounded by a reference frame (65.6 cd/m^2). The reference frame helped observers localize the coronal plane of the cylinder and served as a reference for observer's depth magnitude judgements. The frame subtended 21.8deg and 14.0deg at the near and far viewing distances, respectively. The distance between the edge of the cylinder and the inner edge of the frame was 5.5deg and 3.5 deg at each respective viewing distance. A standing disparity was added to the half-cylinder and reference frame in the stereoscope viewing condition to ensure the stimuli appeared at viewing distances of 83cm and 130cm for the near and far viewing distances. This manipulation changed the amount of conflict between the accommodation and vergence signals in the virtual viewing conditions. In the PTE, the horizontal actuator was moved in depth between each session.

In the monocular condition, cues such as texture (e.g., density and aspect ratio) and the curvature of the top and bottom edge of the cylinder help to define the shape of the surface. When these cues are combined with an estimate of distance, they provide information regarding the depth of the surface (along the z-dimension). Further, the size-distance scaling at the two viewing distances provides relative distance information. While focal blur is available and could aid estimates in the physical environment, the two viewing distances (83cm and 130cm) were chosen so any focal differences between the edge and peak of the surface should be perceptually indistinguishable (D. M. Hoffman & Banks, 2010; Watt et al., 2005). That is, the difference in focal blur between the reference frame and the peak of the surface at the largest depth of 9cm would be approximately 0.15D and 0.06D at the near and far viewing distances, respectively. Given the eyes' depth of focus under typical viewing scenarios is

approximately 0.33D (Campbell, 1957; D. M. Hoffman & Banks, 2010; Walsh & Charman, 1988), these features should appear equally sharp. Thus, the information from texture, edge curvature, and focal blur are roughly equivalent for both the virtual environments and the PTE. When viewed monocularly, the critical difference between physical and virtual viewing was the accuracy of the absolute distance signaled by accommodation. Unlike natural viewing, in the virtual viewing environments accommodative distance is fixed to the focal distance of the device. While vergence eye movements were likely made under monocular viewing, such movements are substantially degraded and therefore less reliable in monocular relative to binocular viewing (Erkelens, 2000; Gibaldi & Banks, 2019). In the binocular viewing condition, binocular disparity (e.g., horizontal and vertical disparities), and vergence cues were available to aid depth estimates. The amount of relative binocular disparity along the surface was equivalent in the physical and virtual viewing environments. While horizontal disparities provide information regarding surface curvature, vertical disparities provide information regarding the absolute distance to the object, though this information has been shown to be available primarily for stimuli that fill a large visual field (e.g., 70 deg) at viewing distances below 50cm (Backus et al., 1999; Bradshaw et al., 1996; Rogers & Bradshaw, 1995). Thus, vertical disparities are unlikely to play a significant role in our study.

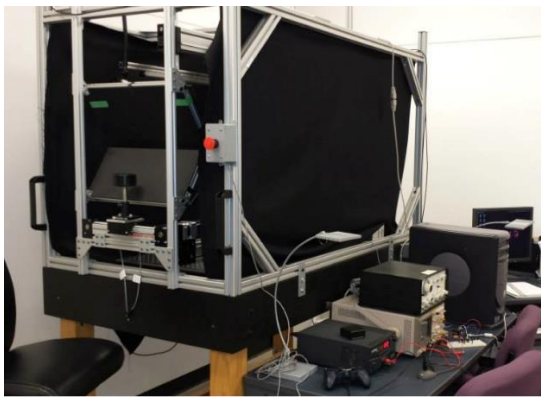
2.2.3 Apparatus

The virtual cylinders were generated and displayed in two virtual environments, (1) mirror stereoscope, and (2) HMD. In the mirror stereoscope, cylinders were generated using OpenGL 3D graphics within the Psychtoolbox package (Brainard, 1997; Pelli, 1997) for MATLAB on a Mac OSX computer. The modified Wheatstone mirror stereoscope consisted of with two LCD monitors (Dell U2412M) with a resolution of 1920 by 1200 pixels and a refresh rate of 75Hz. Each monitor had a viewing distance of 74cm from the observer and a horizontal field-of-view of 25deg. At this resolution and viewing distance each pixel subtends 1.26 arcmin of visual angle. The geometry of OpenGL's projection matrix was designed to replicate the viewing geometry of our modified Wheatstone mirror stereoscope to ensure the two frustums converge at a distance equivalent to the stereoscope's screen plane. The horizontal offset of each stereopair at the virtual screen was derived from the observer's interpupillary distance to ensure the correction representation of each monocular image across observers.

In the HMD condition, the virtual cylinders were generated with the same dimensions as the 3D models in Unity version 5.6.1 using a Windows 10 computer with an NVIDIA GeForce GTX 1080 graphics card. The images were presented using an Oculus Rift CV1 HMD. The Oculus Rift has two organic light-emitting diode displays, each with a resolution of 1080 by 1200 pixels per eye with a refresh rate of 90Hz

and focal distance of approximately 200cm. At a horizontal field-of-view of 94deg, each pixel subtends 4.7 arcmin of visual angle (assuming an equal distribution across the field-of-view). Prior to testing the interpupillary distance of the lenses was adjusted to match each observer's interpupillary distance (rounded to the nearest millimetre). Observers rested their head on a fixed chin rest to stabilize their head position.

The physical stimuli were presented in a computer-controlled environment using our PTE (Figure 2.2, see also Hartle & Wilcox, 2021). This apparatus consists of a collection of linear actuators (Macron Dynamics MGA-M6S) mounted on an optical bench within a light-tight enclosure. Each linear actuator has a positional repeatability of +/- 0.025 mm and a positional error of 0.4mm per metre of travel. Each actuator was driven by a stepper motor controlled by a Galil DMC-4050 motion controller. Stimulus visibility was controlled via the computerized LED lighting within the PTE. The lighting setup minimized shadows and shading along the surface of the half-cylinders. Physical cylinders were 3D printed using a LulzBot TAZ 6 3D printer with the same dimensions as their virtual counterparts. Physical cylinders were mounted on a 3.8cm thick polystyrene board (122cm by 61cm) using magnets embedded in the board and in the flat face of the printed cylinders. A matte heavyweight paper poster was printed and glued to the polystyrene board. The poster displayed a uniform grey background (72.6 cd/m^2) and three reference frames with the same dimensions as the virtual frames. Each cylinder could be mounted and replaced in the center of each reference frame. The board was mounted onto the horizontal linear actuator in the PTE, which moved the cylinders into position between trials. A fixed chin rest was attached to the front of the PTE to stabilize the observer's head position. An adjustable square aperture was placed 50cm in front of the observer to limit the horizontal field of view to 29deg and 19deg in the near and far viewing conditions, respectively. The edge of the aperture obscured the observer's view of the adjacent cylinders on the board.



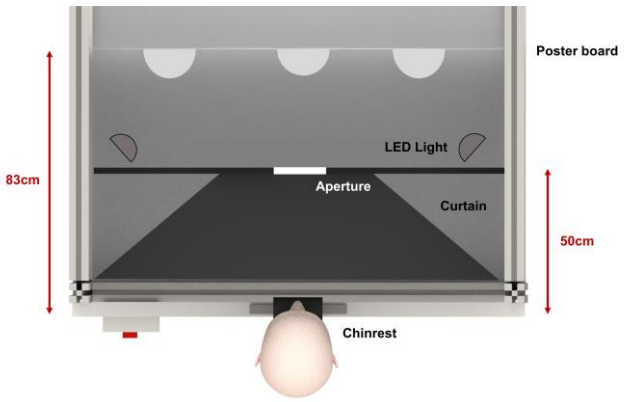


Figure 2.2. The left image shows a picture of the PTE apparatus. The right image is a top-down illustration of the PTE apparatus at the near viewing distance. The poster board is shown 83cm from the observer. An aperture made from a black poster board was positioned 50cm from the observer between the LED light fixtures and the enclosure.

2.2.4 Procedure

Observers were asked to estimate the depth of the surface peak relative to the reference frame in the (1) mirror stereoscope, (2) HMD, and (3) PTE conditions under monocular (left eye patched) and binocular viewing. In all conditions, depth was estimated using a previously validated custom-built pressure-sensitive strip. We have shown previously that measurement methods that use finger displacement (either via sensory strip or direct measurement) are as accurate as methods that use a visual reference, such as a ruler (Hartle & Wilcox, 2016). To make their estimates, observers rested their thumb against a knob at one end of the sensor strip and pressed their index finger along the length of the sensor to indicate the magnitude of perceived depth. The stimulus remained visible until observers submitted their response via a button press. Following practice trials, all observers completed the monocular viewing condition first to avoid order effects caused by the cue-rich binocular viewing conditions. To compensate for order effects, half of the observers completed the PTE condition first, while the other half completed the two virtual conditions first.

The data was analyzed using a linear mixed-effects model using the nlme package in R (Pinheiro et al., 2015) that examined the individual differences in depth estimates using nested random intercepts. This model accounts for repeated-measure variables using random intercepts arranged in a hierarchy. The model was fit using maximum likelihood estimation. A likelihood ratio chi-square test determined the significance of fixed effects (slope of depth estimates, viewing distance, viewing apparatus, viewing condition, and their interactions). The structure of the nested random intercepts was chosen a priori based on the nested design of the experimental conditions (i.e., two viewing distances within three viewing apparatuses within two viewing conditions). Planned a priori comparisons for each fixed effect were evaluated using t tests. An approximation of Pearson's correlation coefficient (r) was used as a measure of effect size (Field et al., 2012). The analysis focused on the comparison of the slope of the functions (estimated vs. predicted depth) obtained at two viewing distances (near and far), the two viewing conditions (monocular and binocular), and the three viewing environments (mirror stereoscope, HMD, and PTE). The slope of depth estimates for each observer was then used to estimate the inferred viewing distance in each environment and fit each observer's data using linear regression. A maximum likelihood estimation (MLE) method was used to estimate inferred viewing distance for each observer and condition according to the following conventional formula for perceived depth that relates interpupillary distance, binocular disparity, and viewing distance (see Howard & Rogers, 2012, pp. 154):

$$\Delta d = \frac{D^2 * \delta}{IPD - \delta * D}$$

Given the binocular disparity of the surface peak (δ), interpupillary distance (IPD), and perceived depth judgements (Δd) were predetermined for each observer, the slope of each observer's function was determined by their estimate of inferred viewing distance (D). Thus, the estimate of inferred viewing distance from each observer's function provides insight into their assumed absolute viewing distance in each viewing condition. We evaluated the differences in perceived viewing distance in each viewing environment by assessing differences in inferred viewing distance for the binocular condition only. For instance, if the cue rich physical stimuli improve the accuracy of absolute distance perception, then the depth scaling in the physical environment will be steeper relative to the virtual environment.

2.3 Results

The linear mixed-effects model with planned comparisons was run as outlined above with post-hoc tests to evaluate the impact of test order. Our expectation was that the counterbalancing of virtual and physical test conditions would control for any impact of order on observers' depth estimates. Instead, we found that this factor had a substantial effect on the depth magnitude estimates and, given the effect size, it was necessary that we consider all the results in the context of which type of condition was tested first. For completeness, the results of the original analysis (independent of condition order) are included in Appendix 2.A, Tables 2.A1 and 2.A2 along with summary plots of the monocular and binocular data (Figures 2.A1 & 2.A2 respectively). The data was re-analyzed using a linear mixed-effect model that included a between-subject variable for condition order. This variable split the observers into two groups based on whether they completed the depth estimation task for physical or virtual half-cylinders first (n=8 for each group). The 4-way analysis (slope of depth estimates, type of apparatus, viewing distance, and order factors) was applied to both the monocular and binocular datasets. This analysis also allowed us to evaluate the individual three-way interactions between test order, type of apparatus, and viewing distance on the slope of depth judgements. The analysis was conducted separately for the monocular and binocular viewing conditions.

2.3.1 Monocular Viewing

In the monocular viewing condition, the analysis revealed a lack of significant 4-way interaction showed that the relationship between the predicted and perceived surface depth did not depend on the type of apparatus, viewing distance, and condition order, $X^2(28)=0.33$, $p=0.85$. However, the analysis did show a significant 3-way interaction that suggested the order of test conditions impacted the slope of depth estimates for the different types of apparatuses, $X^2(23)=9.75$, $p<0.01$. No significant 3-way interaction was found between the slope of depth estimates, viewing distance, and the order of conditions, $X^2(24)=1.75$, $p=0.19$. This pattern of results suggests that the while the condition order effected the slope of depth estimates in the different apparatuses, the relative difference in slope for the apparatuses across the two types of observers was the same at both viewing distances. As outlined above, to examine the effect of condition order on the slope of depth estimates, the analysis was split between the virtual-first and physical-first observers.

2.3.1.1 Virtual-First

Figure 2.3 shows depth estimates as a function of predicted depth for each apparatus and viewing distance under monocular viewing for observers that completed the virtual condition first. As expected, when only monocular cues were available, perceived depth was greatly underestimated with weak scaling of perceived depth with surface depth. To assess the impact of the type of apparatus and viewing distance, planned contrasts compared the slope obtained for each set up independent of viewing distance, and for the two viewing distances independent of apparatus. The contrasts for the virtual-first observers revealed that there was no significant difference in the slope of depth estimates for the two virtual apparatuses, $b=-0.01$, $t(170)=-0.15$, $p=0.88$, $r=0.01$. However, these slopes were significantly greater for physical stimuli than either the stereoscope, $b=-0.10$, $t(170)=-2.18$, $p=0.03$, $r=0.16$, or HMD apparatuses, $b=-0.10$, $t(170)=-2.06$, $p=0.04$, $r=0.16$. There was no significant difference in the slope of depth estimates in the near and far viewing distances, $b=-0.07$, $t(170)=-1.32$, $p=0.19$, $r=0.10$.

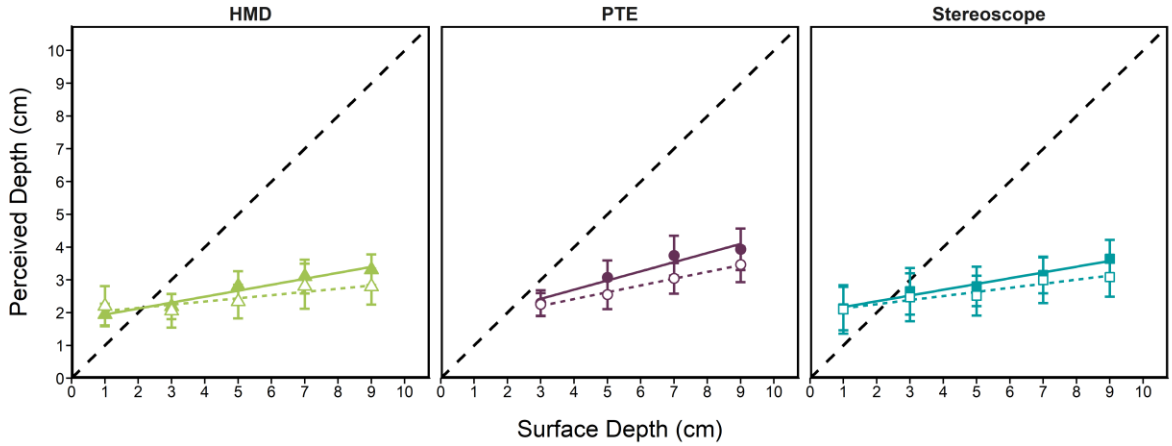


Figure 2.3. *Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under monocular viewing conditions for virtual-first observers (n=8). The dashed line represents accurate depth estimates and error bars represent the standard error of the mean.*

2.3.1.2 Physical-First

Figure 2.4 depicts estimated depth as a function of predicted depth for each apparatus and viewing distance under monocular viewing for observers that completed the physical condition first. As in the virtual-first condition, depth was substantially underestimated, and the resultant slopes were shallow. The contrasts for the physical-first observers revealed a similar pattern across the apparatuses as the virtual-first observers when only monocular cues were available. The slope of depth estimates in the two virtual apparatuses were not significantly different, b=0.12, t(170)=1.86, p=0.06, r=0.14. The slope of the depth estimates for physical half-cylinders were significantly greater than the HMD condition, b=-0.28, t(170)=-3.43, p=0.001, r=0.25, but not the stereoscope condition, b=-0.15, t(170)=-1.92, p=0.06, r=0.15. In addition, the slopes in the near and far viewing distances were not significantly different for observers that completed the physical condition first, b=-0.10, t(170)=-1.09, p=0.27, r=0.08.

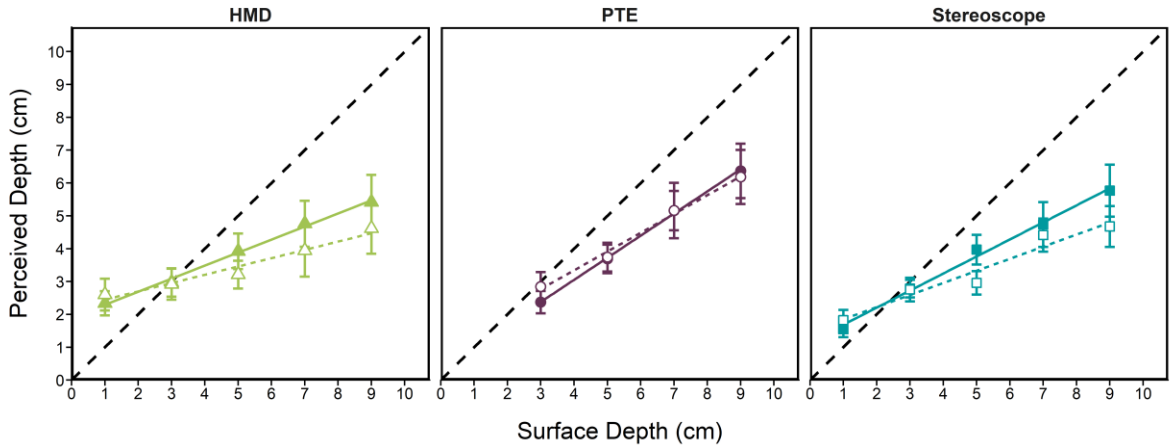


Figure 2.4. *Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under monocular viewing conditions for physical-first observers (n=8). The dashed line represents accurate depth estimates and error bars represent the standard error of the mean.*

2.3.1.3 Summary

When only monocular information was available the relative differences for the apparatuses within each observer group were similar. The slope of depth estimates was higher for physical stimuli relative to the two virtual apparatuses (with exception for the stereoscope condition for physical-first observers), while the slopes were equivalent for the two virtual conditions. Within the observer groups there was no relative difference in slope for the two viewing distances. However, comparison of Figures 2.3 and 2.4 show that that for all three conditions, observers that estimated the depth of physical stimuli first demonstrate better depth scaling and were more accurate than observers who estimated depth of virtual stimuli first. Our analysis showed that the slope of depth estimates for physical-first observers was significantly steeper overall relative to virtual-first observers, $b=-0.39$, $t(340)=-5.18$, $p<0.0001$, $r=0.27$. Contrasts confirmed that the slope was steeper for physical-first observers for depth estimates for physical stimuli, $b=-0.38$, $t(110)=-6.21$, $p<0.0001$, $r=0.51$, the stereoscope, $b=-0.29$, $t(142)=-6.69$, $p<0.0001$, $r=0.49$, and HMD conditions, $b=-0.19$, $t(142)=-4.30$, $p<0.0001$, $r=0.34$, as well as the near, $b=-0.30$, $t(206)=-6.26$, $p<0.0001$, $r=0.40$, and far viewing distance, $b=-0.25$, $t(206)=-4.75$, $p<0.0001$, $r=0.31$.

2.3.2 Binocular Viewing

Unsurprisingly, compared to the monocular conditions, when binocular cues were available perceived depth estimates were more accurate (see Figures 2.5 and 2.6). Unlike the monocular analyses, the analysis of the binocular data revealed a significant 4-way interaction showed that the relationship between the predicted and perceived surface depth depended on the type of apparatus, viewing distance, and condition order, $X^2(28)=10.94$, $p<0.01$. This suggests that the relative differences in the slopes of depth estimates for the two observer groups differs significantly as a function of the type of apparatus and viewing distance. As discussed above, the virtual-first and physical-first data was analyzed separately to better understand their impact on and interaction with the other variables.

2.3.2.1 Virtual-First

Figure 2.5 shows depth estimates as a function of predicted depth for each apparatus and viewing distance in the binocular viewing condition for observers that estimated the depth of virtual half-cylinders first. For these observers, there was no significant difference in the slope of depth estimates for

physical stimuli and the stereoscope, b=-0.12, t(170)=-1.84, p=0.07, r=0.14. The slope of depth estimates was significantly more shallow for virtual stimuli in the HMD relative to physical stimuli, b=-0.24, t(170)=-3.68, p<0.001, r=0.27. However, unlike the monocular condition, the slope of depth estimates for virtual stimuli in the HMD were significantly more shallow than in the stereoscope, b=-0.12, t(170)=-2.25, p=0.03, r=0.17. In addition, the slope of depth estimates at the far viewing distance were significantly more shallow compared to the near viewing distance, b=-0.20, t(170)=-2.63, p=0.01, r=0.20.

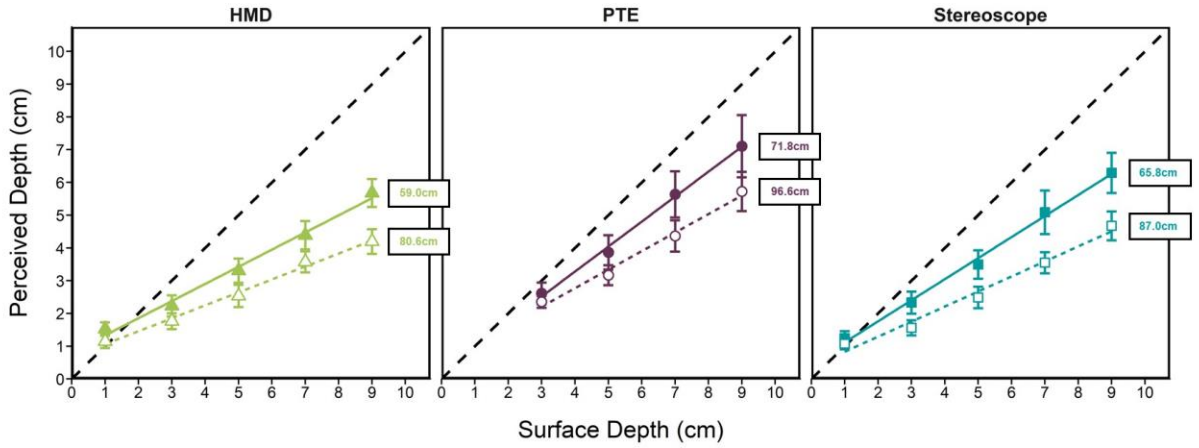


Figure 2.5. *Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under binocular viewing conditions for virtual-first observers (n=8). The inferred viewing distance is annotated for each condition (in cm). The dashed line represents accurate depth estimates and error bars represent the standard error of the mean.*

2.3.2.2 Physical-First

Figure 2.6 depicts depth estimates as a function of predicted depth for each apparatus and viewing distance in the binocular condition for observers that viewed physical stimuli first. When binocular cues were available, observers that completed the depth estimation task for physical stimuli first showed a different pattern of results compared to observers that completed the task for virtual stimuli first. There was no significant difference in slopes obtained using physical stimuli vs. the stereoscope, b=-0.09, t(170)=-1.03, p=0.30, r=0.08, or vs. the HMD apparatus, b=-0.15, t(170)=-1.81, p=0.07, r=0.14. Further, there was no significant difference in the slopes obtained using the two virtual apparatuses, b=0.07, t(170)=0.96, p=0.34, r=0.07. Similarly, there was no significant change in slopes as a function of viewing distances, b=-0.01, t(170)=-0.10, p=0.92, r=0.01. In brief, when binocular information is available, the slope of depth estimates for observers that viewed the physical stimuli first are the same, regardless of the type of apparatus or viewing distance.

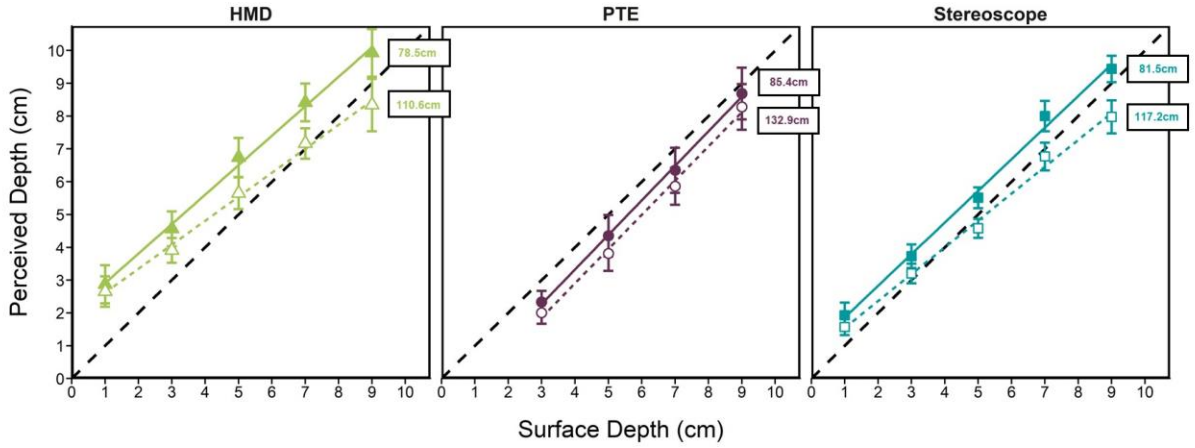


Figure 2.6. Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under binocular viewing conditions for physical-first observers (n=8). The inferred viewing distance is annotated for each condition (in cm). The dashed line represents accurate depth estimates and error bars represent the standard error of the mean.

2.3.2.3 Summary

When binocular cues were available, there was a clear difference in the pattern of results depending on whether the physical or virtual stimuli were seen first. If the depth estimation task for physical stimuli was completed first, then the slopes of depth estimates were equivalent regardless of the type of apparatus or the viewing distance. However, if the depth estimation task for virtual stimuli was completed first, then the slopes of depth estimates were significantly steeper for physical stimuli relative to virtual stimuli, and for stimuli at the near relative to the far viewing distance. In addition to these relative differences between the two groups of observers, we confirmed that the slope of depth estimates for physical-first observers were significantly steeper than virtual-first observers under binocular viewing, $b=-0.29$, $t(340)=-3.32$, $p=0.001$, $r=0.18$. This difference was consistent for the physical, $b=-0.38$, $t(110)=-5.21$, $p<0.0001$, $r=0.44$, the stereoscope, $b=-0.34$, $t(142)=-6.39$, $p<0.0001$, $r=0.47$, and the HMD apparatuses, $b=-0.36$, $t(142)=-6.61$, $p<0.0001$, $r= 0.49$. In addition, the slope of estimates for physical-first observers were also significantly steeper than virtual-first observers at the near, $b=-0.27$, $t(206)=-4.15$, $p<0.0001$, $r=0.28$, and far viewing distances, $b=-0.32$, $t(206)=-6.21$, $p<0.0001$, $r=0.40$. Thus, like the monocular condition, observers that estimated the depth of physical half-cylinders first had consistently better depth scaling than observers that completed the virtual stimulus condition first in all test conditions.

2.3.2.4 Depth Scaling

We took advantage of binocular viewing geometry to represent the changes in depth scaling (i.e., slope of depth estimates) as a measure of viewing distance. The well-known relationship between perceived depth, binocular disparity, viewing distance, and interpupillary distance was used to estimate the inferred viewing distance of the observer's depth estimates. Although the relationship between binocular disparity and perceived depth is non-linear, given the narrow range of disparities presented in our study (0.5deg maximum), we could reliably fit a linear regression line to each observer's data. For example, given the observer's depth estimates, the binocular disparity at the peak of the half-cylinder, and the observer's interpupillary distance, we determined a maximum likelihood estimate of viewing distance and fit the resulting line to each observer's data. This method allowed us to estimate each observer's *inferred* viewing distance for each type of apparatus and displayed viewing distance. Figures 2.5 and 2.6 show the inferred viewing distance in centimetres for each linear fit, and Figure 2.7 shows the individual inferred viewing distance estimates for each observer group in all viewing conditions.

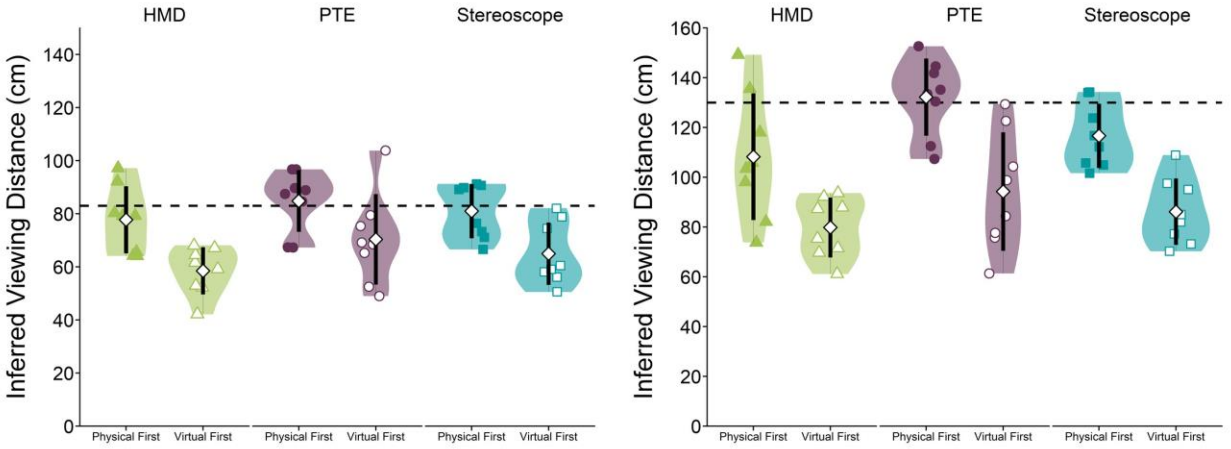


Figure 2.7. *Inferred viewing distance estimates for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near (left plot) and far viewing distances (right plot), for the physical-first and virtual-first observers (filled and open symbols, respectively). The dashed line represents the viewing distance to the reference frame in the near and far conditions (83cm and 130cm, respectively). The white diamond represents the mean and the black rectangle represents the standard error of the mean. The shaded distribution represents a density estimation that was fit using a Gaussian kernel with a smoothing bandwidth using Silverman's rule-of-thumb (or 0.9 times the minimum standard deviation and interquartile range divided by 1.34 times the sample size to the negative one-fifth power). This density estimation is plotted twice, once on each side of the boxplot for each condition.*

To evaluate the accuracy of depth scaling, the slope of depth estimates must be compared to the slope of theoretical predictions (i.e., the dashed line in Figure 2.7). To do so, we compared the slope of depth estimates to ground truth by creating an ideal observer model using randomly generated data. We set the accuracy of the ideal observer to the true depth of each half-cylinder and matched the standard

error of the generated data to that obtained from our observers at each cylinder depth. The ideal observer model was compared to the data for each type of observer for all apparatuses and viewing distances using the same linear mixed effect model as the main analysis above. The results of this analysis are shown in Table 2.1.

Table 2.1: *Accuracy of Depth Scaling Relative to Ideal Observer Model*

	Estimate	DF	*t*	*p*	*r*
Virtual-First					
Near Viewing Distance					
PTE	-0.24	46	-1.50	0.14	0.22
Stereoscope	-0.36	62	-4.61	<0.0001	0.51
HMD	-0.48	62	-5.26	<0.0001	0.56
Far Viewing Distance					
PTE	-0.43	46	-3.04	0.04	0.41
Stereoscope	-0.54	62	-8.36	<0.0001	0.73
HMD	-0.60	62	-7.79	<0.0001	0.70
Physical-First					
Near Viewing Distance					
PTE	0.05	46	0.34	0.73	0.05
Stereoscope	-0.04	62	-0.44	0.66	0.06
HMD	-0.10	62	-1.00	0.32	0.13
Far Viewing Distance					
PTE	0.04	46	0.30	0.76	0.04
Stereoscope	-0.18	62	-2.64	0.01	0.32
HMD	-0.27	62	-2.82	0.01	0.34

Note. Each test compares the slope of the ideal observer model to the slope of the data.

Table 2.1 shows that the depth scaling for observers that completed the virtual conditions first was only consistent with theoretical predictions for physical stimuli presented at the near viewing distance. Virtual stimuli presented at the near viewing distance, and all stimuli at the far viewing distance showed significant deviations from the theoretical slope. For observers that estimated the depth of physical stimuli first, the slope of depth estimates for all three conditions were consistent with theoretical predictions at the near viewing distance. At the far viewing distance, only the slope of estimates for physical stimuli was consistent with the theoretical prediction.

Overall, the slope of depth estimates for physical stimuli was always consistent with theoretical predictions, irrespective of test order, at the near viewing distance. However, this equivalence was only

seen at the far distance for observers that performed the physical depth estimation task first. When the virtual condition was tested first the slopes were consistently shallower than predicted for all virtual test conditions. While observers that completed the physical test condition first showed slopes consistent with theoretical predictions for virtual stimuli at the near viewing distance, but not for the far viewing distance.

2.3.2.5 Stereoscopic Depth Constancy

Depth constancy is said to occur when perceived depth magnitude is constant across viewing distance. Note that the estimates need not be veridical, there may be a constant offset at all distances. To determine if observers attained depth constancy, we compared the intercepts and slopes of each function between the near and far viewing distances for each group of observers. For observers that estimated the depth of physical half-cylinders first, the slopes of depth estimates for virtual stimuli in the stereoscope, b=-0.15, t(62)=-2.33, p=0.02, r=0.28, and HMD, b=-0.17, t(62)=-2.05, p=0.04, r=0.25, were significantly different at the near and far viewing distances. However, for physical stimuli the intercept, b=-0.38, t(7)=-0.67, p=0.52, r=0.25, and slope, b=-0.01, t(46)=-0.12, p=0.91, r=0.02, of depth estimates were consistent at the two viewing distances. For observers that viewed virtual half-cylinders first, the slopes of perceived depth estimates for the stereoscope, b=-0.18, t(62)=-3.44, p=0.001, r=0.40, HMD, b=-0.13, t(62)=-3.30, p=0.002, r=0.39, and physical viewing conditions, b=-0.20, t(46)=-2.08, p=0.04, r=0.26, were significantly more shallow in the far viewing distance than to the near viewing distance. In sum, stereoscopic depth constancy was only seen in the physical-first conditions when physical stimuli were being tested. Despite the similarity of the stimuli, constancy was never attained in any of the virtual test conditions.

2.4 Discussion

2.4.1 General Summary

The aim of this series of experiments was to assess the impact of display-based cue conflicts inherent to computerized displays on the scaling of depth from binocular disparity. Overall, our results showed that depth estimates were more accurate, and observers achieved better scaling when surfaces were defined by binocular cues than monocular cues alone. The accuracy and scaling of perceived depth for physical stimuli (presented in a controlled full-cue PTE) was better than for virtual stimuli presented in either an HMD or a stereoscope. Further, the scaling of perceived depth was similar for virtual stimuli presented in both virtual apparatuses.

Importantly, our post-hoc analyses showed that the order of completion of the virtual and physical test conditions had a substantial impact on the constancy and scaling of perceived depth. Given the impact of condition order on perceived depth, to evaluate depth judgements for physical and virtual viewing conditions independent of condition order we focused on the conditions that observers completed first. That is, we compared the perceived depth judgements of physical objects for physical-first observers, and virtual objects for virtual-first observers. These two groups estimated the depth of our stimuli prior to the influence of any other test conditions. Later, we discuss the impact of condition order on subsequent depth estimates.

2.4.2 HMD vs. Traditional Stereoscopic Display

One potential issue with displaying large stimuli in HMD systems is the geometric distortions introduced by the magnifying lenses. HMDs use an inverse distortion correction to cancel the distortion caused by the lens by assuming the user maintains a fixed forward gaze. This creates a relatively undistorted high-resolution image in the center of the display, but at large gaze angles and eccentricities prism distortions become evident (for review of the impact of prism distortions see Ogle, 1952). Our comparison of depth judgements for virtual stimuli rendered in the stereoscope vs. the HMD reveals the impact of these distortions on depth perception. Under monocular viewing, for both groups of observers, depth scaling was the same for virtual stimuli displayed in the HMD and stereoscope. Thus, under these conditions the optics of the HMD do not play a significant role. Further, the similarity of the data suggests that other non-visual factors such as the weight of the HMD, or knowledge of the distance of the display from the face, do not systematically influence depth estimation. This is an important validation given the increasing use of HMD systems in vision science.

Under binocular viewing, depth was slightly overestimated at the smallest and underestimated at the largest half-cylinder depths when viewed in the HMD relative to the stereoscope (see Figure 2.5 – HMD & Stereoscope). The overestimation at the smallest cylinder depth was also seen in the combined analysis in Appendix 2.A (Figure 2.A2). Given that this difference was not present under monocular viewing, it likely reflects resolution limits of the HMD that impact precise rendering of binocular disparities. In our study, each pixel subtended approximately 4.7 and 1.3 arcmin of visual angle (assuming an even distribution of pixels across the central field-of-view) in the HMD and stereoscope condition, respectively. The peak disparity of the 1cm half-cylinder was 2.7 and 1.1 arcmin for the 83cm and 130cm viewing distances, respectively. Most of the population can detect depth differences defined by only 30 arcsec of disparity (Coutant & Westheimer, 1993; Westheimer & McKee, 1977), with some detecting

differences as small as 5 arcsec (McKee, 1983). Thus, even with built in anti-aliasing in the HMD apparatus, the fine binocular disparities required for the shallowest depths may not have been supported, which may have limited the overall scaling of perceived depth.

2.4.3 Display-based Cue Conflicts

The primary aim of this series of experiments was to assess the impact of both the presence and magnitude of display-based cue conflicts (inherent to 3D computerized displays) on the scaling of perceived depth. There are two primary ways in which the accommodative response could impact depth estimation of objects as a function of distance. First, accommodation helps support distance estimation (albeit weakly) and consequently the accuracy of depth perception from binocular disparity. Second, under some conditions (e.g., at close viewing distances) the presence of, and changes in, accommodative blur could help the observer gauge the 3D extent of the stimuli (Watt et al., 2005). Conflicts between vergence and accommodation would be expected to disrupt both uses of this information. As outlined in the Methods section (see Section 2.2.2) we selected our stimuli and viewing distances so that the difference in focal blur between the far edge and peak of the surface in the physical stimuli would be imperceptible; accommodative blur could not be used to judge the 3D extent of the object. However, it is possible that in binocular test conditions having correct accommodative distance information could aid scaling of disparity. So, we will focus on accommodation as a distance cue.

There is extensive evidence that virtual displays place an unequal demand on the accommodation and vergence systems (Eadie et al., 2000; D. M. Hoffman & Banks, 2010; Shibata et al., 2011). Further, manipulation of vergence and accommodation signals could affect the amount of depth perceived from horizontal binocular disparities (Frisby et al., 1995; Wallach & Zuckerman, 1963). We assessed whether the magnitude and direction of the vergence accommodation conflict influenced depth scaling by comparing depth judgements of virtual objects at two viewing distances in apparatuses with different focal distances (i.e., 74cm and 200cm for the stereoscope and HMD, respectively). At the near test distance of 83cm the accommodative plane was similar to the focal plane of the stereoscope, but different from that of the HMD. The reverse was true at the larger test distance of 130cm. If the discrepancy between the focal plane of the apparatus and rendered stimulus impacted depth estimation, we would expect that depth scaling would be more accurate in the near test condition for the stereoscope and the far test condition for the HMD. That is, the slope relating perceived depth to distance would be different in the two virtual test conditions. Instead, the pattern of results was the same; accuracy decreased as a function of viewing distance, to the same extent, in all virtual test

conditions. Reports of such under constancy are not uncommon in the literature (Foley, 1985; Gogel & Tietz, 1979; Johnston, 1991; Johnston et al., 1994), but the consistency of the slopes across our binocular test conditions shows that the magnitude of the mismatched accommodative distance is not responsible for the relationship. However, this is not to say that accommodation did not play any role in the depth estimation in our virtual test conditions. Recall that in the physical test condition binocular depth estimates were accurate and scaled well with distance. This was not the case in either of the virtual conditions. A parsimonious explanation for these results is that the simple act of decoupling vergence and accommodation in the virtual test conditions adds uncertainty to perceived distance which, in turn, disrupts depth scaling.

We found that depth scaling for physical stimuli was consistent with theoretical predictions at both viewing distances under binocular viewing (Figure 2.6 – PTE). The data also exhibited depth constancy, that is, perceived depth was equivalent at the two viewing distances. This is consistent with previous evaluations of stereoscopic depth constancy that show that at close viewing distances, near-accurate depth constancy is seen with physical stimuli with appropriate size-distance scaling in natural viewing environments (Durgin et al., 1995; Frisby et al., 1996; Willemsen et al., 2008). Unlike physical stimuli, depth judgments of virtual stimuli were less accurate and failed to achieve depth constancy irrespective of the virtual apparatus.

It is not surprising that under cue-rich conditions where there is little to no conflict between monocular and binocular sources of depth information, the visual system is able to scale depth with distance. The introduction of inconsistencies between these sources becomes the most likely explanation for inaccurate depth percepts. The presence and consistency of accommodative information appears to play a significant role in this 'physical advantage' as performance is degraded when it is absent (Frisby et al., 1995), or fixed to a plane (Ono & Comerford, 1977; Watt et al., 2005). In addition to the benefits of binocular fusion and reduction in visual discomfort, when vergence and accommodative distances are equivalent in virtual environments perceived depth estimates are more accurate than if they are in conflict (Hoffman et al., 2008). This effect is often attributed to an improvement in the distance estimate used to scale binocular disparities (Garding et al., 1995; Watt et al., 2005). The significant improvement in depth scaling for physical stimuli and equivalent scaling in both virtual apparatuses under monocular and binocular viewing suggests that the consistency of these ocular distances with the true distance of stimulus plays a significant role in the scaling of depth.

It is worth noting that while the reduction in accuracy as viewing distance increases for virtual stimuli was expected and has been confirmed in numerous studies of depth perception (Bradshaw et al.,

1996; Brenner & Landy, 1999; Collett et al., 1991; Foley, 1980; Glennerster et al., 1998; Johnston, 1991; Tittle et al., 1995), at first pass it may seem surprising that we see this reduction under binocular, but not monocular viewing. It is likely that this is due to the strong circular texture cues (Knill, 1998) in our stimuli which, unlike in most studies of cue integration, provided strong foreshortening cues that varied naturally with viewing distance in all test conditions (see also Collett et al., 1991; Frisby et al., 1996). Further, even though the reliability of accommodation degrades as the distance to the object increases its reliability as a distance cue is poor (Baird, 1903; Foley, 1977; Mon-Williams & Tresilian, 2000), especially at distances beyond 50cm (Watt et al., 2005). Thus, in the monocular viewing condition where distance cues were weak, observers likely relied on texture and size-scaling cues which did not degrade with viewing distance. In the case of binocular viewing in the virtual test conditions, again it appears that observers were disadvantaged by the presence of the conflict between ocular distance cues; the reduction in accuracy as a function of viewing distance was seen here, but not in the physical environment.

2.4.4 Order effects

To this point, our discussion of perceived depth for virtual and physical objects has focused on the conditions that observers completed first, and therefore without the influence of prior experience with the task. As outlined in the results section observers' experience with physical and virtual viewing environments did play an important role in their subsequent scaling of perceived depth. Observers who experienced the physical condition first showed an overall improvement in depth scaling under monocular viewing, however the relative differences in performance between apparatuses and viewing distances was unaffected by test order. All observers were more accurate when viewing physical stimuli than in either of the virtual conditions when only monocular cues were available (Figure 2.3 & 2.4). However, in the binocular viewing condition, experience played a significant role in both estimation accuracy and the impact of viewing distance (depth constancy).

To determine the impact of experience with physical stimuli on subsequent virtual depth judgements under binocular viewing, the virtual depth judgements for physical-first observers (Figure 2.6 – HMD & Stereoscope) were compared to virtual judgements for virtual-first observers (Figure 2.5 – HMD & Stereoscope). Observers with experience with the physical stimuli showed accurate depth scaling for all virtual depth judgements at near viewing distances, but observers without this experience did not (Figure 2.7). Despite the improvement in accuracy, these physical-first observers did not achieve depth constancy for virtual stimuli. Interestingly, experience with virtual stimuli had the opposite effect on subsequent physical depth judgements (compare Figures 2.4 and 2.5 – PTE). These virtual-first observers scaled depth

accurately at the near viewing distance, but their performance deteriorated as viewing distance increased, despite the presence of a cue-rich environment.

The compelling experience effects seen here have also been reported for other types of tasks at larger viewing distances. For instance, similar effects are seen in distance estimation using blind walking tasks in virtual and physical environments. In these studies, experience with a virtual environment first led to greater underestimation of distance in the physical environment (than seen without the prior exposure), and vice versa (Witmer & Sadowski, 1998; Ziemer et al., 2009). Further, when the virtual and physical environments are made more similar, distance judgements were equivalent in both environments when the physical space was experienced first (Interrante et al., 2006). Similarly, using a much smaller range of depth offsets, we have shown that experience with stereoscopic displays can significantly impact the accuracy of depth magnitude estimation, and observers' susceptibility to monocular conflicts (Hartle & Wilcox, 2016).

There are several potential explanations for the impact of previous test experience on depth judgements here. The simplest is that participants memorized the appearance of each stimulus or the range of their finger displacements on the sensor strip in the first test session and repeated these responses in subsequent sessions. However, if this were the case, then the results in the second session should have closely mirrored those of the first session. This is clearly not the case as we found significant effects of viewing distance that differed substantially between virtual and physical test conditions irrespective of test order.

Another, albeit unlikely, possibility is that that the disruptive effect of viewing the virtual environment first is due to a temporary disruption of binocular function caused by 'unnatural' vergence in HMDs (M. Mon-Williams et al., 1993). If this were the case, observers would have recalibrated their perception of space when subsequently testing in the physical environment which explains their improvement in the physical-second test conditions (Feldstein et al., 2020). While the explanation is consistent with some aspects of our data, it does not explain the substantial reduction in accuracy for physical judgements for virtual-first observers. Furthermore, our virtual and physical test sessions were conducted separately, and the intervening time would provide more than enough time to restore any disruption of binocular function.

Instead, it seems more likely that observers formed an internal representation of the distance to and size of the stimuli in the first session and apply this representation in subsequent sessions because of the similarity of the task, stimuli, and surroundings. Studies on the effect of binocular vision on grasping have shown similar effects of learned stimulus attributes, for instance in Keefe and Watt's (2009)

assessment of grip aperture. We cannot rule out the likely possibility that there is more than one cause of these test order effects, but it is clear that they can be significant and do not simply reflect overall improvements due to practice; additional research is needed to understand the factors critical to this phenomenon. For instance, another possibility is observers could be acquiring a prior for interpreting depth variation in the context of the virtual or physical environments (Kerrigan & Adams, 2013). However, it is clear that whatever the cause, advance experience with a well-matched physical version of a depth-based task can improve the accuracy and scaling of judgements with distance in a virtual version.

2.5 Conclusion

To assess the impact of display-based cue conflicts on depth scaling, we replicated a physical viewing environment in virtual counterparts. Under optimal viewing conditions the accuracy and constancy of depth estimation was equivalent in our two virtual environments. This shows that under these conditions HMD optics (e.g., magnification and lens distortion) have little impact on perceived depth. However, when contrasted with matched physical environments, depth judgements made in virtual test conditions are less accurate and do not achieve depth constancy. Given that our physical and virtual test conditions were otherwise the same, we conclude that the suboptimal performance is due to the decoupling of accommodation and vergence in these devices; the degree of conflict does not appear to modulate these differences, just its presence. Finally, our results show that observer's experience with physical and virtual viewing environments has a strong effect on the accuracy and constancy of their depth judgements. Thus, it is important to consider, and perhaps control, participants' familiarity with test environments in judgements of distance and depth, especially if they are being used for skill set training. Further, performance in virtual environments can be enhanced by brief exposure to a related physical task. The extent to which this training scenario must duplicate the virtual environment remains an open question.

CHAPTER 3: CUE VETOING IN DEPTH ESTIMATION: PHYSICAL AND VIRTUAL STIMULI

Preface

The previous chapter demonstrated that careful control of physical and virtual test conditions revealed systematic effects of display-based cue conflicts on distortions of stereopsis. The aim of this chapter was to expand upon that work by examining the role of display-based cue conflicts on the combination of other depth cues, such as motion parallax, with stereopsis. Like the previous chapter, this study used a highly controlled physical test environment as a baseline comparison for virtual stimuli rendered in a head-mounted display system. Specifically, the assessment of the accuracy and precision of depth judgements of virtual and physical objects was used to evaluate how depth from binocular disparity and motion parallax cues are combined in these two test environments. The objective of the direct comparison between the virtual and physical test environments was to determine if the previous failures of linear cue combination models was driven by the limitations of computerized displays systems.

The contents of this chapter was published as a manuscript titled, *Cue vetoing in depth estimation: Physical and virtual stimuli,* in volume 188 of Vision Research (Hartle & Wilcox, 2021). Both authors contributed the conceptualization, design of methodology, review, and editing of the manuscript. Brittney Hartle completed all programming, performing of experiments, data collection, statistical analysis, and prepared the original draft of the published work. Laurie M. Wilcox supervised, managed, and coordinated responsibility for the research.

3.1 Introduction

The ability to accurately estimate the depth and distance of objects is critical to our interpretation of and interaction with the world around us. Here, depth refers to the extent of an object along the z-dimension, while distance refers to the amount of space from the eye to a point on the object's surface. It is well established that when estimating the 3D shape of an object in a complex real scene, the visual system uses multiple monocular and binocular sources of depth information. For instance, depth perception is supported by static monocular cues such as perspective, relative size, and occlusion; when information about absolute distance is available binocular disparity allows observers to judge the amount of depth between objects. Stereopsis is based on the positional disparity between images of an object on the retinae, an observers' interocular distance, and the egocentric or absolute distance to the object. In virtual stimuli (i.e., imaged on computer displays), distortions in relative depth

from binocular disparity have been documented over a wide variety of stimuli, tasks, and viewing distances (Foley, 1967; Foley, 1980). These distortions are often attributed to unreliable or erroneous estimates of absolute viewing distance (Foley, 1980; Rogers & Bradshaw, 1993). That is, in impoverished viewing environments with few distance cues, there is little binocular information available to support reliable estimates of absolute distance apart from the pattern of vertical disparities and the vergence angle of the eyes (Foley & Richards, 1972; Foley, 1985; Rogers & Bradshaw, 1993; Wallach & Zuckerman, 1963). Unsurprisingly, if absolute distance estimates are based on a variable vergence signal or limited vertical disparities, then depth from binocular disparity will also be degraded (Gogel, 1977; Johnston, 1991).

Absolute viewing distance information can also be provided by monocular cues to distance, such as accommodation, familiar size cues, or a combination of relative distance cues. When combined with binocular disparity cues, this monocular information could help improve the accuracy of depth judgements. However, monocular and binocular depth cues are often in conflict in computerized displays. For instance, in conventional stereoscopic display systems accommodative distance always specifies the distance to the screen plane rather than the distance to the 3D object which may be positioned at some distance in front of or behind the screen. This discrepancy results in conflict between vergence and accommodation specified distance that increases as objects are positioned further from the screen plane; particularly when the screen is placed at near viewing distances (Fry, 1939). In physical viewing environments, accommodation and vergence responses are coupled regardless of the distance of the object. The coupling of these cues may increase the accuracy of scaling of depth from binocular disparity and other monocular distance cues (for review see Ono & Comerford, 1977). Studies of stereopsis using physical wireframe stimuli have shown that observers exhibit systematic biases in perceived depth when vergence and accommodative distance are shifted relative to the true viewing distance by trial lenses (Wallach & Zuckerman, 1963). However, as discussed below, many of the biases seen in stereoscopic displays are not evident when physical targets are used. When additional monocular distance cues are available in physical viewing environments, participants are effective at judging depth from binocular disparity at viewing distances up to 3m (Durgin et al., 1995).

Like binocular disparity, motion parallax can be used to determine the relative depth of objects (Rogers & Graham, 1983). For a given lateral head movement, the linear motion parallax between two parts of an object fixed in depth at different points in time varies inversely with the square of their distance (Howard & Rogers, 2012). The same sources of absolute distance that are required for scaling depth from binocular disparity can be used to scale motion parallax. However, additional information

about eye, head, and body position is required to determine relative depth from motion parallax alone (Helmholtz, 1925; Howard & Rogers, 2012). While a few studies have found that accuracy is similar for depth judgements from binocular disparity and motion parallax (Bradshaw et al., 2000; Johnston et al., 1994), others (Durgin et al., 1995; McKee & Taylor, 2010) have shown the estimates of depth from motion parallax for virtual and physical objects are less accurate than binocular disparity. Furthermore, the absolute distance information from motion parallax, if any, tends to be very weak (Gogel & Tietz, 1973; but also see Gogel & Tietz, 1979). More recently, it has been suggested that the observed distortions in the perceived depth of virtual stimuli maybe influenced by unmodeled conflicts between focus and motion cues (Scarfe & Hibbard, 2011), or unmodeled texture cues (Hillis et al., 2004).

Assessments of the integration of stereopsis and motion parallax often use virtual stimuli which are susceptible to distortions of perceived absolute distance and display-based cue conflicts (Landy et al., 1995; Norman & Todd, 1995). These studies typically show that depth is misperceived even when both binocular disparity and motion parallax are available (Scarfe & Hibbard, 2011; Todd, 1985; Todd & Norman, 2003). It is important that we exercise caution when drawing general conclusions based solely on stereograms. For instance, while the integration of depth from binocular disparity and texture cues has been shown to depend on the orientation of surface curvature for virtual stimuli, physical stimuli with accommodative blur cues show no such anisotropy (Buckley & Frisby, 1993; Frisby et al., 1995). Real-world viewing conditions can be approximated using 3D accommodative display systems (Akeley et al., 2004). However, even under these conditions, some limitations remain; for instance, due to display restrictions disparity must be interpolated between a limited number of accommodative planes. Another approach to eliminating cue conflicts is to use physical stimuli presented under controlled viewing conditions. The few experiments that have been conducted show that physical stimuli exhibit some of the same perceptual biases reported for virtual targets (Bradshaw et al., 2000; Todd & Norman, 2003). This is most likely due to the fact that these studies typically used impoverished stimuli (e.g., points of light) with very few distance cues (Bradshaw et al., 2000), or a static viewpoint with a moving object (for review see Landy & Brenner, 2001). Arguably to evaluate the interaction between multiple sources of depth information, more complex environments are needed. Here we used a head-mounted display (HMD) system to update the rendered images according to the observer's head position, so observers generated the motion parallax. We evaluate the contribution of binocular disparity and user-generated motion parallax to the perceived depth of volumetric stimuli; by comparing carefully matched virtual and physical test conditions we determine the relative impact of monocular and binocular depth cues.

In this series of experiments, the accuracy and precision of depth estimation was assessed for virtual and physical stimuli in three cue conditions, (1) motion parallax alone, (2) binocular disparity alone, and (3) both cues present. In all viewing conditions, the information from each cue was consistent with the true depth of the stimulus. The virtual stimuli were rendered in the Oculus Rift HMD and the full-cue physical stimuli were presented in an automated physical test environment (PTE). The depth of truncated square pyramids was measured using a discrimination task, in which observers indicated whether the perceived depth between the base and front base of the pyramid was greater or less than the width of the front surface. To model cue integration, we applied a Bayesian model with either (1) linear, (2) veto, or (3) correlated combination methods. To evaluate the impact of display-based conflicts on depth cue integration we compared the best-fitting Bayesian observer models for virtual and physical objects. To anticipate our results, we found that depth estimates were markedly similar for virtual and physical stimuli in all three cue conditions and depth estimates were most precise when depth was defined by binocular disparity or the combination of binocular disparity and motion parallax. Our modelling shows that observers tend to veto the less reliable motion parallax cue in both virtual and physical viewing environments.

3.2 Methods

3.2.1 Observers

Eight observers were recruited from York University. The stereoacuity of all observers was assessed using the Randot^{TM} stereoacuity test to ensure observers could detect depth from binocular disparities of at least 40 arcseconds. All observers had normal to corrected-to-normal vision, and if necessary, wore their corrected lenses during testing. The research protocol was approved by York University's Research Ethics Board.

3.2.2 Stimuli

The stimuli consisted of truncated square pyramids with a random texture comprised of white circles on a grey background (Figure 3.1). The dimensions of the virtual and physical pyramids were equivalent. The size of the front face and base for all pyramids was 6 cm by 6 cm and 12 cm by 12 cm, respectively. At a viewing distance of 83 cm, the visual angle of the base was 8.27 deg, and the visual angle of the front surface ranged from 4.30 to 4.64 deg depending on the pyramid depth. The distance from the base to the front face of the pyramid (i.e., the pyramid's depth) was sampled around the 6 cm

reference pyramid at step sizes of 0.5 cm or 1.0 cm. Each observer's step size was determined in a short practice session prior to the full experiment.

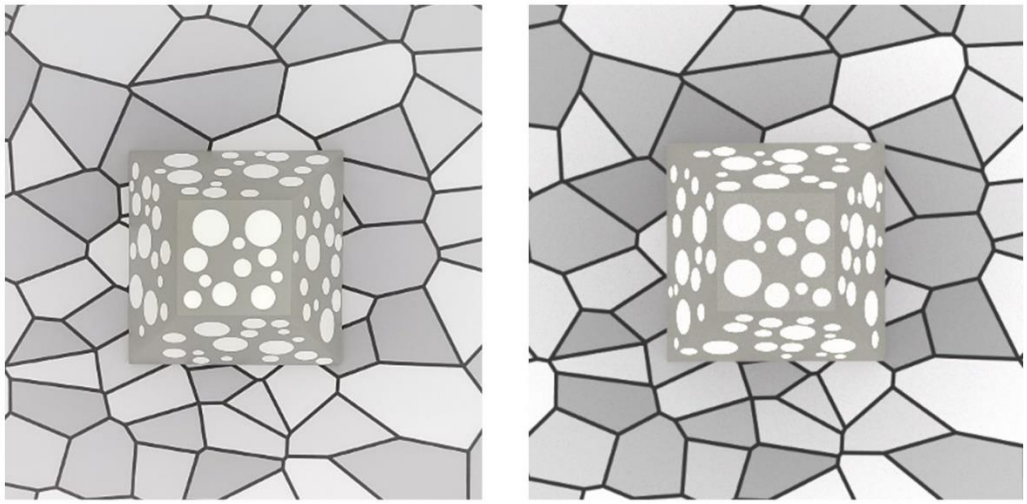


Figure 3.1. *The left image shows an unedited picture of the 6 cm physical pyramid in the PTE apparatus. The right image shows an illustration of the 6 cm virtual pyramid rendered for viewing in the HMD.*

Each pyramid was textured with a random array of non-overlapping circular elements of four sizes: 0.64, 0.95, 1.30, and 1.90 cm (Figure 3.1). The size of these texture elements ranged from 0.44 to 1.47 deg depending on the pyramid depth. The distribution of each of the four texture element sizes (from smallest to largest) on the front surface was 2, 3, 3, 3. For the side surface, which had a larger surface area, there was 3, 4, 5, 4 elements of each size. In the physical pyramids, the luminance of the texture elements was 171.0 cd/m^2, and the luminance of the front and side faces were 59.7 cd/m^2 and 53.8 cd/m^2, respectively. The luminance of the texture elements and the pyramid surface of virtual pyramids were adjusted to match the contrast between the edge of the texture elements and the pyramid surface of the physical pyramids. The virtual textures were generated in MATLAB while the texture elements for the physical pyramids were cut from white vinyl sticker sheets and affixed to the objects that were spray painted with a matte grey paint. To prevent observers from using the absolute position of the texture elements as a reference, unbeknownst to the observer each pyramid was randomly rotated between viewing conditions. All pyramids were presented on a Voronoi background texture generated using the voronoin() function in MATLAB with low contrast grey elements. The position of the points were randomly sampled from a standard uniform distribution and the Delaunay

triangulation parameter was set to the default 'Qbb'. The background texture provided a stable reference for observer's depth judgements.

3.2.3 Apparatus

Virtual pyramids were created in Blender and presented in the Oculus Rift CV1 HMD using the PsychXR library in Python (Cutone & Wilcox, 2018). The Oculus Rift headset was connected to an Alienware Windows 10 computer with a NVIDIA GeForce GTX 1080 graphics card. The Oculus Rift has two organic light-emitting diode displays, each with a resolution of 1080 by 1200 pixels per eye with a refresh rate of 90Hz and a horizontal field-of-view of 94 deg. Each pixel subtends 4.7 arcmin of visual angle. Python code was optimized for presentation in the Oculus Rift headset, such that dropped frames were limited to less than 0.01% of total frames during each virtual cue condition. Prior to testing, each observer's interpupillary distance was measured using a digital pupillometer (GR-4) and the interocular separation of the HMD lenses was adjusted to match this separation. Observers rested their head on a chin rest to stabilize their head position. The chin rest was mounted on a horizontal motion platform that recorded its lateral position. The same motion platform was used in the virtual and physical viewing conditions. The maximum travel distance of the motion platform was 13 cm, which allowed the observer's head to move 6.5 cm to the left and right of the center position. Observers synchronized their movements to a 60 bpm metronome tone, such that their head reached the end of the platform when the tone sounded. To match the visual cues in each environment as closely as possible, the structure of the PTE apparatus was modelled in its entirety in the virtual viewing environment, including the aperture and poster board that was visible to the observer.

The physical stimuli were presented under controlled lighting conditions in our computer-controlled physical test environment (for details see Hartle & Wilcox, 2016). Physical pyramids were 3D printed using a LulzBot TAZ 6 3D printer with the same dimensions as their virtual counterparts. Pyramids were mounted on a 3.8 cm thick polystyrene board (122 cm by 61 cm) using magnets embedded in the board and the base of the 3D printed pyramids. The polystyrene board was positioned 83 cm from the observer. The Voronoi background texture was printed on matte heavyweight paper and glued to the polystyrene board. Four pyramids were mounted on the board during each block, and the horizontal actuator in the PTE centered the pyramids in the aperture in front of the observer on each trial. A Cameron RL-160 Bi-Color LED ring light illuminated the central pyramid on each trial. A 16.7 cm square aperture was placed 48 cm in front of the observer to limit the field-of-view to 19.7 deg at a viewing

distance of 83 cm (Figure 3.2). Limiting the size of the visual field to 19.7 deg ensured that the adjacent pyramids on the poster board in the PTE apparatus were not visible on each trial.

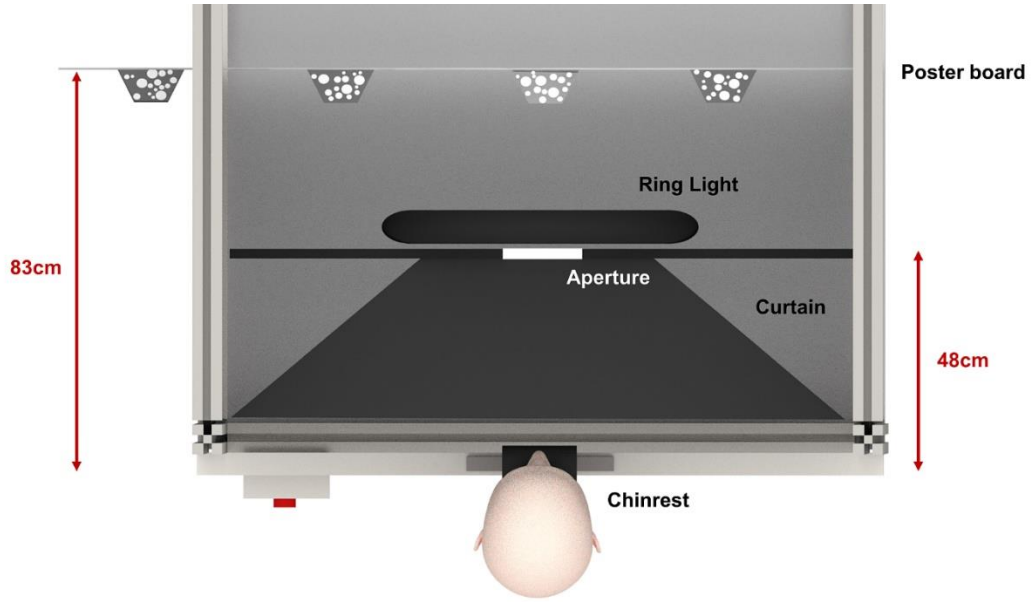


Figure 3.2. An illustration of a top-down view of the PTE apparatus. The poster board was placed 83 cm from the observer. A 16.7 by 16.7 cm opening was cut into a matte black poster board and positioned 48 cm from the observer between the ring light and the enclosure curtain. The aperture limited the observer's field-of-view by blocking their view of both the ring light and adjacent pyramids mounted on the poster board. The matte black curtains framed the apparatus, blocking residual light and the observers' view of the inside of the enclosure.

3.2.4 Procedure

An internal-reference discrimination task was used to assess the perceived depth between the base and front face of the pyramid using the method of constant stimuli. Observers indicated whether the depth between the base and front face of the pyramid was greater or less than the width of the front face of the surface under three viewing conditions, (1) motion parallax, (2) binocular disparity, and (3) combined cue. In the motion parallax condition, observers moved their head laterally to the beat of the metronome while wearing an eye patch on their left eye. In the binocular disparity condition, observers rested their head on a stationary chin rest fixed in the center of their field-of-view and viewed the stimulus binocularly. In the combined cue condition, observers moved their head with the metronome while viewing the stimulus binocularly. In all conditions, observers controlled when the stimulus appeared by pressing a button on the gamepad (an Xbox One wireless controller). This gave observers time to synchronize their head movements to the metronome before the stimulus was presented. The stimulus remained visible until the observers submitted their response using the gamepad. Observers were instructed to maintain their gaze on the front surface of the pyramid for the duration of each trial. It is

possible, but unlikely that observers could respond based solely on the width of the front surface, rather than using the perceived width to estimate the depth of the pyramid. As a check, we added a catch trial using a pyramid with the same depth as the reference, but a different (larger) front surface size. These catch trials were interleaved with the standard test conditions. Each cue condition included trials with a pyramid depth of 6 cm with the size of the front surface modified to match the visual angle of the largest pyramid in the range (i.e., 4.55 or 4.64 deg depending on the observer's step size). Each pyramid (including the catch trial condition) was presented 20 times, resulting in 160 trials per cue condition.

Our task requires that observers use the width of the front surface to estimate the depth of the pyramid. While the physical dimensions of the front surface were constant (6 cm) the perceived width may have varied across observers. Given the internal-reference discrimination task is a relative comparison between the width of the front face and the depth of the pyramid, without an assessment of the perceived width of the front face, the results of the discrimination task alone do not provide information regarding the amount of perceived depth. To simulate the depth-width discrimination task in a Bayesian framework, a measure of the perceived reference width for each observer was required. To do this, in separate sessions we used a magnitude estimation task to assess the perceived width of the front face of the virtual and physical pyramids. To obtain these estimates, pyramids with depths of 4.0, 5.0, 5.5, and 6.0 cm from the base to the front face were placed on poster board on raised platforms, such that their front faces were located 5.5, 6.0, 6.5, and 7.0 cm in front of the poster board. Observers rested their head on a stationary chin rest and viewed the stimuli binocularly. Given past assessments of perceived 3D line length show similar accuracy when stimuli are defined by motion, stereopsis, or a combination of both cues (J. Farley Norman et al., 1996), in our study observers estimated the perceived width of the front surface with a stationary head position binocularly. During a trial, the stimulus remained visible until observers submitted their response using a custom-built pressure-sensitive strip (for details and validation see Hartle & Wilcox, 2016). Observers rested their thumb against one end of the sensor strip and pressed their index finger along the length of the sensor to indicate the magnitude of their estimate.[4] Each pyramid was presented 10 times, for a total of 40 trials. The perceived width of the front surface for each observer was then used as the reference width for their discrimination judgements in the Bayesian model (for details see section 3.3.1 Bayesian Observer Model).

[4] We have previously shown that depth estimation accuracy for cross-modal finger displacement tasks (either via sensor strip or direct measurement) is the same as that obtained using an intra-modal task such as a virtual ruler (Hartle & Wilcox, 2016).

3.3 Results

The perceived width judgements provided a magnitude estimate of the perceived width of the front surface, which was used to simulate the depth-width discrimination task in the Bayesian observer model by combining the distribution for the reference width with the posterior distribution for each observer (for model details see Figure 3.8). The difference between the mean of the reference distribution and the point of subjective equality (PSE) for each observer was used as a relative measure of perceived depth distortions. For example, if the mean of an observer's perceived width judgements was 3 cm, then they underestimated the reference width by 50%. When they performed the depth-width discrimination task, this perceived reference width was compared to the pyramid depth. If their PSE was close to 6 cm, then the observer underestimated the perceived depth and width by equal amounts (i.e., a 3 cm difference). If the PSE fell exactly on the reference width of 3 cm, then the observer was perceiving the depth veridically while still underestimating the perceived width. The relative comparison between these two distributions provides insights into the depth distortions in each of the three cue conditions in the virtual and physical viewing environments.

Figure 3.3 shows the average perceived width of the front face for the virtual and physical pyramids and the individual means for each observer. Both the individual and mean perceived depth estimates in Figure 3.3 show that observers systematically underestimated the perceived width of the front surface in virtual relative to physical stimuli. To evaluate if the perceived widths of the virtual and physical stimuli were significantly different, the data were analyzed by fitting a linear mixed-effects model using the nlme package in R (Pinheiro et al., 2015). The repeated-measure variables were accounted for by using nested random intercepts. Significance was determined using planned a priori comparisons between stimulus type using t-tests, and an approximation of Pearson's correlation coefficient (r) was used as a measure of effect size (Field et al., 2012). The analysis confirmed that there was no significant difference in perceived width between pyramid depths, $X^2(9) = 4.89$, p=0.18, but there was a significant difference between the physical and virtual pyramids, $X^2(6) = 12.05$, p=0.001. The perceived width estimates for the physical pyramids were significantly larger relative to virtual pyramids, b=0.53, $t(7)$=3.77, p=0.007, r=0.82. However, on average the width of the front surface was underestimated for both virtual and physical pyramids. The mean and standard deviation of the perceived width estimates for each observer determined the mean (μ_{ref}) and standard deviation (σ_{ref}) of the Gaussian distribution for the reference width in the Bayesian observer model.

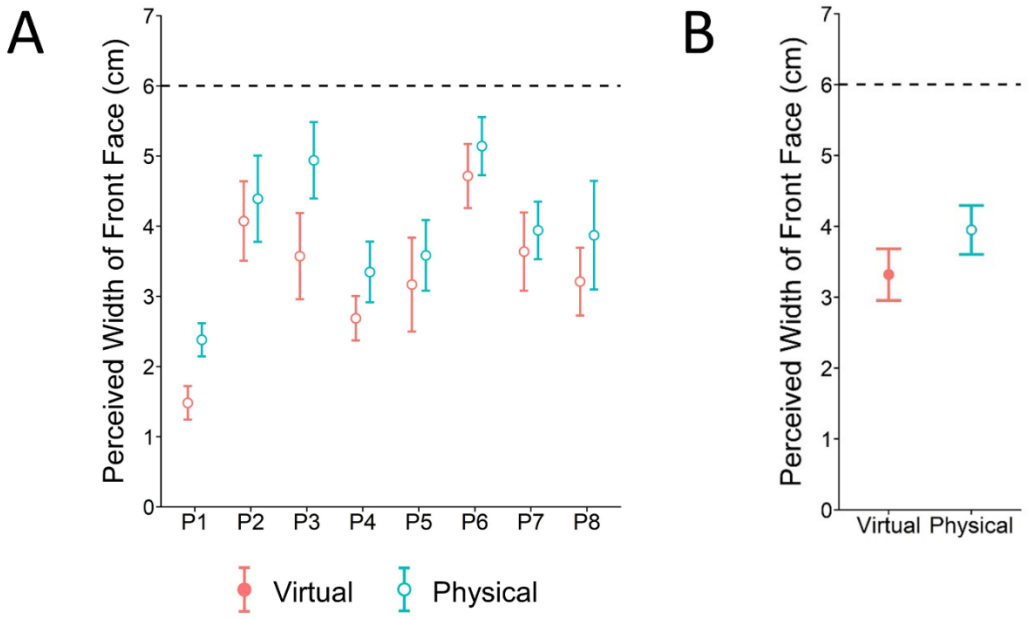


Figure 3.3. *Graph A shows the average perceived width estimates of the front surface for the virtual and physical pyramids for each observer. Graph B shows the average perceived width of the front surface for the virtual and physical pyramids. The error bars represent the standard error of the mean. The black dotted lines represent the true width of the front surface.*

Figure 3.4 shows the proportion of responses for each type of pyramid in the binocular disparity, motion parallax, and combined cue conditions for virtual and physical pyramids. The average responses show that there was little difference between the proportion of responses for the original and catch trial pyramids. To determine if the change in visual angle of the front surface over the range of pyramid depths impacted observer's judgements, we compared the proportion of responses for the 6 cm pyramid depth to the catch trial with the modified size of the front surface. The data was fit with a mixed-effect logistic regression with a logit (binomial) link function using the lme4 package in R. The repeated-measure variables were accounted for using nested random intercepts. Effect sizes were converted from log odds ratios into Cohen's standardized mean difference (d) values using the transformations proposed in Borenstein et al. (2009). A likelihood ratio chi-square test determined if the difference between the proportions for each pyramid type reached significance in each of the three cue conditions for virtual and physical pyramids. The results showed that there was no significant difference in the proportion of responses between the two types of pyramids, $X^2(9) = 0.33$, $p=0.56$. There was no significant effect of viewing condition (i.e., virtual or physical) on the proportion of response for each pyramid, $X^2(14) = 0.74$, $p=0.39$, nor was there an effect of cue condition, $X^2(13) = 0.29$, $p=0.86$. Lastly, there was no significant three-way interaction between the type of pyramid, viewing condition, and cue condition on the

proportion of responses, $X^2(16) = 0.60$, p=0.74. These effects were confirmed as all planned a priori comparisons for all fixed-effects (including all interactions) were non-significant. Thus, the change in visual angle of the front surface over the range of pyramid depths had no impact on the proportion of observers' responses.

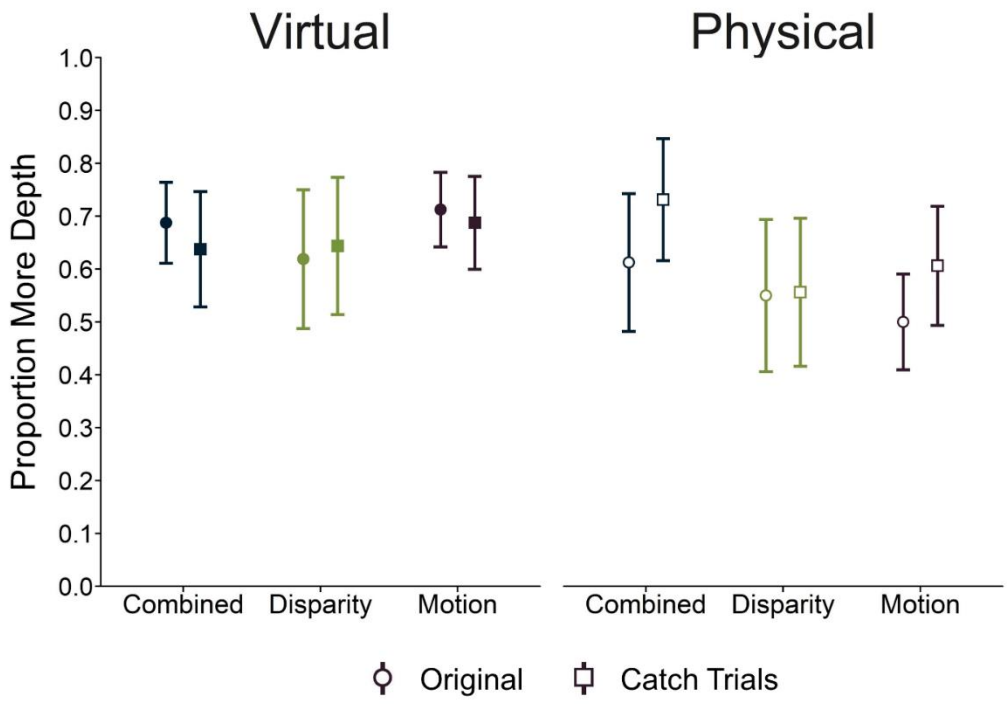


Figure 3.4. The mean proportion of responses of "more depth" for the pyramid depth of 6 cm (circles) and the catch trial stimulus (squares) from all observer's psychometric data. The catch trial pyramid had the same depth as the standard stimulus, but the width of the front surface matched the visual angle of the largest pyramid in the range. The proportion is shown for each of the three cue conditions: binocular disparity only (green), motion parallax (purple), and their combination (blue). Error bars represent the standard error of the mean.

For each of the eight observers, a maximum likelihood method was used to fit a cumulative normal distribution to the empirical psychometric function for the binocular disparity, motion parallax, and combined cue conditions. An example of one observer's psychometric function for the virtual viewing condition is shown in Figure 3.5. The PSE was computed as the 50% point for each test condition for each observer and the just noticeable difference (JND) was computed as the difference threshold between 75% and 25% divided by 2. Bootstrapped 95% confidence intervals (CI) were calculated for the PSE and JND measurements using Monte Carlo simulation methods run 10,000 times for each dataset (Wichmann & Hill, 2001a, 2001b). To evaluate if the PSEs and JNDs were significantly different in the cue conditions for virtual and physical pyramids, the data were analyzed by fitting a similar linear mixed-effects model as the analysis for perceived width. The repeated-measure variables were accounted for by using nested

random intercepts in a hierarchy. These variables modeled the correlation of the variance of the intercepts for each observer within each type of pyramid for each cue condition. A likelihood ratio chi-square test determined the significance of the fixed-effects. Planned a priori comparisons for each fixed-effect were evaluated using t-tests, and an approximation of r was used to measure effect size.

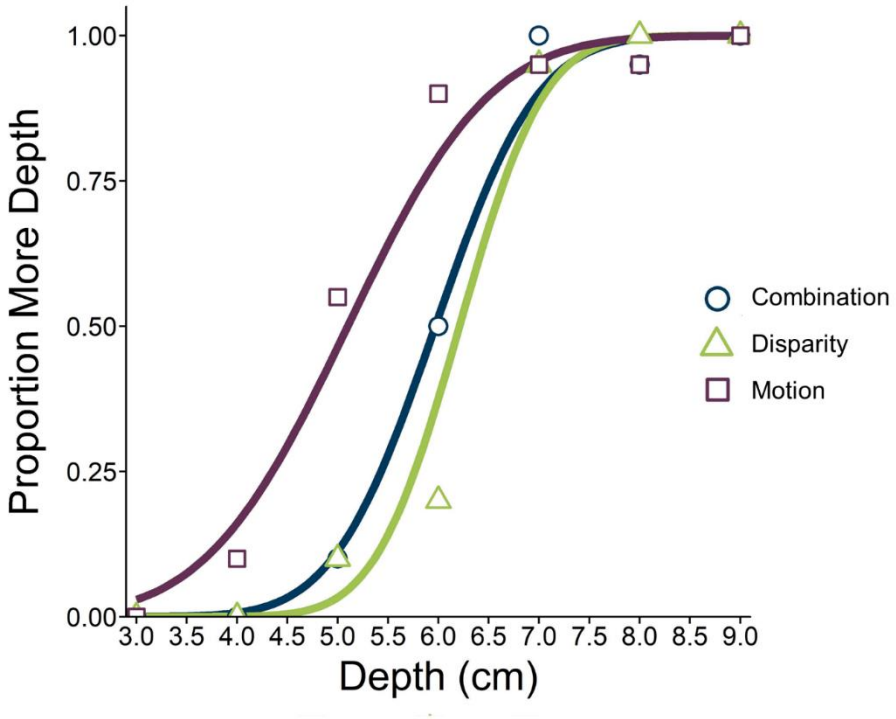


Figure 3.5. *An example of one observer's psychometric functions for the virtual viewing condition. The proportion of "more depth" responses for each pyramid depth in centimeters are shown for each of the three cue conditions: binocular disparity (green triangles), motion parallax (purple squares), and their combination (blue circles).*

Figure 3.6 shows the average PSEs for the single (binocular disparity, motion parallax) and combined cue conditions for the virtual and physical pyramids (individual PSEs are shown in Appendix 3.A). If the observer estimated depth and perceived width veridically, then the PSE should fall exactly on the true reference width of 6 cm. However, the analysis of perceived size data in Figure 3.3 determined that observers underestimated the perceived width of the front surface. In this case, if the PSE falls on the true reference width of 6 cm, then the observer underestimated the perceived depth and width by equal amounts. If observers perceived the depth of the pyramid veridically, but underestimated the perceived size of the reference width, then their PSEs should fall on the mean of their perceived width estimates. Figure 3.6 shows that the mean PSE was larger than perceived reference width. On average, observers were responding "less depth" more often. Thus, all observers underestimate the perceived depth of both the physical and virtual pyramids. Our analysis showed that there was no significant

difference between the PSEs in the physical and virtual pyramids, $X^2(8) = 3.75$, p=0.05, in the three cue conditions, $X^2(7) = 0.16$, p=0.92, or in the three cue conditions as a function of pyramid type, $X^2(10) = 1.45$, p=0.49. In addition, to determine if there was a difference in the magnitude of the depth distortions for virtual relative to physical stimuli, we compared the difference between perceived width estimates for each observer to their average PSE across the three cue conditions. This analysis also revealed no significant difference between virtual and physical stimuli, $X^2(5) = 1.21$, p=0.27.

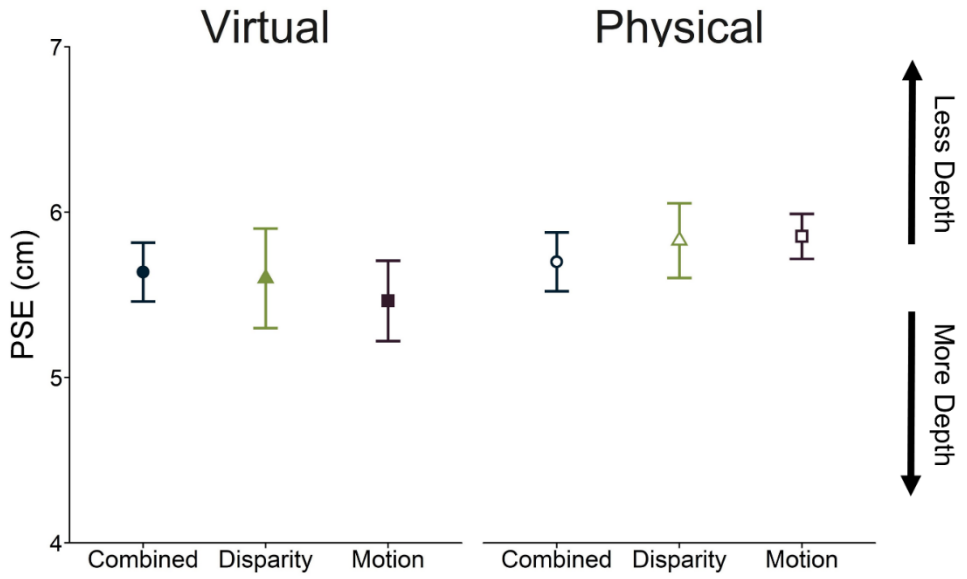


Figure 3.6. Average PSEs (n = 8) are shown here for each of the three cue conditions: binocular disparity only (green triangles), motion parallax (purple squares), and their combination (blue circles). Error bars represent the standard error of the mean.

The average JNDs for the binocular disparity, motion parallax, and combined cue conditions for the virtual and physical test conditions are shown in Figure 3.7 (individual JNDs are shown in Appendix 3.A). The analysis revealed a significant two-way interaction between the type of stimulus and cue condition, $X^2(10) = 6.26$, p=0.04. Orthogonal contrasts revealed that the JNDs for the motion parallax condition were significantly elevated relative to the combined, b=0.29, $t(14)$=4.82, p<0.001, r=0.79 and binocular disparity conditions, b=0.37, $t(14)$=6.10, p<0.0001, r=0.85. In addition, JNDs for the motion parallax condition were significantly smaller for physical relative to virtual pyramids, b=-0.24, $t(7)$=-2.77, p=0.03, r=0.72. However, there was no difference in JNDs for the combined, b=-0.14, $t(7)$=-2.08, p=0.08, r=0.62, and binocular disparity conditions, b=-0.03, $t(7)$=-0.95, p=0.38, r=0.34, between the physical and virtual pyramids. Thus, observers were more precise when depth was defined by binocular disparity or the combination of binocular disparity and motion parallax, than when depth was defined by motion parallax alone. This was true for both virtual and physical stimuli.

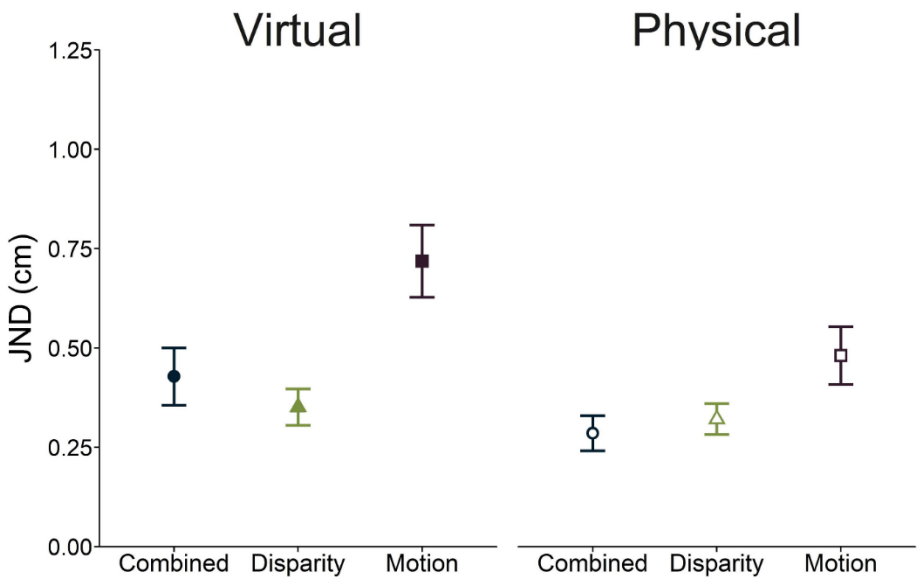


Figure 3.7. Average JNDs (n = 8) for each of the three cue conditions: binocular disparity only (green triangles), motion parallax (purple squares), and their combination (blue circles). Error bars represent the standard error of the mean.

3.3.1 Bayesian Observer Model

Bayesian decision theory has been widely used as a basis for modelling how sensory information from multiple sources, with differing reliability, are integrated (Landy et al., 1995). Among these, weighted linear combination methods assume that each depth cue is processed separately and integrated into a combined estimate with greater weight placed on the more reliable cues (Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003). However, in scenarios where the visual information is noisy or incomplete, alternative combination methods have been proposed (Maloney & Landy, 1989). To assess how depth cue integration is achieved, multiple depth cues must be presented in different combinations, in scenarios with different view geometry and supplementary visual information.

We created a Bayesian observer to model the integration of depth from binocular disparity and motion parallax for both virtual and physical pyramids. A Bayesian observer estimated the pyramid depth on a trial-by-trial basis according to the 3D geometry for each observer's data. Each observer's interocular distance was used to calculate the binocular disparity between the left and right eye images in degrees. The predicted oscillation of lateral head movements was defined by the sinusoidal function, $F(cm) = a * \sin(2pi/\omega t)$, with an amplitude ($a$) of 6.5 cm and period ($\omega$) of 2 seconds. The depth between the base and front face of the pyramid (Δd) was calculated using the following equation:

$$\Delta d = \frac{D^2 * \delta}{IOD - \delta * D}$$

where IOD is the observer's interocular distance, D is the distance from the observer to the front face of the pyramid, and δ is the horizontal angular disparity (Howard & Rogers, 2012, pp.154). This equation assumes symmetrical convergence and the small angle approximation, where the tangent of an angle is approximately equal to the angle in radians. If angular disparity is specified in degrees, then to equate disparity to units of distance (e.g., centimetres or metres) it must be converted from degrees to physical disparity using $\tan(degrees * (pi/180))$. For motion parallax with lateral head motion, the relative angular velocity between the base and front surface is equivalent to disparity (δ) and head velocity is equivalent to IOD in binocular vision (Gillam et al., 2011; Ono et al., 1986). This portion of the model is deterministic and does not introduce any noise to the depth estimates the pyramid.

The Bayesian observer interprets the imprecise depth information considering prior experience. For each possible stimulus, the Bayesian observer considers the probability of each hypothesized depth given the depth of the stimulus (i.e., the likelihood) and the prevalence of the stimulus from experience (i.e., the prior). The information provided by binocular disparity and motion parallax about the perceived depth of the pyramid is given as the posterior probability, $p(d|b; m; \sigma_b; \sigma_m; \sigma_p)$. Using Bayes' rule and assuming that the sensory noise associated with binocular disparity and motion parallax are independent, we can write the posterior probability as the product of the likelihood functions of each cue and the prior distribution. Where binocular disparity and motion parallax are defined as,

$$p(d|b; \sigma_b, \sigma_p) \propto p(b|d; \sigma_b)\ p(d; \sigma_p) \text{ and}$$

$$p(d|m; \sigma_m, \sigma_p) \propto p(m|d; \sigma_m)\, p(d; \sigma_p), \text{ respectively.}$$

Here, the likelihood functions are $p(b|d; \sigma_b)$ and $p(m|d; \sigma_m)$ for binocular disparity and motion parallax, respectively. Each likelihood function was modelled as a Gaussian centered on the true depth of the pyramid (i.e., b and m) with the spread of the distribution (i.e., σ_b and σ_m) fit using the JNDs of the observer's empirical psychometric function in each single-cue condition. For each observer, the sigma for each cue condition (i.e., σ_b and σ_m) and the sigma for the prior distribution (i.e., σ_p described below) were fit to the observer's data in conjunction.

The likelihood distribution for each individual cue is combined with the prior distribution, $p(d; \sigma_p)$ that represents cues to flatness. Residual flatness cues are associated with a prior for fronto-

parallel surfaces in limited cue situations and/or cues to flatness inherent to a flat monitor, such as accommodation (Watt et al., 2003). These cues were modelled as a Gaussian centered at zero depth with a standard deviation σ_p. The standard deviation of the prior (σ_p) reflects the relative strength of the prior for fronto-parallel and the reliability of residual flatness cues. In each single cue distribution, the likelihood was combined with residual flatness cues to produce the posterior distribution. The same σ_p was used in the binocular disparity and motion parallax cue conditions.

Once the posterior distribution was determined for each single cue by combining their likelihoods with the prior distribution, the Bayesian observer simulated the same perceptual discrimination task completed by the human observers. The width of the reference surface for the Bayesian observer was a Gaussian distribution with a mean (μ_{ref}) and standard deviation (σ_{ref}) determined from each observer's width (size) estimates. To determine the Bayesian observer's psychometric function, we calculated the probability that the depth (d) for each single cue distribution was greater than the reference width (r) for each hypothesized depth value (i),

$$p(d > ref) = \sum_{ref_i} p(ref_i)p(d > ref_i).$$

Figure 3.8 shows an example of a simulated trial for the Bayesian observer model. The psychometric function for the Bayesian observer model for the binocular disparity and motion parallax cue conditions depended on the μ_{ref} and σ_{ref} of the reference width for each observer, the σ_p of the prior distribution, and the standard deviation of the likelihood distributions for each cue (σ_b or σ_m). The estimates of σ_b, σ_m, and σ_p were fit by comparing the psychometric function for the Bayesian observer to the observer's empirical psychometric functions in each single cue condition. The σ_b, σ_m, and σ_p that best fit the empirical psychometric function for each observer were found for both virtual and physical pyramids.

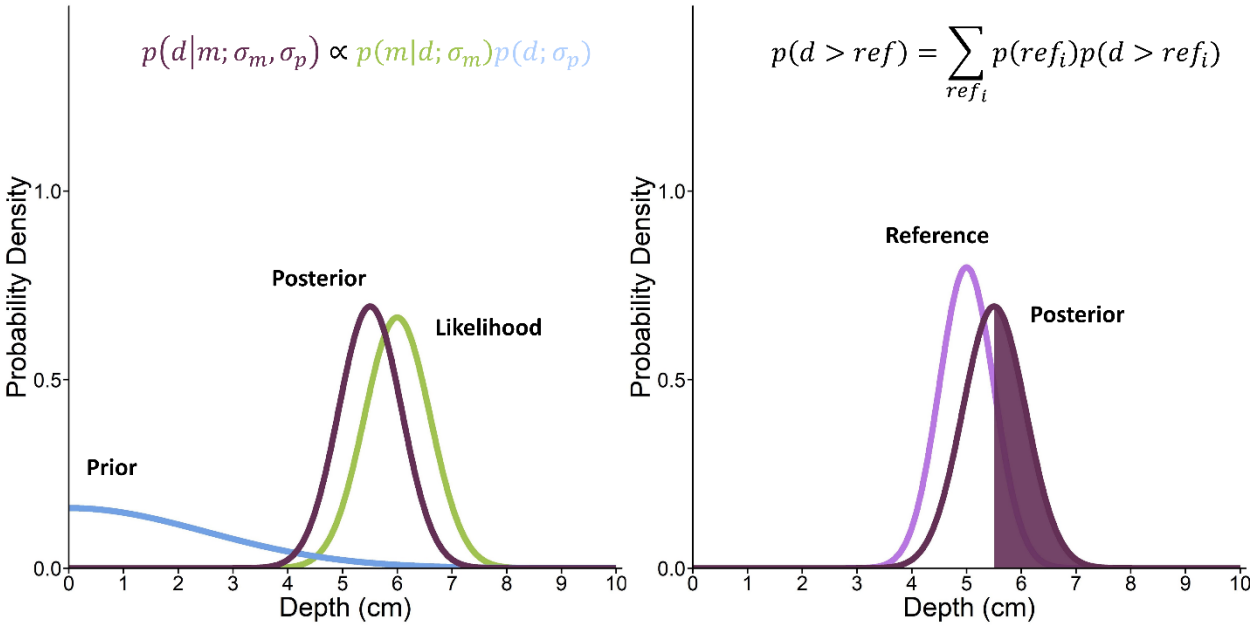


Figure 3.8. *An example of a simulated trial in which the perceived width of the front face is compared to a pyramid defined by motion parallax with a depth of 6 cm. The left illustration shows the first step of the Bayesian model that combines the likelihood of the motion parallax cue,* $p(m|d;\sigma_m)$ *and the prior distribution,* $p(d;\sigma_p)$*. The right illustration shows the comparison of the perceived depth of the pyramid to the perceived width of its front face defined by the mean* μ_{ref} *and standard deviation* σ_{ref} *from the human observer's size estimates. The shaded region shows the probability that the pyramid depth is greater than the reference for a hypothesized depth of 4.5 cm. Given the sum of this probability for all hypothesized depth values is large, the Bayesian observer would respond greater depth on this trial.*

To determine the combination method that best fit the human observers' performance in the combined cue condition, we calculated the combined condition for the Bayesian observer using (1) a linear, (2) a veto, and (3) a correlated combination method. When the likelihood and prior distributions are all Gaussian, the posterior is the weighted sum of the means of the two likelihood (binocular disparity and motion parallax) and the prior distribution. Here the weights for each distribution are proportional to the inverse of the variances of each distribution, so greater weight is placed on the more reliable cue (Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003). Using this approach, the cues are integrated linearly, and optimal cue integration maximizes reliability (Ernst & Banks, 2002; Landy et al., 1995).

If one cue is highly unreliable, a viable strategy for the visual system may be to veto the less reliable of the two cues, akin to removing outliers in statistics (Landy et al., 1995). In this case, instead of averaging the two cues as in linear integration, a single more reliable cue is used, while the other is ignored. To assess if a veto strategy better predicts human performance, our veto model combined the likelihood of the most reliable cue with the prior to produce the combined posterior distribution, disregarding the least reliable cue. For seven out of the eight observers, depth from binocular disparity

was more reliable than depth from motion parallax, so for most observers their posterior distribution for the veto model was determined by,

$$p(d|b;\sigma_b,\sigma_p) \propto p(b|d;\sigma_b)\, p(d;\sigma_p).$$

The third approach applied here, the correlated error model proposed by Oruç et al. (2003), adjusts the optimal reliability according to the estimated correlation (ρ) between binocular disparity and motion parallax cues. The corrected reliability of the combined cue condition is defined as,

$$r_c = \frac{r_b + r_m - 2\rho\sqrt{r_b r_m}}{1-\rho^2},$$

where r_b and r_m are the reliabilities of binocular disparity and motion parallax cues defined by the inverse of the variances for each distribution, and ρ is the correlation between the two cues. As the correlation ρ between binocular disparity and motion parallax increases, the weight applied to the more reliable cue increases. This model accounts for a linear, but suboptimal choice of weights by correcting the reliability of each single cue condition by $-\rho\sqrt{r_1 r_2}$. Thus, the inclusion of this model will capture if observers are using a linear combination method using suboptimal weights.

3.3.2 Human vs. Bayesian Observer Performance

To compare the best-fit model predictions to the empirical psychometric functions of each observer, the Bayesian observer's psychometric functions for the three combination methods above were fit using the same method applied to our human observers. Then, the Bayesian information criterion (BIC) for the combined cue condition between the observed and predicted models was calculated for each combination method (linear, veto, or correlated), for virtual and physical stimuli. The BIC accounts for differences in the number of parameters in each model by correcting for the number of degrees of freedom. To determine which Bayesian observer model best fit the observed data, we subtracted the BIC of the linear model from each combination model for virtual and physical stimuli (Appendix 3.B, Table 3.B1). If the minimum BIC difference was greater than 10, this was considered strong evidence that the model with smallest BIC difference was the best-fit to the human observers' performance (Raftery, 1995). For virtual stimuli, seven of eight observers' combined cue data were best fit by a veto model, while the remaining observer's data could be explained by either the correlated or veto models. None of the

observers showed a pattern consistent with linear integration for virtual stimuli. For physical stimuli, the outcomes were more variable: a veto strategy explained the results of six of eight observers (though one of these observers' data was also consistent with the correlated model). The remaining two observers' data was best fit by different models (one correlated cue and the other the linear). Overall, the veto model was the best-fitting model for the majority of observers in all viewing conditions. The predictions of the three models relative to each observer's PSE (Figure 3.B1) and JND (Figure 3.B2) are shown in Appendix 3.B.

3.4 Discussion

The aim of the current study was to evaluate the impact of unmodelled display-based cue conflicts, such as the conflict between vergence and accommodative distance on the depth integration of binocular disparity and motion parallax. To accomplish this, we assessed the relative accuracy and precision of depth estimates between virtual and physical truncated square pyramids in three cue conditions, (1) motion parallax, (2) binocular disparity, and (3) both cues combined. The purposeful replication of the physical environment in the virtual counterpart was essential to isolating the impact of these display-based cue conflicts from other, potentially confounding differences between the two environments. While the presence of display-based cue conflicts (such as accommodation) is commonly proposed as an explanation for underestimation of perceived depth in virtual environments, they are rarely studied directly. Given the similarity of performance across all conditions in virtual vs. physical environments, it is clear that the depth underestimates are not due to depth cue conflicts in the virtual stimuli.

3.4.1 Precision of stereopsis and motion parallax

Our results showed no difference in the precision of depth estimation from binocular disparity alone or when both motion parallax and binocular disparity were available, regardless of viewing condition (Figure 3.7). Further, observers were less precise when depth was defined by only motion parallax for virtual and physical targets. This is consistent with previous studies that show depth thresholds for disparity-defined corrugations are typically half of those defined by motion parallax (Bradshaw & Rogers, 1996; Bradshaw & Rogers, 1999; Rogers & Graham, 1982). The inclusion of motion parallax improves monocular depth discrimination thresholds in natural environments, but they typically remain higher than binocular thresholds (McKee & Taylor, 2010). The difference in reliability may be due in part to the somewhat awkward viewing conditions which require that observers maintain side-to-side

head motion at a constant rhythm while making discrimination judgements. While stereopsis only requires an estimate of absolute viewing distance and interocular distance to estimate depth, motion parallax requires an estimate of eye, head, and body position over time, along with absolute viewing distance. Further, if the object is moving, observers must correctly register and compensate for that motion (Helmholtz, 1925; Howard & Rogers, 2012). Thus, the precision of depth judgements for binocular disparity alone and motion parallax alone also depend on the precision of absolute distance information from vergence and the precision of motor and proprioceptive information from eye, head, and body movements, respectively.

Perceived depth from motion parallax was also less precise in the presence of display-based cue conflicts in the virtual environment relative to the physical environment. While this is likely due to the presence of conflicts in the virtual condition, even in the physical condition where cue conflicts are absent, depth estimates based on motion parallax remained less reliable than those from binocular disparity. As noted in the Introduction, most previous assessments of the precision of depth estimates based on motion parallax used virtual stimuli (Bradshaw & Rogers, 1996; Bradshaw & Rogers, 1999; Rogers & Graham, 1982). One difference between virtual and physical stimuli in the current study was the presence of update latencies that are inherent to HMDs. It is well-established that update latencies can be deleterious to performance in HMDs. Thresholds for latency detection range from 40 to 60 ms in the average observer (Adelstein et al., 2003) but can be as low as 17 to 33 ms (Adelstein et al., 2003; Ellis et al., 1999; Zhao et al., 2017). Even when latencies are subthreshold there is the potential for impact on performance of some tasks (Jay et al., 2007). According to the Oculus Rift SDK Performance Summary HUD, the motion-to-photon latency was approximately 19.4ms during our virtual motion parallax conditions. We cannot directly rule out the possibility that this update latency may have contributed to the reduction in precision in the virtual test conditions. However, the fact that the same loss of precision was evident in *both* the virtual and physical motion parallax conditions argues against this explanation.

3.4.2 Accuracy of stereopsis and motion parallax

Our size estimation task showed that observers underestimated the width of the front surface of the pyramids in virtual stimuli, relative to their physical counterparts (Figure 3.3). The catch trials confirmed that the perceived size of the front surface did not significantly change over the range of pyramid depths. Further, for virtual and physical stimuli observers estimated the surface width to be 55% and 66% of the true width, respectively. Size constancy of around 50% is typical for virtual stimuli presented on 3D displays and HMDs (Brenner & Van Damme, 1999; Hornsey et al., 2020). These results

are consistent with previous studies that show failures of size constancy for 3D lines presented at 85 cm even in natural scenarios with binocular disparities, motion parallax, shading, texture gradients, and accommodative blur cues present (Norman et al., 1996). Thus, it is not surprising that observers underestimate the width of the front surface of these stimuli, underscoring the utility of measuring this for each observer to improve the accuracy of our modelling.

Interestingly, while the perceived width of the reference surface differed for the virtual and physical stimuli, the resultant PSEs for virtual vs. physical pyramids revealed no relative difference in the magnitude of perceived depth (Figure 3.6). That is, although the reference appears larger for physical relative to virtual stimuli, once this is taken into account, the judgements of depth based on this reference are consistent across the two environments. This is likely due to the careful rendering of the virtual environment that minimized conflicts with other cues to depth and scale.

We found that depth judgments for virtual and physical viewing conditions (Figure 3.6) were made with similar accuracy. While a few studies have reported that depth judgements based on binocular disparity and motion parallax for virtual stimuli are comparable (Johnston et al., 1994; Rogers & Graham, 1979), others have shown that observers are more accurate in estimating depth when relying on stereopsis, compared to motion parallax, even for physical stimuli (Durgin et al., 1995; Sherman et al., 2012). We suggest that the apparent discrepancy reflects differences in the availability and consistency of monocular and binocular cues in these studies. For instance, Johnston et al. (1994) used cylindrical stimuli defined by circular texture cues (e.g., density and texture gradient) that were always consistent with the motion cue. Textures of homogenous regular circles provide strong foreshortening and linear perspective cues for texture scaling that improve accuracy of estimates of surface relief (Todd et al., 2007). Rogers & Graham (1979) tested observers in a dimly lit room; as a result, observers likely had sufficient absolute distance information to accurately scale depth from motion parallax and binocular disparity. On the other hand, the physical cone stimuli used by Durgin et al. (1995) were textured with a fine random dot pattern with minimal texture cues and Sherman et al. (2012) used physical trapezoids that formed a corner with strong perspective cues along the edges, and a random painted texture gradient with consistent density and perspective cues. Both of these studies presented their physical stimuli in isolation in a darkened room. Thus, the similarity of depth judgement accuracy for binocular disparity and motion parallax in our current study likely reflects the presence of strong homogenous texture cues, such as foreshortening, that provide additional support for depth percepts.

When both binocular disparity and motion parallax cues were present, depth estimates were as accurate as when either cue was viewed in isolation and as precise as when binocular disparity was

presented alone. Previously it has been suggested that the visual system could exploit the combination of depth from binocular disparity and motion parallax to obtain veridical depth estimates by using invariant properties of motion parallax to facilitate the interpretation of binocular disparities (Richards, 1985; Rogers & Graham, 1982). However, most studies show that depth distortions remain, even when both binocular disparity and motion parallax are available (Tittle et al., 1995; James T. Todd, 2004). While studies like that of Johnston et al. (1994) demonstrate that their combination results in more accurate judgements, their results may have been influenced by conflicts between binocular disparity and geometric cues. In our stimuli, depth from disparity ranged from 0.16 to 0.52 deg, consistent with the largely suprathreshold disparities common in natural scenes. It has been suggested that binocular depth judgements within this range do not tend to benefit from the presence of motion parallax, that is, that motion parallax only affects perceived depth in the presence of stereopsis for fine disparities below 0.13 deg (Rogers & Collett, 1989). However, it stands to reason that the point at which perceived depth benefits from motion parallax under binocular viewing is determined by the experimental context. For instance, Sherman et al. (2012) found that despite having fine binocular disparities (approximately 0.06 deg), when observers reported relative depth (rather than completing the matching task used by Rogers & Collett, (1989) perceived shape was not influenced by the simultaneous presence of motion parallax and binocular disparity. Our study supports the conclusion that the combination of depth from motion parallax and binocular disparity does not improve the accuracy of depth judgements more than either cue in isolation for virtual *or* physical objects over a wide range of disparities (Bradshaw et al., 2000). Further, the lack of difference between the virtual and physical judgements suggests presence of display-based cue conflicts in virtual environments does not appear to impact the combination of binocular disparity and motion parallax.

3.4.3 Combination Models

To determine the depth cue integration model that best fits the empirical data, we compared a Bayesian observer model with a (1) linear, (2) veto, and (3) correlated combination methods to human performance when binocular disparity, motion parallax, or both were present. The fact that precision did not improve when motion parallax information was combined with binocular disparity suggests that motion parallax does not aid depth estimates under binocular viewing (Durgin et al., 1995; Sherman et al., 2012). This result is contrary to the predictions of a weighted linear model which would predict that observers should be more accurate when multiple cues provide depth information. In the current study, if observers combined cues linearly then the presence of multiple cues should increase the accuracy of

depth judgements (i.e., the PSEs should be closer to each observer's perceived reference width). However, this is not the case (see Figure 3.B1 in Appendix 3.B).

Instead, our results show that most observers veto the information from the less reliable motion parallax cue and base their judgments entirely on the depth information from binocular disparity when both cues are present. There is some evidence that when binocular disparity and motion parallax cues are present and consistent, binocular disparity is weighted more heavily than motion parallax (Rogers & Collett, 1989; Tittle & Braunstein, 1993); however, our results are surprising as the only strong evidence of vetoing in the literature is obtained when these cues are presented in conflict. Commonly in such studies the conflict is extreme, for instance where the two cues define different surfaces, resulting in a rivalrous stimulus (Girshick & Banks, 2009; J. Farley Norman & Todd, 1995). In contrast, in our experiments all cues were consistent with the true depth of the pyramid for virtual and physical stimuli, but we find observers continue to rely exclusively on depth from binocular disparity. Our result echoes that of Norman et al. (1996) who showed that the combination of binocular disparity and motion parallax are only as accurate as the best individual modality in judgements of 3D line length.

However, other assessments of depth cue integration using binocular disparity and motion have shown that their combination in a depth-matching paradigm (when cues are consistent with the depth of the surface) results in improved accuracy and precision of relative depth judgements (Domini et al., 2006). A potentially important difference between our study and Domini et al.'s (2006) experiments is the type of motion parallax used. That is, in our experiments the motion parallax signal was generated by observer's head movements not by rotational or translational motion of the stimulus. While perceived depth has been shown to be similar under some conditions for the two types of motion (self vs. object) the equivalence depends critically on viewing distance and speed, both of which impact eye-movements (see Nawrot et al., 2014). Other studies have shown that parallax induced by self-motion enhances the accuracy of perceived slant (van Boxtel et al., 2003) and depth (Ono & Steinbach, 1990) by providing additional nonvisual information about the amount of relative motion between the observer and the display. While head velocity may play a smaller role than the object deformation in Bayesian models of slant estimation (Caudek et al., 2011), non-visual information from self-motion is likely used to stabilize the retinal image for a better measurement of optic flow (Cornilleau-Pérès & Droulez, 1994). Further, as outlined previously, theoretically head-motion provides the baseline distance (equivalent to the interocular distance for binocular disparity) used to compute depth. One explanation for the dominance of binocular disparity in our study is that the presence of binocular information was sufficient to scale and interpret the motion parallax information. For instance, motion parallax alone provides reliable

information about relative depth, but the non-visual information about viewing geometry from head movements are noisy relative to binocular convergence. Unlike previous assessments of binocular disparity and motion parallax (such as Domini et al., 2006), in our study strong homogenous texture cues provided additional support for depth percepts in all viewing conditions. The presence of this monocular information may have been sufficient to determine perceived depth in combination with binocular cues even if motion parallax was also being used in some manner. It is important to note that good stereoacuity was an inclusion criterion for our experiments, therefore none of the observers were stereo-deficient. It is possible that if observers have impaired stereovision, they may rely more heavily on motion parallax in scenarios where binocular cues are also present.

It is likely that the number of visual cues available, the nature of the task being performed, and the viewing geometry significantly affects the point at which the presence of motion parallax aids perceived depth judgements (Bradshaw et al., 1998). For instance, if observers are given sufficient audio and visual feedback on their performance on a 3D motion task, they can learn to exploit small motion parallax cues from head jitter (Fulvio & Rokers, 2017). Here we deliberately replicated the physical environment in its virtual counterpart. By generating stimuli that reproduce the real-world viewing geometry, minimize cue conflicts, and have at least one other consistent cue, we show that the failure of linear models in previous assessments of depth integration are not simply due to the presence of display-based conflicts. Our results show that the accuracy of depth estimates for virtual and physical objects are equivalent, and the method of combination does not differ between virtual and physical objects. This is good news for experiments that use displays. If virtual environments are carefully constructed, and other factors such as experience with stereoscopic displays are taken into account (Hartle & Wilcox, 2016), then the outcomes are generalizable to depth perception in the real world. However, the abundance of metric and ordinal depth information present in natural viewing environments could allow observers to utilize more complex methods of cue integration. We acknowledge that the Bayesian combination methods evaluated here may not capture all aspects of a full cue natural environment. The current study provides a great starting point, but further work is needed to understand the complexities of cue integration in complex cue rich natural environments.

3.5 Conclusion

We showed that depth estimates defined by binocular disparity, motion parallax, and their combination were remarkedly similar for virtual and physical stimuli. The accuracy of depth estimates was the same irrespective of the cue condition or whether the stimulus was virtual or physical. Depth

estimates were most precise when depth was defined by binocular disparity or the combination of binocular disparity and motion parallax for both virtual and physical stimuli. Depth estimates from motion parallax were less precise for virtual stimuli. Under natural conditions where 3D geometry is rendered correctly for suprathreshold volumetric stimuli, human observers do not combine depth information in an optimal linear fashion, instead they veto the information from the less reliable motion parallax cue. This occurs irrespective of the presence of display-based cue conflicts, such as conflict between vergence and accommodation, suggesting that previous failures of linear models of cue combination are not likely due to the presence of such conflicts, but instead to model assumptions.

CHAPTER 4: SCALING STEREOSCOPIC DEPTH THROUGH REACHING

Preface

The studies presented in Chapters 2 and 3 suggest that (1) depth distortions in virtual environments are due in part to an unreliable estimate of absolute distance, and (2) the combination of depth percepts from stereopsis and observer-induced motion parallax are subject to systematic biases in *both* virtual and physical stimuli. Like most studies of depth scaling, these studies have relied on stereoscopic displays with purely visual information. Given these well-established depth distortions from visual stimuli, the current chapter investigated the possibility that non-visual cues, such as distance information from reaching in depth, can be used to aid the scaling of stereoscopic depth. This was accomplished using a ring placement task with error-based feedback that depended on the accuracy of distance judgements.

The contents of this chapter was submitted for publication as a manuscript titled, *Scaling stereoscopic depth through reaching,* in the Journal of Vision and is currently under review (Hartle & Wilcox, JOV-08378-2022). Both authors contributed the conceptualization, design of methodology, review, and editing of the manuscript. Brittney Hartle completed all programming, performing of experiments, data collection, statistical analysis, and prepared the original draft of the published work. Laurie M. Wilcox supervised, managed, and coordinated responsibility for the research.

4.1 Introduction

Everyday activities that involve reaching for and picking up objects require that we estimate the object's position relative to our body and its three-dimensional (3D) shape. One of our most powerful sources of this information is stereopsis which can determine the relative position and 3D structure of an object using binocular disparities (in combination with interpupillary distance) if information regarding absolute distance is accurate and reliable (Wheatstone, 1838). Most studies of depth scaling have relied on stereoscopic displays with purely visual information from a static head position. The results consistently show systematic distortions of perceived depth, especially in virtual environments (Johnston, 1991; Rogers & Bradshaw, 1993). These biases in stereoscopic depth perception have often been attributed to incorrect scaling of binocular disparity via absolute distance.

There is an abundance of literature that shows our estimates of absolute distance in virtual environments are unreliable (Foley & Richards, 1972; Foley, 1980, 1985; Johnston, 1991; Rogers & Bradshaw, 1993; Wallach & Zuckerman, 1963; Witmer & Kline, 1998). The reported distortions in

perceived distance are often attributed to either (1) limited information signaling distance, or (2) display-based cue conflicts. The first of these stems from efforts to isolate stereopsis to obtain better experimental control, usually by removing other cues to depth. Unfortunately, this also limits the information available to support distance information; often only vergence remains, which alone is known to be unreliable (Brenner & van Damme, 1998; Foley & Richards, 1972; Rogers & Bradshaw, 1995). The second common source of distortion arises from well-known properties of 3D display systems, in conjunction with stimulus attributes (e.g., displayed distance and relative disparity). For instance, the well-known vergence-accommodation conflict occurs when an observer converges on an object presented at some distance from the focal plane of a display. Under these conditions, the normally synchronized accommodation and vergence signals conflict because accommodation is always fixed at the focal plane of the device (Hoffman et al., 2008; Wann et al., 1995). Of course, this is not the case with physical objects. Considerable empirical attention has been paid to the impact of the conflict between vergence and accommodation on depth perception. However, there are other potential sources of information that could help calibrate absolute distance that are commonly available in natural environments that have not been generally considered in this context. For instance, while depth scaling studies occasionally use the observer's hand to make their responses, these studies often rely on purely visual estimations of depth and distance. In the real world observers typically have the opportunity to reach out and touch objects in near space. This haptic interaction not only provides information about object shape, but may also provide information about the distance of objects from the observer.

The influence of misaligned visual feedback on sensory recalibration has been known for over a century (Helmholtz, 1909). It is thought that the visual manipulation (i.e., via prisms) prompts the recalibration and realignment of proprioception to match the distorted presentation of the hand (Redding & Wallace, 2006). Arguably, reaching to virtual objects in a virtual reality environment with known distance distortions is similar to reaching to targets with misaligned visual feedback from prisms. Evidence of such sensory recalibration has been observed following reaches along the depth dimension in virtual reality environments (van Beers et al., 2002). In such studies, observers adjust their reaching movements until the visual representation of the hand reaches the desired end point (Redding & Wallace, 1996), which occurs quickly and without the observer's awareness (Krakauer et al., 2000). This sensory recalibration depends heavily on error-based feedback (Shadmehr et al., 2010; Wei & Kording, 2009). Action-contingent visual feedback has been shown to be critical for learning for many sensory cues (Fulvio & Rokers, 2017). Here we examined the possibility that in virtual environments where distance estimation based on visual cues has been shown to be quite variable, the introduction of reach-based distance

information will improve the accuracy of distance judgements and consequently the accuracy of depth judgements from stereopsis.

During natural prehension movements it appears that reach distance and grip aperture rely on vergence and binocular disparity cues, but are calibrated independently (Bingham et al., 2007; Coats et al., 2008). For instance, vergence contributes mainly to reach planning, while binocular disparity contributes to grasping (Melmoth et al., 2007). Thus, the presence of stereopsis improves the accuracy of reaching and grasping actions (Jackson et al., 1997; Marotta et al., 1995; Marotta & Goodale, 1998; Servos et al., 1992; Servos & Goodale, 1994), but is the reverse also true? Can the act of reaching to objects provide additional cues to distance that, in turn, aid the scaling of binocular disparities? There is reason to think this is likely, for instance, distance information from proprioceptive cues from reaching-in-depth are potentially more precise than visual cues, such as vergence (Snijders et al., 2007; van Beers et al., 2002). There is also evidence that extending the perceived reach of the forearm using an adaptation paradigm increases the magnitude of observer's depth judgements (Volcic et al., 2013). While there is potential for reach to be an informative distance cue, reach could still be susceptible to similar depth distortions as stereopsis. For instance, depth judgements measured by reaching with an unseen hand show similar systematic distortions of depth (Foley, 1980, 1985). Recent studies have shown reaching and grasping demonstrate failures of depth constancy similar to those seen with stereopsis (Bozzacchi & Domini, 2015), but these systematic errors can be corrected with haptic feedback (Campagnoli et al., 2017). However, Bozzacchi and Domini (2015) showed a similar disruption of depth constancy in virtual and physical environments, which suggests that results may not be due to the absence of cues or presence of cue conflicts in virtual environments. Therefore, if observers are provided with error feedback to recalibrate reaching movements, then reaching to objects in a virtual environment could provide additional distance cues to aid the scaling of perceived depth. The ability to reach to objects could be one of the reasons we are better at judging the distance to objects in natural reaching and grasping tasks.

4.1.1 Experiment 1

As outlined above, the aim of the current study was to determine if we use distance information from reaching in depth to scale stereoscopic depth. To accomplish this, we designed a ring placement task that depended on the accuracy of distance judgements and gave observers incentive to improve their accuracy over the course of the task via feedback. Our ring placement task was inspired by a buzz-wire game, where a wire loop was moved along a path without touching the loop to the wire.

Performance on this type of task has demonstrated a strong binocular advantage (Read et al., 2013) and possible dependence on stereoacuity in past assessments (Murdoch et al., 1991). Our ring task incorporated a reach component by having observers guide a ring to a target peg using their index finger. Testing proceeded in 4 phases: (1) pre-reach depth magnitude, (2) proprioceptive assessment, (3) ring placement, and (4) post-reach depth magnitude. The pre-reach depth magnitude task was completed first to determine observer's baseline depth estimation accuracy for comparison with the post-reach depth estimation performance. The goal of the proprioceptive assessment task was to assess the accuracy of observer's blind reaches using the head-mounted display (HMD) and handheld controllers to ensure that using the handheld controllers did not introduce important biases in observer's reach judgements.

4.2 Methods

4.2.1 Observers

Fifteen observers with little to no prior experience with VR were recruited from York University. Their stereoacuity was assessed using the Randot^{TM} stereoacuity test to ensure observers could detect depth from binocular disparities of at least 40 arcseconds. All observers had normal to corrected-to-normal vision, and if necessary, wore their corrective lenses during testing. Fourteen of the fifteen observers were right-handed. The research protocol was approved by York University's Research Ethics Board.

4.2.2 Stimuli

4.2.2.1 Depth Magnitude

Figure 4.1 shows the stimuli for the depth magnitude task, which consisted of a rectangle (5.2deg by 7.4deg) surrounded by a reference frame (11.4deg by 13.7deg with a thickness of 2.3deg). The Voronoi texture for the rectangle and frame was generated using the voronoin() function in MATLAB with low contrast grey elements. The position of the points in the texture were randomly sampled from a uniform distribution and the Delaunay triangulation parameter was set to the default 'Qbb'. The position of the reference frame was jittered by +/- 0.5deg between trials, so the absolute position of the frame and rectangle could not be used as a reference for depth judgements. The reference frame was always rendered at a distance of 50cm and the central rectangle was always presented some distance in front of the reference frame. The position of the rectangle and frame were head locked, so any translation or rotation from the headset was nulled and the stimuli remained fixed to the center of the field of view. The distance along the z-dimension from the reference frame to the rectangle (i.e., the depth of the

rectangle) was from 0 to 5.0cm in half centimeter steps for a total of 11 depth steps. The disparity of the rectangle was calculated for each observer using their interpupillary distance and conventional binocular viewing geometry (see Howard & Rogers, 2012, pp.152-154).

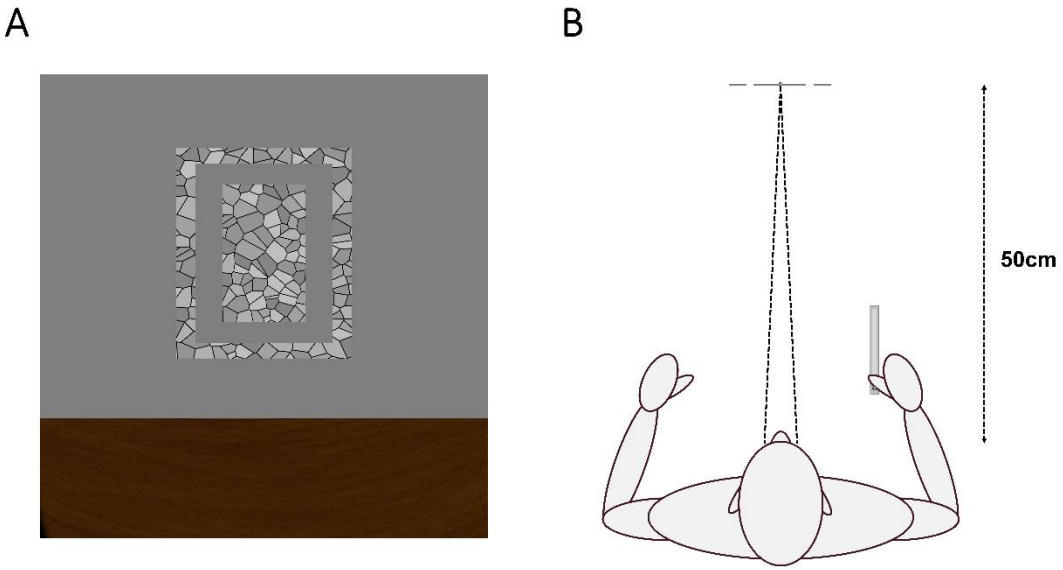


Figure 4.1. *Image A shows a screenshot of the rectangle and reference frame used in the depth magnitude task. Image B shows a top-down illustration of the stimuli for the depth magnitude task. The rectangle and reference frame were rendered 50cm from the observer.*

4.2.2.2 Ring Placement

In the ring placement task, rings and a target peg were rendered on a virtual wooden table (Figure 4.2). The position of the virtual table coincided with a physical table in the testing space. Five rings were randomly positioned on the table. A target peg was positioned 50cm in front of the observer on top of a rectangular base. The target peg diameter was 0.5cm with a height of 7.5cm. Thus, the top of the target peg was 14.5cm below the headset origin. The lateral position of the peg was randomly varied +/- 5cm between each trial. The ring positions were sampled from a uniform distribution between +/- 30cm in X, and 25cm to 50cm in Z. If the position of any two rings overlapped or intersected with the base of the target peg, their position was resampled. To make it easier to 'grab' the rings, they were placed 5cm above the top of the table.

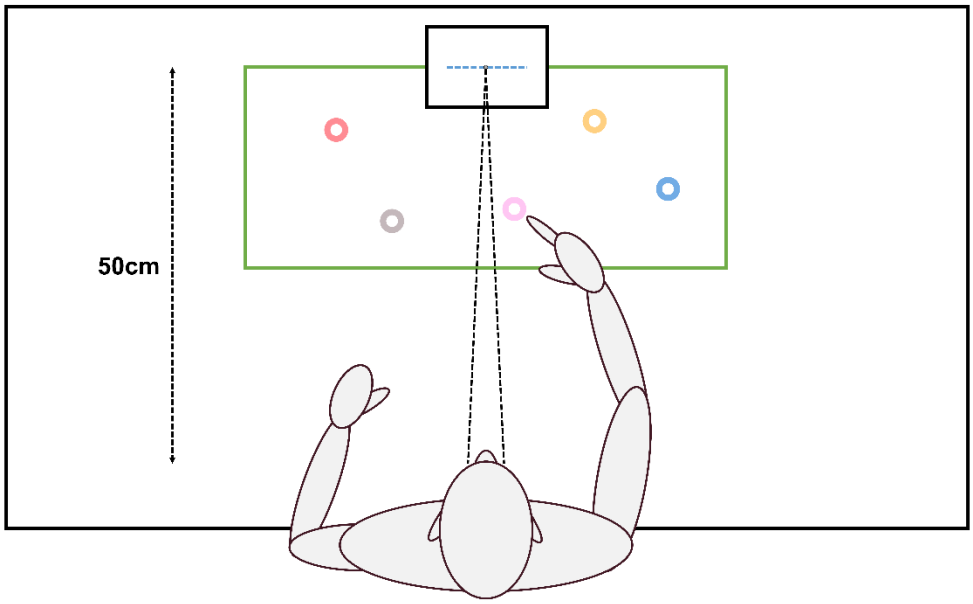


Figure 4.2. *A top-down illustration of the stimuli for the ring placement task. Each ring had an outer radius of 1.5cm and an inner radius of 0.75cm. The surface of the virtual table was 120cm wide by 65cm deep and positioned 32.5cm in depth such that its closest edge was positioned 32cm below the headset origin. The green rectangle represents the 25 by 60cm space in which the five rings were randomly rendered. The small black rectangle represents the 10 x 10 x 15 cm base on which the peg was placed. The grey horizontal dashed line represents the +/- 5cm space where the peg could be randomly placed.*

4.2.3 Apparatus

All stimuli were presented in the Oculus Rift CV1 HMD using the psychXR library in Python (Cutone & Wilcox, 2018). The Oculus Rift headset was connected to an Alienware Windows 10 computer with a NVIDIA GeForce GTX 1080 graphics card. This headset has two organic light-emitting diode displays, each with a resolution of 1080 by 1200 pixels per eye with a refresh rate of 90Hz and a horizontal field-of-view of 94 deg. Each pixel subtends 4.7 arcmin of visual angle (assuming a uniform distribution across the field-of-view). Prior to testing, each observer's interpupillary distance was measured using a digital pupillometer (GR-4) and the interocular separation of the HMD lenses was adjusted to match this separation. The Oculus touch controllers were used to track observers hand movements. The position accuracy error for the touch controllers were within acceptable ranges for measuring human kinematics, with an average error of 3.5 +/- 2.5mm for large 50cm movements along the z-axis (Shum et al., 2019). Observers rested their head on a chin rest to stabilize their head position for all test conditions.

4.2.4 Procedure

The four tasks were always conducted in the same order, (1) pre-reach depth magnitude, (2) proprioceptive assessment, (3) ring placement task, and (4) post-reach depth magnitude tasks.

4.2.4.1 Depth Magnitude

A suprathreshold depth estimation paradigm was used to assess the perceived depth between the rectangle and reference frame. Observers were asked to estimate the amount of depth between the rectangle and reference frame using a custom-built pressure-sensitive strip (for details Hartle & Wilcox, 2016). Observers rested their thumb against one end of the sensor strip and pressed their index finger along the length of the sensor to indicate the magnitude of their estimate. If the observers perceived zero depth between the rectangle and frame, they were instructed to place their thumb and index finger tightly together at the bottom of the sensor strip. We have successfully used this method to assess the accuracy of depth judgements in previous experiments (Hartle & Wilcox, 2016, 2021, 2022). The rectangle and frame remained visible until observers submitted their response. When observers pressed the spacebar, the recorded voltage was converted into millimeters. Observers completed the depth magnitude task before and after the ring placement task.

The data from the depth magnitude task was analyzed using a linear mixed-effects model using the nlme package in R (Pinheiro et al., 2015). This model examined the individual differences in depth estimates in repeated-measure conditions using random intercepts arranged in a hierarchy. A likelihood ratio chi-squared test determined the significance of fixed effects. Planned a priori comparisons for each fixed effect were evaluated using t tests. An approximation of Pearson's correlation coefficient (r) was used as a measure of effect size (Field et al., 2012). The slope of depth estimates for each observer was used to estimate the inferred viewing distance. We fit each observer's data using linear regression. A maximum likelihood estimation (MLE) method was used to estimate inferred viewing distance for each observer by fitting the slope of their linear functions according to the conventional formula for perceived depth that relates interpupillary distance, binocular disparity, and viewing distance (see Howard & Rogers, 2012, pp. 154).

$$\Delta d = \frac{D^2 * \delta}{IPD - \delta * D}$$

For each observer the binocular disparity of the rectangle (δ), and interpupillary distance (IPD) were known. When combined with their perceived depth judgements (Δd), these values were used to compute their inferred viewing distance (D). The estimate of inferred viewing distance for each observer provides insight into their perceived absolute distance for a given test condition.

4.2.4.2 Proprioceptive Assessment

We used a proprioceptive task to gauge the accuracy of observer's blind reaches using the handheld controllers. Observers held a physical target peg (mounted on a board) with one hand with their thumb on top of the peg. The target peg was an Edmund Optics post holder with a stainless-steel mounting post positioned in front of the participant at 50cm. Their task was to locate the top of the target peg by touching the index finger of their other hand to the top of a board placed above the target.

During testing, observers held the target peg with one hand and the touch controller in the other. At the beginning of each trial, observers placed their index finger (guided by a tracking dot with a diameter of 0.4cm) inside of a green sphere (radius = 1.5cm), before pressing a button to start the trial. The sphere was placed 5cm from the midline, 20cm below the head origin at a viewing distance of 25cm. This was done to reset their hand position and ensure the extent of their reach was similar in each trial. To keep the tracking dot roughly on the observer's fingertip observers wore a simple brace that held their index finger straight and the distance to the tip of their index finger was measured for each observer. A strip of Velcro attached the brace to the controller. Once the trial started the tracking dot was replaced with a uniform grey background. The observer was asked to reach out with their index finger and touch the board where they thought the peg (and their other hand) was located. When they were content with their estimate, they pressed a controller button to confirm their response. After their response was submitted, the next trial did not start and the tracking dot did not become visible until their index finger was 10cm away from the top of the target peg. Thus, the absolute position of the tracking dot relative to the peg could not be used as a reference for their judgements. Observers completed this task with both hands in separate blocks. Half of observers started with their right hand, and vice versa. Observers completed 15 trials with each hand, for a total of 30 trials.

4.2.4.3 Ring Placement

On each trial observers viewed five rings positioned quasi-randomly on a virtual wooden table. Their task was to place all five rings on the target peg as quickly as possible without touching the inner edge of the ring to the peg. A red tracking dot (0.4cm diameter) indicated the position of the index finger

in the virtual space. To pick up a ring observers guided the tracking dot to the ring and held the trigger on the controller. The same brace and tracking dot that was used in the proprioceptive assessment was used in this task. While holding the trigger button, observers guided the ring (rendered at the tip of their index finger) to the top of the peg, slid it down the peg until it reached 2.5cm from the base where it counted as a successful placement. If the inner edge of the ring "touched" the target peg, the controller vibrated, and the observer lost "accuracy points". Accuracy was calculated as the amount of time the ring did not touch the peg divided by the total time to successfully place each ring (as a percentage). Once all five rings were placed on the target peg, observers received feedback about the time and accuracy of each ring and the total for the trial. Observers were instructed to try to beat their previous time on each subsequent trial, while keeping their accuracy above 75%. If observers accidentally dropped the ring by releasing the trigger button, its position was reset. Observers could place the rings in any order they chose, and between each trial the lateral position of the peg and the position of the five rings were varied. The observers completed 12 trials for a total of 60 ring placements per session.

4.3 Results

4.3.1 Depth Magnitude

Figure 4.3.A shows the mean perceived depth as a function of predicted depth for the pre-reach and post-reach sessions. Observers overestimated the perceived depth of the rectangle in all viewing conditions. After completion of the ring placement task, the slopes of the depth estimates were significantly more shallow, $b=-0.23$, $t(268)=-3.59$, $p < 0.001$, $r=0.21$. To assess the accuracy of depth judgements, we calculated the root-mean-square error (RMSE) relative to perfect accuracy (i.e., the distance between observed points and the theoretical line in Figure 4.3.A). The average RMSE values (Figure 4.3.B) were significantly lower for depth estimates after the ring placement task, $F(1,14) = 4.72$, $p=0.047$, $r=0.19$. Thus, perceived depth judgements were more accurate after observers performed the ring placement task.

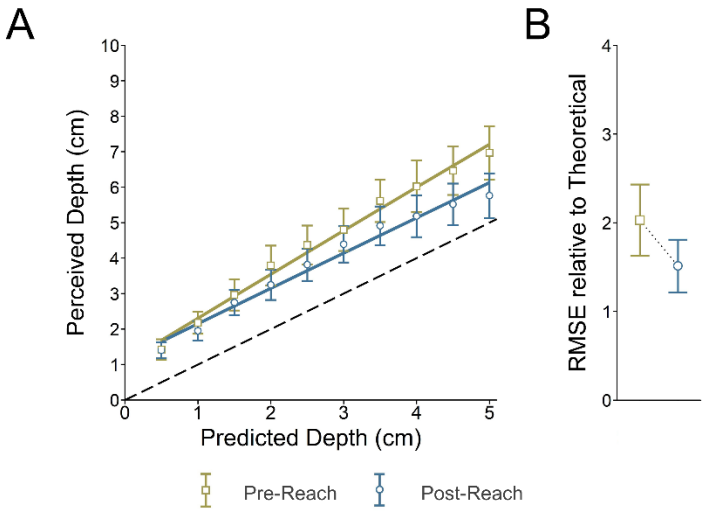


Figure 4.3. *Graph A shows the average perceived depth as a function of predicted depth for the pre-reach (green squares) and post-reach (blue circles) sessions for the ring placement task. Graph B shows the average root-mean-square error for the pre-reach and post-reach sessions. The error bars represent the standard error of the mean. The black dashed line in Graph A represents ideal performance.*

4.3.2 Inferred Viewing Distance

To compare depth estimates to distance, we calculated inferred viewing distance based on the magnitude of each observer's estimates as described in the Methods. Figure 4.4 shows the individual (Figure 4.4.A) and average (Figure 4.4.B) inferred viewing distance estimates for the pre-reach and post-reach sessions. On average, prior to reaching viewing distance was overestimated, but after completing the ring placement task, the slope was significantly more shallow, $F(1,14) = 8.38$, $p=0.012$, $r=0.26$, and more consistent with the simulated viewing distance. Seven of the fifteen observers showed a significant reduction in the slope of depth estimates after the ring placement task, while eight observers showed no significant change in slope (Figure 4.4.A).

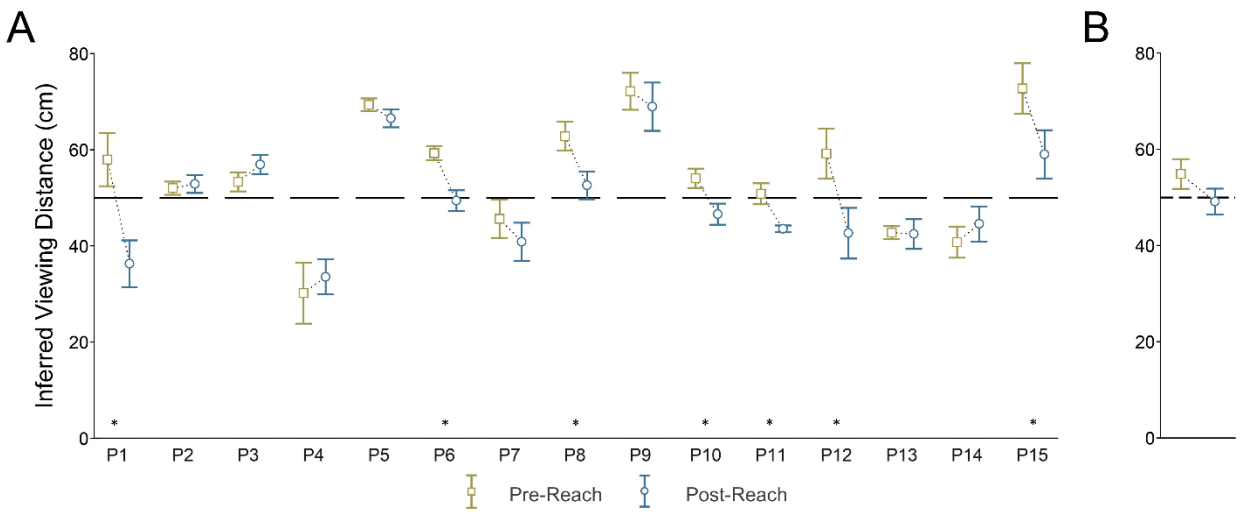


Figure 4.4. *Graph A shows the estimated inferred viewing distance values for the pre-reach (green squares) and post-reach (blue circles) sessions for each observer. Error bars represent 95% confidence intervals. Graph B shows the average inference viewing distance for all observers. Error bars represent the standard error of the mean. The horizontal dashed lines in both graphs represent the true viewing distance of the reference frame. The asterisks show which observers had a significant reduction in inferred viewing distance after the ring placement task.*

4.3.3 Proprioceptive Assessment

Figure 4.5 shows the blind reach estimates for the proprioceptive assessment task for both right and left hand reaches. Figure 4.5.A shows that the distance of the majority of reaches in the initial proprioceptive assessment were underestimated, and too close to the midline for both hands. The reach errors were less accurate on average for the left hand (7.12cm) than the right hand (5.28cm), which is expected given most observers were right-handed (Figure 4.5.B).

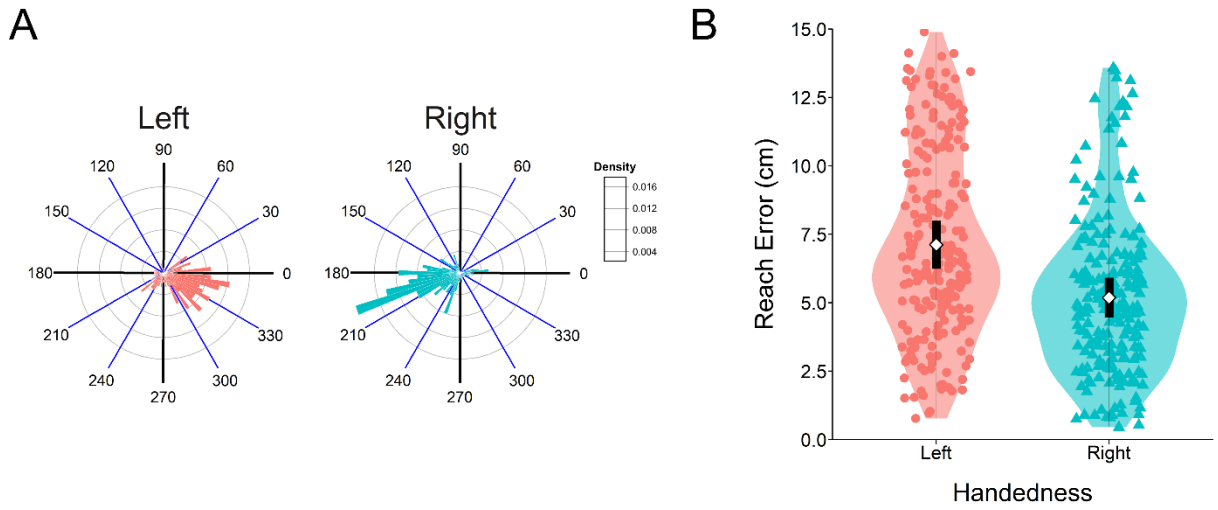


Figure 4.5. *Graph A shows a radial histogram of the blind reach estimates for each hand for all observers in the proprioceptive assessment task. A value of 90 deg is an error further than the target peg, and 270 deg is an error too far in front of the target peg. Graph B shows the magnitude of all reach errors made in the proprioceptive*

assessment task in cm for both hands. The white diamond represents the mean of the distribution, and the black box represents the standard error of the mean. The shaded distribution represents a density estimation that was fit using a Gaussian kernel with a smoothing bandwidth using Silverman's rule-of-thumb (or 0.9 times the minimum standard deviation and interquartile range divided by 1.34 times the sample size to the negative one-fifth power). This density estimation is plotted twice, once on each side of the boxplot for each condition.

4.3.4 Ring Placement

Figure 4.6.A shows an example of one observer's reach data as they placed rings on the target peg during a single trial. The observer was quick to correct an error when they received the vibrational feedback from the handheld controller. This result was typical of many observers. Figure 4.6.B shows the direction of all reach errors made during the ring placement task. Most errors were underreaches, consistent with the observation that observers commonly underestimate distances in virtual environments.

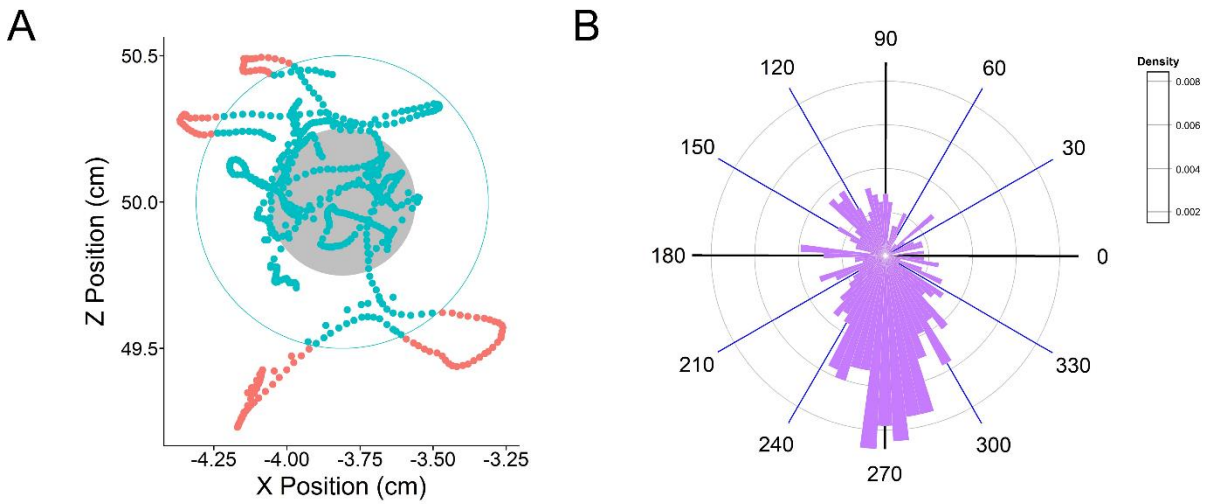


Figure 4.6. *Graph A shows an example of one observer's position data from a single trial. The points represent the positions of the center of the rings as they move down the target peg. The grey circle represents the peg itself and each point represents the position of the center of the ring. Each red dot outside the blue circle represents an "error", where the inner edge of the ring "touches" the target peg. Graph B shows a radial histogram that shows the direction of all errors made during the ring placement task. A value of 90 deg is an error further than the target peg, and 270 deg is an error too far in front of the target peg.*

4.4 Experiment 1 Discussion

Within peripersonal space, observers overestimated the amount of depth between two surfaces and the viewing distance to objects (Figure 4.3). After completing the ring placement task, observer's depth estimates were significantly more accurate (Figure 4.3), and the slope of their estimates were more shallow, consistent with theoretical predictions (Figure 4.4). Seven of fifteen observers showed a significant reduction in slope after the ring placement task, while the others showed no significant difference in the pre- and post-reach conditions. Interobserver variability is typical in assessments of

perceived depth judgements (Mon-Williams et al., 2000), which can depend on the level of experience with 3D displays (Hartle & Wilcox, 2016), especially when cues other than stereopsis are limited (Allison & Howard, 2000; Sato & Howard, 2001; K. A. Stevens & Brookes, 1988).

The majority of reach errors made during the ring placement task were underreaches, consistent with the finding that observers underestimate perceived distance in virtual environments. These reach errors were similar to those reported in our proprioceptive assessment where observer's blind reaches were underestimated and too close to the midline for reaches with both the left and right hand. This pattern of reach errors is consistent with biases in egocentric distance reported for reaches where the vision of the hand is obscured (Bozzacchi et al., 2014, 2016). The consistency between these studies confirms that our results are not specific to the use of the Oculus touch controllers.

Overall, our results are consistent with systematic distortions of perceived depth and distance reported in virtual environments, where the perception of distance is compressed, and depth perception is underestimated as distance increases. For example, Campagnoli et al. (2017) also showed that reaches at a similar viewing distance (45cm and 55cm) were consistently underestimated, while perceived depth at the same distances was overestimated. The results of Experiment 1 show that if observers have an opportunity to reach to virtual objects with error feedback their depth estimation accuracy improves.

4.4.1 Experiment 2

A potential concern with our interpretation of Experiment 1 is that our observers completed the magnitude estimation task twice, so the improvement in accuracy could be due to practice. To evaluate the role of practice, in Experiment 2 we replicated the study, using the same pre and post estimation task, but with comparable versions of an intervening reach and no-reach task. The new reach and no-reach tasks were visually similar and took approximately the same amount of time to compete. For the reach task, observers reached to a rectangle with a reference frame, while in the no-reach task observers performed the same task without reach using head movements alone.

4.5 Methods

4.5.1 Observers

Twelve and eight observers were recruited from York University for the reach and no-reach tasks respectively. Their stereoacuity was assessed using a Randot^{TM} stereoacuity test to ensure observers could detect depth from binocular disparities of at least 40 arcseconds. All observers had normal to corrected-to-normal vision, and if necessary, wore their corrective lenses during testing. Of the twelve

observers that participated in the reach condition, eleven were right-handed. The research protocol was approved by York University's Research Ethics Board.

4.5.2 Stimuli

The same depth magnitude and proprioceptive tasks were used in this experiment as described in Experiment 1. The depths and distances used in both tasks were identical. The stimuli for both the reach and no-reach tasks consisted of the same Voronoi reference frame (11.4deg by 13.7deg with a thickness of 2.3deg) as the previous depth magnitude task. A dark grey target square (1.15deg) was presented at a random position within a central 7.4deg by 5.7deg space within the reference frame. The reference frame and target square were always presented at a viewing distance of 50cm.

4.5.3 Apparatus

Stimuli were presented in the same Oculus Rift CV1 HMD used in the previous experiment. Observers rested their head on a chin rest to stabilize their head position in all conditions.

4.5.4 Procedure

4.5.4.1 Depth Magnitude

The same suprathreshold depth estimation paradigm used in Experiment 1 was used in Experiment 2. The stimuli, procedure, and analysis were identical to Experiment 1.

4.5.4.2 Reach Task

In the reach task, observers were asked to reach out and "touch" a target square that randomly appeared within a reference frame. The target and reference frame were rendered on a uniform grey background. As in the proprioceptive assessment in Experiment 1, at the beginning of each trial observers had to place their index finger inside a green sphere. The position and size of the green starting sphere and tracking dot was the same as that of the proprioceptive assessment. In addition, we matched the dynamic relative depth cues present in Experiment 1. That is the start and end point of the reaches in Experiment 2 spanned the same relative distance between 25 and 50cm from the observer. Further, as in Experiment 1, the size of objects (e.g., tracking dot) scaled with distance providing relative size (i.e., looming), occlusions, absolute stimulus disparity (i.e., changes in vergence induced by changing disparity), and relative disparity in relation to other stationary features (Howard & Rogers, 2012). Observers wore the brace described in the previous experiment. After the observer started their trial, their task was to

guide their index finger to the target square. The tracking dot was visible during the entire trial. When the tracking dot was within a radius of 0.5cm from the center of the target square the controller vibrated and once the observer held their finger there for 1.5 seconds it counted as a successful reach. The observers completed a total of 60 trials.

4.5.4.3 No-Reach Control

In the no-reach task, observers were asked to "look at" the target square by rotating their head. The head direction was tracked by casting a ray from the headset origin along a path parallel to the observer's line of sight. At the beginning of the trial, observers had to rotate their head towards the green target sphere to start the trial. Once the ray intersected with the target sphere (radius of 4deg), and the observers pressed a controller button, the trial started. On each trial, the target square was randomly presented within the reference frame. When the ray casted from the headset intersected an invisible sphere with a radius of 4.6deg at the position of the target square for 2 seconds, it counted as a successful response and testing proceeded to the next trial. The size of the spheres for ray casting were chosen such that the task was difficult enough that the duration of the no-reach task was similar to the reach task. As in the reach task, observers completed 60 trials in total.

4.6 Results & Discussion

4.6.1 Reach Control

Figure 4.7.A shows the mean perceived depth as a function of predicted depth for the pre-reach and post-reach sessions in cm. There was no significant difference in the intercepts between the pre- and post-reach sessions, b=-0.17, t(11)=-0.56, p=0.59, r=0.17, nor was there a significant difference in the slope between the two sessions, b=-0.04, t(214)=-0.44, p = 0.66, r=0.03. Inferred viewing distance was not assessed for this task because there was no difference in the overall slope between pre and post sessions. However, the RMSE was significantly lower in the post-reach session than the pre-reach session, F(1,11) = 5.16, p=0.044, r=0.12 (Figure 4.7.B). The results of Experiment 2 show that when observers reach to virtual objects their subsequent depth estimates are slightly, but significantly, more accurate. Despite the significant improvement in accuracy after reaching in Experiment 2, unlike Experiment 1 (Figure 4.3) there was no change in the slope of the function relating predicted and perceived depth (Figure 4.7). Further, the effect size for the improvement in accuracy (i.e., RMSE) was smaller (i.e., r=0.12 or d=0.24) than the moderate effect in the ring placement task in Experiment 1 (i.e., r=0.19 or d=0.38, for conversion from r to d see Rosenthal, 1994). Taken together, these data suggest that there was

information present in the ring placement task that facilitated the improvement in inferred viewing distance (i.e., slope) that was absent in the reach task in Experiment 2.

As in Experiment 1, the proprioceptive assessment in Experiment 2 showed a bias towards underestimation (see Appendix 4.A). While the observers in Experiment 2 made slightly larger errors, the pattern was the same as in the first study. Thus, the pattern of errors cannot account for differences between the two experiments.

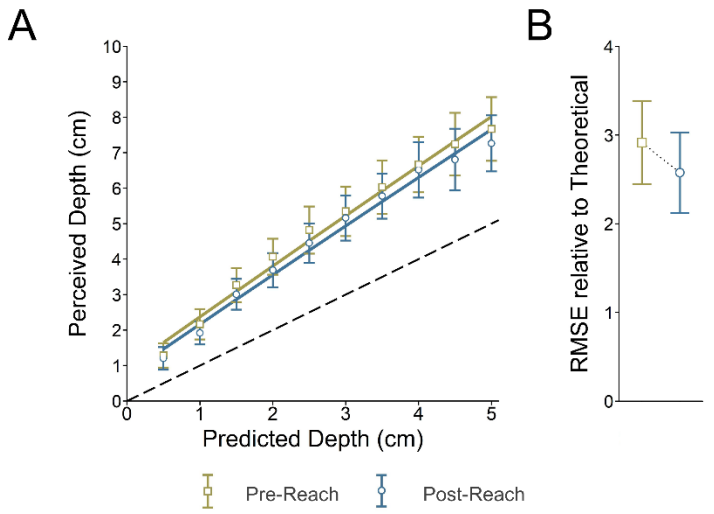


Figure 4.7. *Graph A shows the average perceived depth as a function of predicted depth for the pre-reach (green squares) and post-reach (blue circles) sessions for the reach control task. Graph B shows the average root-mean-square error for the pre-reach and post-reach sessions. The error bars represent the standard error of the mean. The black dashed line in Graph A represents accurate performance.*

4.6.2 No-Reach

Figure 4.8.A shows the mean perceived depth as a function of predicted depth for the pre-no-reach and post-no-reach session in cm. This group of observers were slightly more accurate overall than the group of observers that performed the reach task. There was no significant difference in the intercepts between the pre- and post-no-reach sessions, b=0.27, t(7)=0.88, p=0.41, r=0.32, nor was there a significant difference in the slope in the two sessions, b=-0.15, t(142)=-1.70, p = 0.09, r=0.14. Like the previous condition, inferred viewing distance was not assessed here because the overall effect of slope was non-significant. Unlike the previous reach task, here there was no significant difference in the RMSE between the pre- and post-no-reach sessions, F(1,7) = 0.01, p=0.93, r=0.01 (Figure 4.8.B). When observers completed a no-reach task where they only turned their head to look at a target object, their subsequent depth estimates showed no improvement in accuracy. Thus, Experiment 2 confirmed that it was likely the reach component, not practice that contributed to the improvement in the accuracy of depth judgements.

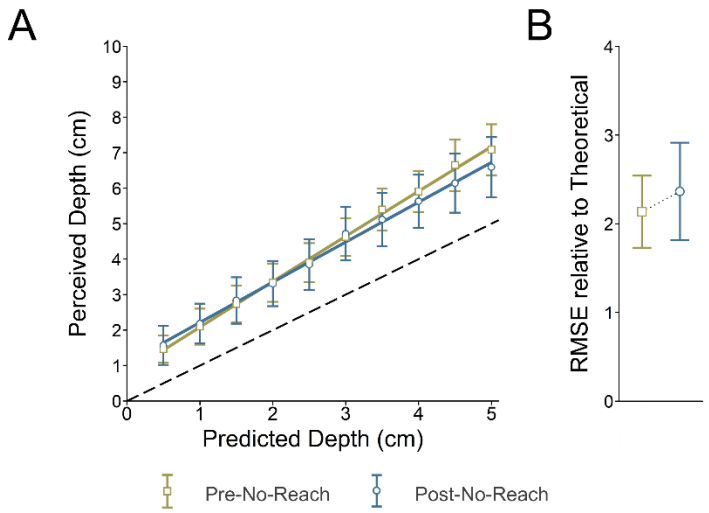


Figure 4.8. *Graph A shows the average perceived depth as a function of predicted depth for the pre (green squares) and post (blue circles) sessions for the no-reach control task. Graph B shows the average root-mean-square error for the pre-no-reach and post-no-reach sessions. The error bars represent the standard error of the mean. The black dashed line in Graph A represents accurate performance.*

4.7 General Discussion

Our experiments show the potential benefits of actively interacting with virtual objects on our ability to estimate depth from binocular disparity within the same region. However, the fact that depth scaling (i.e., slope) was improved in Experiment 1, but not Experiment 2, suggests that there was additional information in Experiment 1 that facilitated the impact of reach. To fully appreciate the impact of reach on perceived depth, these additional sources of depth or distance information and how they potentially interact with reach must be considered. In addition to the task changes, the structure of the virtual environment was different in Experiments 1 and 2. In our first study, the environment resembled a more natural everyday scene that consisted of a virtual wooden table with rings at various distances from the observer (Figure 4.2). Whereas in Experiment 2, observers reached with limited interaction to a target within a rectangular frame rendered on an empty uniform grey background. While reaching-in-depth did improve the overall accuracy of depth judgements in both experiments, only reaching in the structured environment in Experiment 1 subsequently improved depth scaling. These results suggest that there was additional distance information (other than the reach) present in the more natural environment in Experiment 1 that facilitated distance estimation (and therefore subsequent magnitude estimation). The presence of these environmental cues, or their interaction with reach, improved subsequent depth scaling.

One distance cue common in natural reaching tasks that was present in the ring placement task, but not the reaching task in Experiment 2 is the relative height of objects in the visual field. The elevation

of an object in a visual scene provides information regarding its distance (Gardner & Mon-Williams, 2001; Marotta & Goodale, 1998; Ooi et al., 2001), where higher objects are perceived as further away (Dunn et al., 1965; Epstein, 1966). For instance, if the ground plane is level and visually perceived eye level coincides with the true horizon, distance could be determined based on eye height and the angle of declination from horizontal. Observers can learn the relationship between height in the field and distance of objects on a plane below eye level (Marotta & Goodale, 1998). In Experiment 1, observers were also likely more aware of the distance to the virtual table, given it coincided to a physical table in the lab space. (This physical table was present in all experiments, but only rendered in Experiment 1.) The changes in elevation of rings between trials can cause an even larger effect on observer's perceived distance judgements by providing pictorial cues (i.e., local perspective and shape cues) that varied with the distance of the rings (Marotta & Goodale, 1998; Sedgwick, 1986). Further, given that the observer interacted with the rings at multiple distances within the virtual scene they had access to additional samples of visual and proprioceptive cues to distance at multiple positions in the scene. Indeed, most psychophysical assessments of reaching and grasping use a similar setup with objects displayed on a tabletop at various distances, which provides important relative distance cues according to their height in scene (Jakobson & Goodale, 1991; Servos et al., 1992).

It is possible that these strong distance cues facilitated the impact of our reach manipulation by also interacting with other distance cues critical to reach planning. For example, while distance information from vergence alone is known to be unreliable and cause distortions in the scaling of binocular disparities (Brenner & van Damme, 1998; Foley & Richards, 1972), there is good evidence that vergence plays an important role in the planning of reach movements (Melmoth et al., 2007; Mon-Williams & Dijkerman, 1999; Tresilian et al., 1999), particularly when it is combined with other distance cues (Mon-Williams et al., 2000). In comparison, other than vergence, the sparse environment used in Experiment 2 provided little additional support for distance estimation. Importantly, the improvement in the accuracy of depth estimation in Experiment 1 following reaching-in-depth occurred regardless of the presence of display-based cue conflicts between accommodative and vergence distance in the virtual environment. Thus, our results suggest that without a structured virtual environment with other distance cues, the impact of distance information provided by reaching-in-depth is weak. The key to improving accuracy and the scaling of depth judgements appears to be the combination of reaching to multiple positions in the scene and the presence of other reliable distance cues that are often present in natural environments. While outside the scope of this study, it remains to be determined which of these additional sources of information best support distance calibration via reaching.

We demonstrated a significant impact of reach on the accuracy of subsequent depth judgements, however, the mechanism underlying this improvement is unclear. The benefit of reach could be due to an overall improvement in the accuracy of perceived distance within near space or the result of calibrating depth judgements to a specific reaching distance (50cm). In Experiment 1 observers initially overestimated distance (56cm, 95% CI [54.5,57.2]), and after repeatedly reaching to a target at a viewing distance of 50cm, inferred viewing distance was significantly lower and more similar to the actual distance and the reached distance (50cm, 95% CI [48.3, 51.7]). However, this result is consistent with a general calibration or a specific calibration to the reaching distance. These two scenarios would predict different outcomes under conditions where the trained reach distance and distance to the magnitude estimation target were different. While we cannot dissociate these outcomes in the current study, there is evidence that the scaling of binocular disparity is closely tied to the magnitude of reaching distance. For instance, Volcic et al. (2013) showed that reach adaption that extends the perceived reaching distance, increases the magnitude of perceived depth. However, in their adaptation paradigm they manipulated the perceived reach of the forearm by creating a conflict between the distance of visual cues and the observer's reach. The same effect may not be achieved in an environment with consistent cross-modal distance information. Thus, while our results show that reach improves the subsequent accuracy of perceived depth judgements, further experimentation is required to determine the generalizability of this cross-modal calibration.

4.8 Conclusion

We showed that when observers have an opportunity to reach to virtual objects with error-based feedback the accuracy of their subsequent stereoscopic depth judgements improves. This is consistent with reports that the inclusion of haptic feedback allows the visuomotor system to compensate for systematic errors (Campagnoli et al., 2017). We believe the improvement is due to generation of a more accurate sense of absolute distance. We show that the inclusion of strong distance cues that often occur in natural reaching and grasping scenarios as the hand moves objects within the scene further aids depth scaling. Further, the addition of natural haptic feedback (i.e., physical touch) that is present in natural reaching and grasping movements could further improve the contribution of reaching by refining visuomotor interactions (Ozana et al., 2020). In cases where accurate distance and depth perception is necessary for training in virtual environments, it could be beneficial to have an initial calibration phase where observers reach to different peripersonal distances to aid scaling of the virtual space.

CHAPTER 5: GENERAL DISCUSSION

5.1 Summary

Given the complex and diverse literature addressing depth constancy and depth cue integration, it is difficult to determine what information is critical to support an accurate and reliable sense of depth, especially in virtual environments. To breakdown this complex problem into manageable components this series of studies addressed three major questions surrounding the impact of complex viewing environments and limitations of computerized displays on distortions of stereoscopic depth perception. The aim of the series of studies was to compare virtual and physical objects to learn more about the contribution of various monocular, binocular, and extraretinal factors to an accurate and reliable sense of depth.

The experiments in Chapter 2 evaluated if conflicts between oculomotor distance cues produced by computerized display systems drive the underconstancy of stereoscopic depth perception in virtual environments. The results demonstrated that only cue rich physical stimuli supported depth constancy while all depth estimates based on virtual stimuli rendered on computerized display systems showed consistent underconstancy. Given the similarities of the virtual and physical test conditions, the suboptimal performance was likely due to the decoupling of accommodation and vergence in these devices; however, the degree of the conflict did not affect perceived depth magnitude under these conditions. Interestingly, the observers' prior experience with the physical and virtual test environments had a dramatic effect on the accuracy and constancy of their depth judgments. This suggests that attention should be paid to the observer's familiarity with test environments especially when using similar tasks or technology, even after a brief exposure. Further, this highlights the importance of thoroughly examining the consequences of counterbalancing conditions and not just assuming that any order effects in the data are mitigated.

The studies in Chapter 3 determined if the limitations of virtual test environments contributed to the lack of linear depth cue combination between binocular disparity and motion parallax observed in previous studies. The results showed that even under natural viewing conditions, observers did not optimally combine depth information from binocular disparity and motion parallax. Instead, they relied heavily on depth from binocular disparity in lieu of the less reliable motion parallax cue, irrespective of the presence of display-based cue conflicts in virtual stimuli. This suggests that previous failures of linear models were not to the limitations of virtual stimuli.

Lastly, Chapter 4 took a different approach than Chapters 2 and 3, by looking beyond stereoscopic displays with purely visual information. The objective of the studies in Chapter 4 was to determine if interacting with a virtual environment via reaching in depth provides distance information that can be used to scale stereopsis. The results showed that when observers have an opportunity to reach to virtual objects with error-based feedback, the accuracy of subsequent depth judgements improved. If the virtual environment includes strong distance cues common to natural reaching and grasping tasks, then the depth scaling of subsequent judgements becomes more consistent with the viewing geometry. This suggests that a period of interaction with virtual objects benefits the perception of distance and depth in virtual environments even with known conflicts and ambiguities in visual distance information.

Overall, this series of studies demonstrates that it is important to consider not only the quality and quantity of depth information in a scene, but the context in which that information is presented, observers' prior experience with the environment, and how well that information corresponds to natural viewing environments. Despite inherent issues with virtual objects, other factors, such as previous experience with a similar real-world environment or reaching to virtual objects, mitigate depth distortions endemic to virtual environments. The influence of these other factors allows observers to gain a better understanding of the depth and distance of objects, regardless of the limitations of the visual scene.

5.2 What visual cues support stereoscopic depth perception?

As discussed in Section 1.2, studies of stereoscopic depth constancy typically focus on identifying the factors critical to support accurate absolute distance estimation for depth scaling.[5] Stereopsis is often assessed in sparse environments with little information to support accurate absolute distance information apart from the pattern of vertical disparities and the eye's vergence. While studies have shown that both vergence and vertical disparities play an important role in distance perception, especially for large stimuli at short viewing distances below 50cm (Backus et al., 1999; Mon-Williams et al., 2000; Rogers & Bradshaw, 1995), neither vergence, (Foley & Held, 1972; Gogel, 1961; 1977; Komoda & Ono, 1974), nor vertical disparities are sufficient to support constancy on their own (Foley et al., 1975; Rogers & Bradshaw, 1993). However, given that the scaling of stereoscopic depth depends heavily on access to

[5] It has also been proposed that absolute distance estimation and stereoscopic depth estimation are completely separate processes and that stereoscopic depth constancy is support by other invariant information, such as disparity curvature (Rogers & Cagenello, 1989; Vreven, 2006).

accurate and reliable information concerning the absolute viewing distance to the object, accurate depth scaling may require an environment with multiple consistent distance cues.

First, it is critical to clarify what is considered a depth cue in this context, as the term has been criticized for its ambiguity when used to describe the information that aids depth perception and the 'higher-level' assumptions required to use this information (Rogers, 2022). In this dissertation, the terms 'distance' and 'depth cue' are used to describe the type of information available in a visual scene that could be used by the visual system to determine the distance of an object from the observer, or the relative depth of a target, respectively. Some examples mentioned here include, binocular disparity, motion parallax, accommodation, vergence, texture gradient, size, and focal blur. The term information refers to a property of the visual scene that the cue describes. In this document, the focus is on cues that provide information regarding the relative depth and distance of objects, with specific focus on how various distance cues aid the scaling of stereopsis.

Given the breadth of literature documenting depth distortions from stereopsis and the many sources of variability between studies, such as observer, methodological, and stimulus differences, it is challenging to determine the critical sources of information that support depth scaling. One primary objective of this series of studies was to identify the contribution of display-based cue conflicts to distortions of stereoscopic depth by directly comparing depth judgements of carefully matched stimuli in virtual and physical environments. This was accomplished by limiting interobserver differences using a repeated-measures design, and carefully matching the consistency of depth cues between the two types of viewing environments. Further, to give stereopsis the best opportunity to attain accurate depth scaling, studies in Chapters 2 and 3 used more complex stimuli with other consistent monocular cues to surface curvature. This was motivated by other studies that have shown that monocular size and depth information influence stereoscopic depth constancy (Brenner & Van Damme, 1999; Collett et al., 1991; Foley, 1968; Mon-Williams et al., 2000). The current studies used stimuli with large circular texture gradients that varied in size and perspective that provided additional information regarding surface curvature to support stereopsis, but did not support accurate depth scaling in isolation (Figure 2.3, 2.4).

In Chapter 2, the comparison between carefully controlled virtual and physical objects revealed a consistent lack of depth constancy in virtual, but not physical objects, which suggested that the inherent decoupling of accommodation and vergence in the computerized displays was a critical factor. This is consistent with other studies that have shown the coupling of these cues increases the effectiveness of scaling depth from binocular disparity and reduces the degradation of depth perception (Hoffman et al., 2008; Watt et al., 2005). However, there is some debate over the effectiveness of accommodation as a

distance cue, as there are some instances of limiting accommodation in natural viewing environments that have little to no impact on depth constancy (Ritter, 1977). It is important to note that even though the viewing geometry in Chapter 2 was designed to minimize focal differences between the edge and peak of the half-cylinders, we cannot completely rule out that accommodative blur contributed to the improvement in depth judgements for physical objects. While the eye's depth of focus is approximately 0.33D in typically viewing scenarios (Campbell, 1957; Hoffman & Banks, 2010; Walsh & Charman, 1988), well above the estimated focal blur of our stimuli (0.15D and 0.06D at the near and far viewing distances, respectively), blur detection thresholds vary from 0.12 to 1.75D depending on the viewing scenario (Ciuffreda, 1998; Walsh & Charman, 1988). The presence of accommodative blur has been shown to be a critical factor in the improvement of depth judgements in natural viewing environments, especially at near viewing distances (Frisby et al., 1995; Watt et al., 2005). Thus, the attainment of stereoscopic depth constancy in natural viewing environments appears to be driven by the coupling of oculomotor distance cues, and possibly the presence of appropriate accommodative blur.

Chapter 3 took advantage of the comparison of virtual and physical objects to tackle a different open question in the cue integration literature on stereopsis. That is, whether the limitations of computerized display systems contributed to previous failures of linear cue combination models to accurately model the combination of binocular disparity and motion parallax. Previous studies have drawn inconsistent conclusions regarding the relative accuracy of depth from binocular disparity and motion parallax, where some show similar accuracy for both cues (Bradshaw et al., 2000; Johnston et al., 1994), while others show that depth estimates from motion parallax are less accurate than binocular disparity (Durgin et al., 1995; McKee & Taylor, 2010). Typically, studies show that depth is misperceived even when both cues are available (Scarfe & Hibbard, 2011; Todd, 1985; Todd & Norman, 2003), and these studies often attribute these misperceptions of depth to the possible influence of unmodeled conflicts between focus and motion cues (Scarfe & Hibbard, 2011). However, the results in Chapter 3 showed that regardless of the type of test environment, the presence of consistent motion parallax and binocular disparity information did not significantly improve the accuracy or precision of depth judgements beyond performance using binocular disparity alone. The results outlined in Chapter 3 are consistent with studies that show the visual system does not exploit the simultaneous presence of binocular disparity and motion parallax (Bradshaw et al., 2000). The lack of benefit of combining binocular disparity and motion parallax occurs irrespective of the presence of display-based cue conflicts, suggesting that previous failures of linear cue integration are not likely due to the presence of such conflicts.

Overall, Chapters 2 and 3 concluded that the visual cues that were critical to supporting stereopsis were the consistency of oculomotor distance cues. Under these viewing conditions, the addition of motion parallax did not improve the accuracy or precision of depth judgements beyond that seen when binocular disparity was presented alone, even when display-based cue conflicts were present. Thus, the absence of cue conflicts in real-world stimuli is the most likely explanation for achieving stereoscopic depth constancy in natural viewing environments, compared to virtual representations.

5.3 Physical Environments as a Validation Tool

The observed benefits of consistent oculomotor distance cues, such as accommodation (Frisby et al., 1995; Watt et al., 2005) and vergence (Durgin et al., 1995; Mon-Williams et al., 2000) to accurate distance perception, and consequently, stereoscopic depth perception in physical environments, highlights the value of real-world comparisons to laboratory measures of stereopsis. This observation echoes a more general concern regarding the generalizability of behaviour in laboratory studies. An important, but challenging question to address is whether a perceptual effect observed with virtual objects is due to the hypothesis being tested or technological limitations used to render the virtual environment. To answer this question, the comparison of performance between virtual and physical objects serves as an important form of validation.

There are a few examples of perceptual effects that are predominantly seen in virtual stimuli, that are diminished or completely vanish when evaluated in a full-cue physical test environment. One example mentioned in Section 1.2.2 is the dependence of binocular disparity and texture cue integration on the orientation of surface curvature vanishes for physical stimuli with accommodative blur cues (Buckley & Frisby, 1993; Frisby et al., 1995). Even in quite sparse physical environments, the presence of consistent relative size and focal blur cues in combination with binocular disparity eliminated depth distortions and interobserver differences that depend on level of stereoscopic experience (Hartle & Wilcox, 2016). Assessments of the effect of lateral retinal motion on stereoscopic depth discrimination thresholds have shown that thresholds in natural viewing conditions remain stable at higher velocities than their virtual counterparts (Cutone et al., 2019). The assessment of stereoscopic depth constancy in Chapter 2 is consistent with other studies that show that constancy and accurate depth perception is reached even in relatively sparse physical environments (Durgin et al., 1995; Mon-Williams et al., 2000). Thus, if the aim of research is to understand visual functioning under typical conditions, then it is useful to evaluate performance in real-world scenarios using naturalistic images or directly compare performance between virtual and full-cue natural viewing environments. As outlined below, this can be difficult achieve

with the requisite degree of experimental control, but at the very least researchers should be aware of the constraints on generalizability.

5.3.1 Challenges of Real-World Validation

A common assumption is that real-world stimuli sacrifice high experimental control for ecological validity, while virtual stimuli rendered on computerized displays provide the opposite. Chapter 2 and 3 demonstrate that there are ways to attain control in physical environments through well-designed apparatuses and experimental protocols. While generalizable results are particularly important for studies with direct real-world applications, concepts such as ecological validity and 'real world' have been criticized for being too broad and context dependent (Holleman et al., 2020). In general, ecological validity refers to environments and task demands that are characteristic of the real world, such that results tested in laboratory conditions will generalize to an equivalent scenario in the real world. However, what characteristics are critical to attain this generalizability are ill-defined. It is up to researchers to specify and describe the context-specific issue being addressed through real-world comparisons. In this dissertation, Chapters 2 and 3 took advantage of real-world validation to focus on the prominent issue of oculomotor distance conflicts between accommodative and vergence endemic to computerized display systems. If the experiment is well controlled and the research question is clearly specified, real-world validation through the comparison of a carefully matched virtual and physical environment is a powerful tool for determining if perceptual effects persist in real-world scenarios.

However, research in real-world environments present their own technical and perceptual challenges. Despite the results of Chapter 2 demonstrating that depth judgements of physical objects are both accurate and achieve depth constancy, this series of studies also presents two scenarios in which tasks using physical objects demonstrate consistent depth distortions: (1) when previous experience with virtual objects reduced the accuracy of subsequent depth judgements of physical objects (Figure 2.5), and (2) when observers estimated the size and depth of truncated square pyramids (Figure 3.3, 3.6). The first scenario occurred in Chapter 2 where the careful control of depth cues and estimation method between virtual and physical test environments revealed unexpectedly strong order effects. Despite controlling for the test order between environments, the observer's prior experience completing the same task in closely matched virtual and physical test environments strongly affected depth judgements in the subsequent viewing environment to such an extent that the data had to be divided into separate between-subject groups (Section 2.3). While the cause of this order effect could not be determined, it was clear that it was highly significant and did not simply reflect overall improvements due to practice.

Observers with prior real experience with a particular task performed better in a virtual representation, while observers with previous experience in a virtual environment performed worse in a real-world environment than observers without such experience. Thus, it is critical to consider prior experience, especially if VR is used to train observers on a particular task. Observers that are only exposed to a virtual simulation first may need a period of adjustment to match real-world performance. It is critical to consider the observer's prior experience with the task in real-world environments and how the experience gained in the virtual space could impact performance.

The second scenario in which depth judgements of physical objects demonstrated consistent distortions occurred in Chapter 3 when observers estimated the size and depth of truncated square pyramids. Even in the full-cue physical test environment, observers underestimated the perceived size of the front surface (Figure 3.3), and the depth of the pyramids (Figure 3.6). Despite the similarities of depth judgements of virtual stimuli in Chapter 2 and 3, the depth of physical objects were underestimated to the same extent as virtual objects in Chapter 3, but not in Chapter 2.

One possibility is the different stimuli used in Chapter 2 and 3 contributed to differences between virtual and physical environments. Chapter 3 used truncated square pyramids (Figure 3.1), which consisted of linear horizontal and vertical gradients of horizontal disparity and texture along the sloped surfaces. Whereas Chapter 2 used texture half-cylinders (Figure 2.1), which contained curved texture and disparity gradients (i.e., second-order spatial derivatives of surface curvature) that provided additional information regarding the relative depth and slant of local surface curvature (Lappin & Craft, 2000; Norman et al., 1991). However, stimulus differences would not explain the improvement in depth judgements of physical stimuli in Chapter 2. That is, given that second-order disparities are not scaled by viewing distance if they are presented in the median plane (Rogers & Cagenello, 1989), it is unlikely that they contributed to the improvement in depth scaling in physical relative to virtual environments in Chapter 2.

A more likely explanation for the differential effect of viewing environment is *how* depth was estimated in each study. As mentioned in Section 2.1, *how* depth is estimated is an important consideration when comparing the accuracy of depth judgements across conditions (Glennerster et al., 1996). Several methods can be used to measure 'how much' depth an observer perceives, such as manual-pointing tasks (Foley et al., 1975), depth interval bisection tasks (Ogle, 1952b; 1953), ruler adjustment (Tsirlin et al., 2012), haptic matching tasks (Brenner & Van Damme, 1999; Hornsey et al., 2020; Wallach & Zuckerman, 1963), or generative haptic methods (Hartle & Wilcox, 2016). While perceptual estimation methods such as depth matching tasks are limited to equating one perceptual

magnitude to another (Foley et al., 1975), they often demonstrate accurate depth scaling even in sparse real-world viewing environments (Glennerster et al., 1996; Ritter, 1977), while manual tasks tend to show systematic depth distortions (Foley et al., 1975).

Such methodological effects may account for the difference between Chapters 2 and 3. That is, observers relied on a generative haptic method to estimate perceived depth magnitude in Chapter 2, while a discrimination task compared the perceived width of a frontal surface to perceived depth in Chapter 3. A major difference between these studies is the fact that the reference for depth judgements varied in Chapter 3, but not in Chapter 2. As shown in Figure 3.3, the perceived width of the frontal surface was significantly smaller for virtual relative to physical objects. Thus, the size of the width reference changed substantially between viewing environments. When this difference in perceived width was accounted for, the relative reduction in perceived depth was similar in the two environments. However, in Chapter 2, observers always used the distance between their thumb and forefinger to indicate their depth judgements in all test conditions. The mapping of their judgement to the motor response of their fingers remained consistent between viewing environments. Thus, the manual-estimation task was able to capture the differences in perceived depth between the virtual and physical test environments. However, while the different measurement method used in Chapter 2 and 3 may explain the difference between depth judgements of physical stimuli across the studies, it cannot entirely account for the underestimation of size and depth of physical objects seen in Chapter 3. This may be due to other factors, such as interobserver variability (Mon-Williams et al., 2000), or differences in task demands (Glennerster et al., 1996; Scarfe & Hibbard, 2006; Todd, 2004).

The comparison of depth judgements between Chapter 2 and 3 highlight the importance of considering the sources of variability, including interobserver, methodological, and stimulus differences, when comparing perceptual judgments between virtual and physical viewing environments. While real-world environments provide an important form of validation, it is important to recognize that perception in real-world environments is not necessarily veridical. To generalize behaviour on a particular task in virtual environments to performance in the real world, the focus should be on the relative comparison between virtual and physical test environments, rather than absolute accuracy. However, if the aim is to understand absolute accuracy, then performance in the physical test environment could be treated as ceiling with the understanding that even under these conditions, perception may not correspond perfectly to the physical structure of the scene. Studies must optimize control through well-designed experimental protocols and a focused context-specific research question to take advantage of real-world environments as a validation tool.

5.4 Do we perceive metric depth accurately from visual information alone?

If depth distortions exist in both virtual and physical environments, is it appropriate to assume that the human visual system maintains an accurate representation of metric depth? As discussed in Section 1.2, distortions of perceived depth typically consist of overestimation of depth in near space and underestimation of depth at large viewing distances (Foley, 1980). Even at viewing distances less than 50cm, where distance information from vergence, vertical disparities, and accommodation are most reliable (Backus et al., 1999; Mon-Williams et al., 2000; Rogers & Bradshaw, 1995), depth from binocular disparity still tends to be overestimated (Foley, 1980; Gogel, 1977; Norman et al., 1996). Likewise, while binocular disparity supports depth scaling at viewing distances well beyond the useful range of these distance cues (e.g., 4, 8, and 18m), depth from binocular disparity is greatly underestimated at large viewing distances, even in real environments by upwards of 75% (Allison et al., 2009; Palmisano et al., 2010). There are numerous examples of studies that report these depth distortions from stereopsis in virtual stimuli (Bradshaw et al., 1996; Brenner & Landy, 1999; Brenner & Van Damme, 1999; Glennerster et al., 1996; Johnston, 1991; Johnston et al., 1994; Todd & Norman, 2003; Willemsen et al., 2008; Witmer & Kline, 1998). In the current series of studies, virtual assessments confirm these typical patterns of underconstancy, where objects presented at near viewing distances of 50cm were overestimated (Figure 4.3), and objects presented at farther viewing distances of 83 and 130cm were underestimated (Figure 2.5 and 3.6). However, psychophysical assessments of depth perception using physical objects are mixed. Some report that depth distortions in physical environments are similar or less extreme than those observed with virtual stimuli (Bradshaw et al., 2000; Cuijpers et al., 2000; Loomis & Philbeck, 1999; Tittle et al., 1995). While others, including the assessment in Chapter 2, demonstrate that stereoscopic depth constancy and accuracy is attained in physical environments , even in relatively sparse viewing environments (Durgin et al., 1995; Frisby et al., 1996; Glennerster et al., 1996; Hartle & Wilcox, 2016; Ritter, 1977; Willemsen et al., 2008).

There are at least two interpretations of the above findings driving the debate of whether the visual system, represents metric depth information accurately. The first interpretation is the mixed depth estimation results do not necessarily suggest that the visual system fails to recover the metric depth of objects. Instead, the observed depth distortions could be the product of sparse, cue limited laboratory environments, or, in the case of virtual objects, the result of the limitations of computerized display systems, such as the presence of cues to flatness (Young et al., 1993), or conflicts between oculomotor distance cues (Hoffman et al., 2008; Watt et al., 2005). This position is supported by assessments of

physical stimuli without such conflicts that demonstrate near-accurate depth constancy (Durgin et al., 1995; Frisby et al., 1996; A Glennerster et al., 1996; Ritter, 1977; Willemsen et al., 2008). The second interpretation argues that the existence of perceptual distortions, that can persist under full-cue conditions (Bradshaw et al., 2000; Cuijpers et al., 2000; Loomis & Philbeck, 1999; Tittle et al., 1995), is evidence that the visual system does not have a metric representation of a scene. In this case, the failure of accurate metric estimation should be considered an important finding that needs to be addressed (Glennerster et al., 1996; Hecht et al., 1999; Tittle et al., 1995; Todd, 2004; Todd & Norman, 2003). Section 1.3.4 discussed how cue combination models could address these failures by integrating biased estimates using modified linear or alternative models of combination. However, discussion of whether the visual system integrates metric or non-metric information is beyond the scope of this series of dissertation studies.

It may be the case that the visual system does represent metric depth information accurately, but only when there is an abundance of consistent and reliable information regarding the depth and distance of objects; no single depth cue provides an accurate sense of depth. According to this view, only complex viewing environments with a plethora of rich, redundant, and consistent sources of depth and distance information are sufficient for the visual system to achieve accurate depth perception (Bradshaw et al., 2000; Landy et al., 1995). In Chapter 2, while the depth of virtual objects was consistently underestimated, matched physical objects presented at the same viewing distances (83 and 130cm) were perceived accurately. It has also been argued that the visual system only requires a complete and accurate metric representation of space within the range of distances where object interactions take place. For example it is possible that stereoscopic depth perception of object location is fine-tuned to be optimal at a natural grasping distance (Richards, 2009; Volcic et al., 2013). The range of distances in which objects are commonly manipulated, approximately 50 to 70cm, are also the distances where depth and distance cues are quite reliable (Backus et al., 1999; Mon-Williams et al., 2000; Rogers & Bradshaw, 1995). While the current series of studies did not evaluate perceived depth within viewing distances of 50 to 70cm, in the studies reported here, depth judgements of virtual objects presented at 50cm were shown to be overestimated (Figure 4.3) and judgements at 83cm were shown to be underestimated (Figure 2.5). However, in full-cue physical test environments the depth of objects can be perceived accurately beyond reaching distances, often to at least 2m (Foley, 1980; Ono & Comerford, 1977).

Thus, the visual system may generate accurate metric depth percepts, but only under certain viewing conditions. In sparse, cue limited environments the visual system provides accurate metric information over the critical range of distances necessary for basic object interactions. Whereas in natural

viewing environments with abundant sources of information an accurate metric representation is available over a larger range of viewing distances.

5.5 What about other non-visual cues to distance?

During our everyday behaviours, we have few difficulties perceiving the distances to, or the depth of objects. One important advantage of real-world objects could be our ability to interact with them through reaching and grasping actions (Berkeley, 1709; Held & Hein, 1963; Welch, 1969). It is possible that other non-visual cues to distance, such as proprioceptive information from interacting with objects, also contribute to our sense of space. Chapter 4 proposes that given stereopsis improves the accuracy of reaching and grasping actions (Jackson et al., 1997; Marotta et al., 1995; Marotta & Goodale, 1998; Servos et al., 1992; Servos & Goodale, 1994), the act of reaching to objects may also provide additional cues to distance that, in turn, aid the scaling of binocular disparities. Reaching to objects in a virtual environment with known distance distortions is akin to reaching to targets with misaligned visual feedback. While these limitations of virtual environments contribute to distortions of perceived distance and depth, we show in Chapter 4 that if observers have an opportunity to reach to virtual objects, then the accuracy of subsequent depth judgements improve. However, it was apparent that the contribution of reaching-in-depth in a sparse virtual environment was quite weak (Figure 4.7). Only the inclusion of strong distance cues that are often available in natural reaching and grasping in conjunction with distance cues from reaching in depth further improved the scaling of subsequent depth judgements (Figure 4.3). Thus, while reaching-in-depth improves the scaling of subsequent stereoscopic depth judgements, this does not completely account for accurate depth judgements in natural viewing environments as the presence of other consistent distance information is necessary to improve the scaling of depth judgements. Despite the presence of conflicts between oculomotor distance cues in virtual environments, non-visual modalities compensate for unreliable distance information and allow observers to improve their distance perception in virtual environments with many supportive distance cues. Observers could benefit from an interactive phase where they have an opportunity to reach to virtual objects at various distances to gain additional information regarding the position of multiple objects in the scene and aid the scaling of the virtual space.

5.6 Future Directions

5.6.1 Combination of Motion Parallax and Binocular Disparity under Less Reliable Conditions

While the study of depth integration from binocular disparity and motion parallax in Chapter 3 provided a starting point, further work is necessary to understand the complexities of this depth cue integration. The results showed that observers did not benefit from the combination of binocular disparity and motion parallax in virtual or physical viewing environments. However, under the tested viewing conditions, depth estimates from binocular disparity were much more precise than motion parallax (Figure 3.7). Thus, motion parallax may have been at a disadvantage relative to binocular disparity. It would be beneficial to determine if the same vetoing method would be observed if the reliability of depth from binocular disparity was degraded. If the reliability of depth from binocular disparity was less than motion parallax, observers may rely more heavily on motion parallax cues. To test this hypothesis, in a follow-up study we used a Bangerter occlusion foil to decrease the visual acuity in one eye to 20/300 in order to reduce the reliability of depth from binocular disparity. Observers reported that the blurred image was close to detection threshold. As a starting point, observers were only tested in the virtual test condition rendered using the same Oculus Rift CV1 HMD, truncated pyramid stimuli, and procedure as Chapter 3. Seven observers were recruited from York University and passed the required stereoacuity screening.

Figure 5.1 shows the average PSEs and JNDs for the single (binocular disparity, motion parallax) and combined cue conditions for the virtual pyramids. The statistical analysis is summarized in a table in Appendix 5.A. This analysis revealed that the PSE for the combined condition was not significantly different from the binocular disparity or motion parallax condition. However, the PSE for the motion parallax condition was significantly lower than the binocular disparity condition. Unfortunately, the JNDs for the binocular disparity condition were not significantly different relative to the combined or motion parallax conditions. Nor was the difference between the combined and motion parallax conditions. While it is apparent that the occlusion filters did reduce the precision of depth from binocular disparity relative to the previous study (Figure 3.7), the JND was not significantly less than the motion parallax condition. Observers reported that the depth discrimination task was extremely difficult when only binocular disparity was present. Despite depth estimates from motion parallax being slightly more accurate relative to those resulting from binocular disparity alone, there was no difference in accuracy between the combined condition and either single cue condition. The same cue combination analysis used in Chapter 3 revealed a similar result in that there was no evidence that observers combined binocular disparity or motion parallax according to a linear cue integration model.

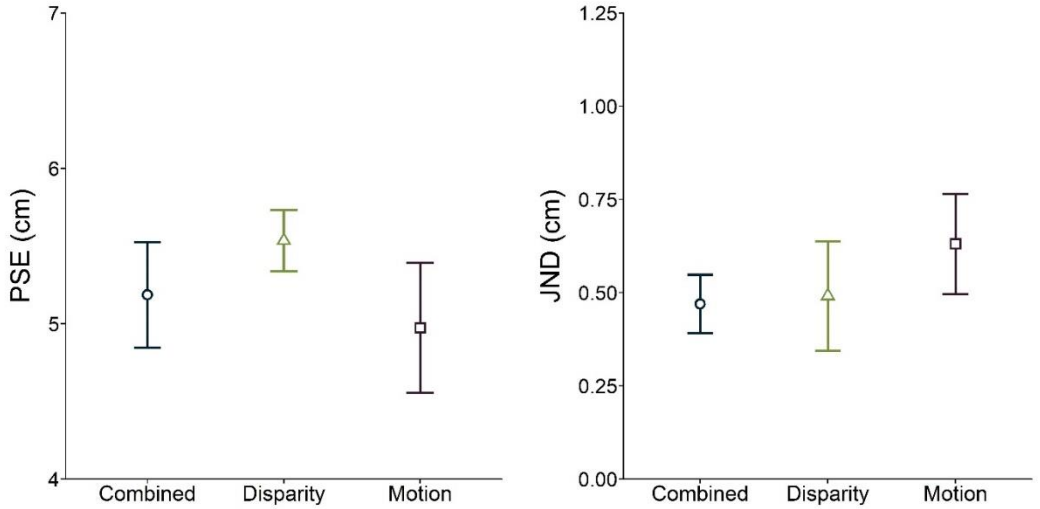


Figure 5.1. Average PSEs (left) and JNDs (right) for the follow-up study using occlusion foils. The averages (n=7) are shown for each of the three cue conditions: binocular disparity only (green triangles), motion parallax (purple squares), and their combination (blue circles). Error bars represent the standard error of the mean.

This follow-up study demonstrated that is quite challenging to reduce the reliability of binocular disparity to a level poorer than that of motion parallax. The absence of evidence of linear cue combination, even under unreliable conditions suggests that either (1) a linear combination model is not an appropriate model for the integration of depth from motion parallax and binocular disparity, or (2) motion parallax may not significantly contribute to depth magnitude estimation in the presence of stereopsis. As discussed in Section 1.3.4, if depth information from individual cues is inherently biased, then the visual system may adopt an alternative cue combination method. While the model tested in Chapter 3 accounts for underestimations in depth judgements using a flatness prior, other more complex models, such as biased integration methods (Peter Scarfe & Hibbard, 2011), may better account for the results. The other possibility is that perhaps the relative depth information providing by motion parallax is relatively weak compared to binocular disparity (Durgin et al., 1995; McKee & Taylor, 2010). That is, because relative depth from motion parallax alone depends on additional information about eye, head, and body position (Helmholtz, 1925; Howard & Rogers, 2012), there could be less precision in depth estimates from motion parallax than binocular disparity. Thus, depth magnitude may not be the right medium to compare the contribution of binocular disparity and motion parallax to depth perception. For instance, motion parallax may play a larger role in depth segmentation than depth magnitude (Yoonessi & Baker, 2011), especially given only a small amount of motion (as little as two frames) is sufficient to segment surfaces (Baker & Braddick, 1982; Nakayama et al., 1985).

5.6.2 Reaching in Simple and Complex Virtual Scenes

One outstanding question from the results of Chapter 4 is whether the benefit of reaching in depth is due to an overall improvement in the accuracy of perceived distance or the result of calibrating depth judgements to a specific reaching distance. In Chapter 4, Experiment 1, reaching to a target at a viewing distance of 50cm, reduced the inferred viewing distance (i.e., slope) of depth judgements so they were consistent with both reaching and viewing distance of the target. However, this would occur whether observers calibrated their depth judgements to the specific reaching distance, or if the accuracy of their absolute distance estimate improved. These two possibilities predict disparate outcomes if the magnitude of the reaching distance and viewing distance to the object are different. For instance, if the reaching distance was further than the 50cm viewing distance to stimulus, two outcomes are possible: (1) if observers calibrate to the reaching distance, then inferred viewing distance would increase, or (2) if the accuracy of the observers' estimate of absolute distance is increased, then we would predict a decrease in inferred viewing distance, similar to the results of Experiment 1.

Further, given that the influence of reaching-in-depth in isolation was weak, and depended on the availability of other distance cues (e.g., height in the field), then varying the number of available cues in the scene may modulate the impact of reaching-in-depth. For instance, by removing specific elements from the virtual scene, the quantity and quality of depth cues can be controlled between reach conditions. Examples of these possible cue conditions are shown in Figure 5.2. In the limited cue condition, by removing the table, the information regarding the distance of the target peg is limited to binocular disparity and vergence cues. This condition would provide a baseline for the influence of reaching-in-depth without the support of additional distance cues. The second, Table Only condition would add the wooden table that was present in the original ring task in Chapter 4 without any objects on the table. This would add texture and perspective cues from the table, as well as relative disparity between the target peg and the surface of the table. Lastly, the cluttered condition would add simple 3D geometric objects to the tabletop. These objects would add distance information from relative height in the field and relative size cues, without the influence of familiar size cues. In all three conditions, the ring would only be tracked to the observer's fingertip (i.e., rings would not appear on the table), and the observer would reset their hand position between trials to control the pattern of reaching movements between cue conditions. The comparison of stereoscopic depth judgements from before and after each of these cue conditions could provide insight into what information is critical to aid depth scaling when reaching to objects in depth.

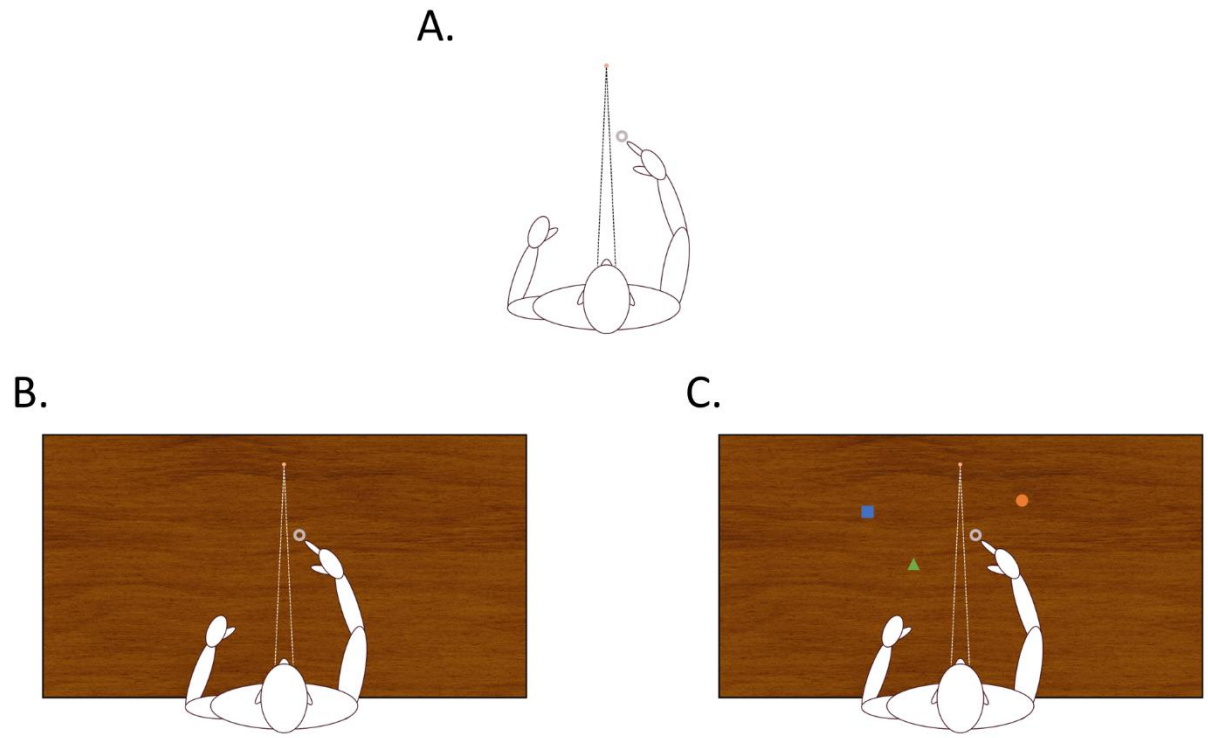


Figure 5.2. *An illustration of potential cue conditions in a follow-up reaching-in-depth study. Scene A represents a cue limited scenario with only binocular disparity and vergence information to indicate the distance of the target peg. Scene B represents a scenario with only texture and perspective cues from the table present. Scene C represents a cluttered scene where additional objects are placed on the table to provide cues to height in the field, relative disparities, and perspective cues.*

5.7 Conclusions

When assessing depth perception, it is important to consider not only the presence of various depth cues in the scene, but the context in which that information is presented, the observer's prior experience, and how that relates to natural viewing environments. It is well established that research on 3D perception is best carried out using naturalistic objects rather than impoverished stimuli on computerized displays. One must be cautious when limiting cues as it only tells us about how the visual systems manages with that particular cue, which may not translate to the real world. However, this series of studies has shown that despite inherent cue conflicts in virtual environments that distort perceived depth, other factors compensate for these issues. Exposure to a similar physical environment prior to the virtual environment allows the user to improve their depth and distance perception of virtual objects. Reaching to objects in a virtual environment provides additional distance cues that aid subsequent depth judgements from binocular disparity. Thus, if tasks performed in virtual environments require accurate

depth and distance perception, then users' performance will benefit from additional practice reaching to virtual objects or brief exposure to a similar physical environment.

REFERENCE LIST

Adelstein, B. D., Lee, T. G., & Ellis, S. R. (2003). Head Tracking Latency in Virtual Environments: Psychophysics and a Model. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(20). https://doi.org/10.1177/154193120304702001

Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, *23*(3). https://doi.org/10.1145/1015706.1015804

Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, *9*(1), 1–14. https://doi.org/10.1167/9.1.10

Allison, R. S., & Howard, I. P. (2000). Temporal dependencies in resolving monocular and binocular cue conflict in slant perception. *Vision Research*, *40*(14), 1869–1885.

Backus, B. T., Banks, M. S., Van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*(6), 1143–1170. https://doi.org/10.1016/S0042-6989(98)00139-4

Baird, J. (1903). *The influence of accommodation and convergence upon the perception of depth*. *14*(2), 150–200.

Baker, C. L., & Braddick, O. J. (1982). Does segregation of differently moving areas depend on relative or absolute displacement? *Vision Research*, *22*, 851–856.

Berkeley, G. (1709). An essay toward a new theory of vision. In *Works on vision* (C.M. Turba). Bobbs-Merrill.

Bingham, G. P., Coats, R., & Mon-Williams, M. (2007). Natural prehension in trials without haptic feedback but only when calibration is allowed. *Neuropsychologia*, *45*(2), 288–294.

Blake, A., Bülthoff, H. H., & Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*, *33*(12), 1723–1737. https://doi.org/10.1016/0042-6989(93)90037-W

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *The Journal of Physiology*, *211*(3), 599–622. https://doi.org/10.1113/jphysiol.1970.sp009296

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). Converting among effect sizes. In *Introduction to meta-analysis*. Wiley.

Bozzacchi, C., & Domini, F. (2015). Lack of depth constancy for grasping movements in both virtual and real environments. *Journal of Neurophysiology*, *114*(4), 2242–2248. https://doi.org/10.1152/jn.00350.2015

Bozzacchi, C., Volcic, R., & Domini, F. (2014). Effect of visual and haptic feedback on grasping movements. *Journal of Neurophysiology*, *112*(12), 3189–3196.

Bozzacchi, C., Volcic, R., & Domini, F. (2016). Grasping in absence of feedback: systematic biases endure extensive training. *Experimental Brain Research*, *234*(1), 255–265. https://doi.org/10.1007/s00221-015-4456-9

Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, *36*(9), 1255–1264. https://doi.org/10.1016/0042-6989(95)00190-5

Bradshaw, M F, Parton, A. D., & Eagle, R. A. (1998). The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. *Perception*, *27*(11), 1317–1331.

Bradshaw, M F, & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research*, *36*(21), 3457–3468.

Bradshaw, Mark F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, *40*(27), 3725–3734. https://doi.org/10.1016/S0042-6989(00)00214-5

Bradshaw, Mark F., & Rogers, B. J. (1999). Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research*, *39*(18). https://doi.org/10.1016/S0042-6989(99)00015-2

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4). https://doi.org/10.1163/156856897X00357

Brenner, E., & van Damme, W. J. (1998). Judging distance from ocular convergence. *Vision Research*, *38*(4), 493–498. https://doi.org/10.1016/S0042-6989(97)00236-8

Brenner, E, & Landy, M. S. (1999). Interaction between the percieved shape of two objects. *Vision Res*, *39*, 3843–3848.

Brenner, Eli, & Van Damme, W. J. M. (1999). Perceived distance, shape and size. *Vision Research*, *39*(5), 975–986. https://doi.org/10.1016/S0042-6989(98)00162-X

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, *33*(7), 919–933. https://doi.org/10.1016/0042-6989(93)90075-8

Campagnoli, C., Croom, S., & Domini, F. (2017). Stereovision for action reflects our perceptual experience of distance and depth. *Journal of Vision*, *17*(9), 1–26. https://doi.org/10.1167/17.9.21

Campbell, F. W. (1957). The depth of field of the human eye. *Optica Acta: International Journal of Optics*, *4*(4), 157–164. https://doi.org/10.1080/713826091

Caudek, C., Fantoni, C., & Domini, F. (2011). Bayesian modeling of perceived surface slant from actively-generated and passively-observed optic flow. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0018731

Ciuffreda, K. J. (1998). Accommodation, the pupil, and presbyopia. In *Borish's clinical refraction* (pp. 93–144).

Coats, R., Bingham, G. P., & Mon-Williams, M. (2008). Calibrating grasp size and reach distance: Interactions reveal integral organization of reaching-to-grasp movements. *Experimental Brain Research*, *189*(2), 211–220.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, *4*(1), 102–118. https://doi.org/10.2307/2984123

Collett, T. S., Schwarz, U., & Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, *20*(6), 733–754. https://doi.org/10.1068/p200733

Cormack, L. K., Landers, D. D., & Ramakrishnan, S. (1997). Element density and the efficiency of binocular matching. *Journal of the Optical Society of America A*, *14*(4), 723–730. https://doi.org/10.1364/josaa.14.000723

Cornilleau-Pérès, V., & Droulez, J. (1994). The visual perception of three-dimensional shape from self-motion and object-motion. *Vision Research*, *34*(18). https://doi.org/10.1016/0042-6989(94)90279-8

Coutant, B. E., & Westheimer, G. (1993). Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics*, *13*(1), 3–7. https://doi.org/10.1111/j.1475-1313.1993.tb00419.x

Cuijpers, R. H., Kappers, A. M. L., & Koenderink, J. J. (2000). Investigation of visual space using an exocentric pointing task. *Perception and Psychophysics*, *62*, 1556–1571.

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, *33*(5–6), 827–838. https://doi.org/10.1016/0042-6989(93)90201-7

Cutone, M. D., Allison, R. S., & Wilcox, L. M. (2019). The impact of retinal motion on stereoacuity for physical targets. *Vision Research*, *161*, 43–51. https://doi.org/10.1016/j.visres.2019.06.003

Cutone, M. D., & Wilcox, L. M. (2018). *PsychXR* (Version 0.1.4). https://github.com/mdcutone/psychxr

Domini, F., & Caudek, C. (2010). Matching perceived depth from disparity and from velocity: Modelling and psychophysics. *Acta Psychologica*, *133*(1), 81–89.

Domini, Fulvio, & Caudek, C. (2009). The intrinsic constraint model and Fechnerian sensory scaling. *Journal of Vision*, *9*(2), 1–15. https://doi.org/10.1167/9.2.25

Domini, Fulvio, Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision Research*, *46*(11), 1707–1723. https://doi.org/10.1016/j.visres.2005.11.018

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, *26*(6), 973–990. https://doi.org/10.1016/0042-6989(86)90154-9

Duane, A. (1917). *Fuchs's text-book of ophthalmology* (5th ed.). Lippincott.

Dunn, B. E., Gray, G. C., & Thompson, D. (1965). Relative height on the picture-plane and depth perception. *Perceptual and Motor Skills*, *21*(1), 227–236.

Durgin, F. H., Proffitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing Depth From Motion With Depth From Binocular Disparity. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 679–699. https://doi.org/10.1037/0096-1523.21.3.679

Eadie, A. S., Gray, L. S., Carlin, P., & Mon-Williams, M. (2000). Modelling adaptation effects in vergence and accommodation after exposure to a simulated virtual reality stimulus. *Ophthalmic and Physiological Optics*, *20*(3), 242–251. https://doi.org/10.1016/S0275-5408(99)00057-5

Ellis, S. R., Young, M. J., Adelstein, B. D., & Ehrlich, S. M. (1999). Discrimination of Changes of Latency during Voluntary Hand Movement of Virtual Objects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(22). https://doi.org/10.1177/154193129904302203

Epstein, W. (1966). Perceived depth as a function of relative height under three background conditions. *Journal of Experimental Psychology*, *72*(3), 335–338.

Erkelens, C. J. (2000). Perceived direction during monocular viewing is based on signals of the viewing eye only. *Vision Research*, *40*(18), 2411–2419. https://doi.org/10.1016/S0042-6989(00)00120-6

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Feldstein, I. T., Kölsch, F. M., & Konrad, R. (2020). Egocentric distance perception: A comparative study investigating differences between real and virtual environments. *Perception*, *49*(9), 940–967. https://doi.org/10.1177/0301006620951997

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE.

Foley, J. M. (1967). Binocular disparity and perceived relative distance: An examination of two hypotheses. *Vision Research*, *7*(7–8), 655–670. https://doi.org/10.1016/0042-6989(67)90073-9

Foley, J. M. (1968). Depth, size and distance in stereoscopic vision. *Perception & Psychophysics*, *3*(4), 265–274. https://doi.org/10.3758/BF03212742

Foley, J. M. (1977). Effect of distance information and range on two indices of visually perceived distance. *Perception*, *6*(4), 449–460. https://doi.org/10.1068/p060449

Foley, J. M., Applebaum, T. H., & Richards, W. A. (1975). Stereopsis with large disparities: Discrimination and depth magnitude. *Vision Research*, *15*(3), 417–421. https://doi.org/10.1016/0042-6989(75)90091-7

Foley, J. M., & Held, R. (1972). Visually directed pointing as a function of target distance, direction, and available cues. *Perception & Psychophysics*, *12*(3), 263–268. https://doi.org/10.3758/BF03207201

Foley, J. M., & Richards, W. (1972). Effects of voluntary eye movement and convergence on the binocular appreciation of depth. *Perception & Psychophysics*, *11*(6), 423–427.

Foley, John M. (1980). Binocular Distance Perception. *Psychological Review*, *87*(5), 411–434. https://doi.org/10.1037/h0021465

Foley, John M. (1985). Binocular Distance Perception. Egocentric Distance Tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(2), 133–149. https://doi.org/10.1037/0096-1523.11.2.133

Frisby, J. P., Buckley, D., & Duke, P. A. (1996). Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *Perception*, *25*(2), 129–154.

Frisby, J. P., Buckley, D., & Horsman, J. M. (1995). Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception*, *24*(2), 181–198. https://doi.org/10.1068/p240181

Fry, G. A. (1939). Further experiments on the accomodative convergence relationship. *Optometry and Vision Science*, *16*(9), 325–336.

Fulvio, J. M., & Rokers, B. (2017). Use of cues in virtual reality depends on visual feedback. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-16161-3

Garding, J. P., Porrill, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, *35*(5), 703–722. https://doi.org/10.1016/0042-6989(94)00162-F

Gardner, P. L., & Mon-Williams, M. (2001). Vertical gaze angle: Absolute height-in-scene information for the programming of prehension. *Experimental Brain Research*, *136*(3), 379–385.

Gibaldi, A., & Banks, M. S. (2019). Binocular eye movements are adapted to the natural environment. *Journal of Neuroscience*, *39*(15), 2877–2888. https://doi.org/10.1523/JNEUROSCI.2591-18.2018

Gillam, B., Palmisano, S. A., & Govan, D. G. (2011). Depth interval estimates from motion parallax and binocular disparity beyond interaction space. *Perception*, *40*(1), 39–49. https://doi.org/10.1068/p6868

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision*, *9*(9). https://doi.org/10.1167/9.9.8

Glennerster, A, Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic Depth Constancy Depends on the Subject 's Task. *Vision Research*, *36*(21), 3441–3456.

Glennerster, Andrew, Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception*, *27*(11), 1357–1365. https://doi.org/10.1068/p271357

Gogel, W C. (1961). Convergence as a cue to the perceived distance of objects in a binocular configuration. *The Journal of Psychology*, *52*(2), 303–315.

Gogel, Walter C. (1977). An indirect measure of perceived distance from oculomotor cues. *Perception & Psychophysics*, *21*(1). https://doi.org/10.3758/BF03199459

Gogel, Walter C., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics*, *13*(2). https://doi.org/10.3758/BF03214141

Gogel, Walter C., & Tietz, J. D. (1979). A comparison of oculomotor and motion parallax cues of egocentric distance. *Vision Research*, *19*(10), 1161–1170. https://doi.org/10.1016/0042-6989(79)90013-0

Hartle, B., & Wilcox, L. M. (2016). Depth magnitude from stereopsis: Assessment techniques and the role of experience. *Vision Research*, *125*, 64–75. https://doi.org/10.1016/j.visres.2016.05.006

Hartle, B., & Wilcox, L. M. (2021). Cue vetoing in depth estimation: Physical and virtual stimuli. *Vision Research*, *188*, 51–64.

Hartle, B., & Wilcox, L. M. (2022). Stereoscopic depth constancy for physical objects and their virtual counterparts. *Journal of Vision*, *22*(4), 9–9.

Hecht, H., van Doorn, A., & Koenderink, J. J. (1999). Compression of visual space in natural scenes and in their photographic counter-parts. *Perception and Psychophysics*, *61*, 1269–1286.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, *56*(5), 872–876.

Helmholtz, H. (1909). Physiological optics, vol. 3, trans. *JPC Southall. Optical Society of America*.

Helmholtz, H. (1925). *Helmholtz's treatise on physiological optics* (Vol. 3; JP). Optical Society of America.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 967–992. https://doi.org/10.1167/4.12.1

Hoffman, D. M., & Banks, M. S. (2010). Focus information is used to interpret binocular images. *Journal of Vision*, *10*(5), 1–17. https://doi.org/10.1167/10.5.13

Hoffman, David M, Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence – accommodation conflicts

hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), 1–30. https://doi.org/10.1167/8.3.33.Introduction

Holleman, G. A., Hooge, I. T., Kemner, C., & Hessels, R. S. (2020). The "real-world approach" and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 721.

Hornsey, R. L., Hibbard, P. B., & Scarfe, P. (2020). Size and shape constancy in consumer virtual reality. *Behavior Research Methods*, *52*(4), 1587–1598. https://doi.org/10.3758/s13428-019-01336-9

Howard, I. P., & Rogers, B. J. (2012). Perceiving in Depth. In *Perceiving in Depth* (Vol. 2). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199764150.001.0001

Interrante, V., Anderson, L., & Ries, B. (2006). Distance perception in immersive virtual environments, revisited. *IEEE Virtual Reality*, 3–10. https://doi.org/10.1145/1620993.1620996

Jackson, S. R., Jones, C. A., Newport, R., & Pritchard, C. (1997). A kinematic analysis of goal-directed prehension movements executed under binocular, monocular, and memory-guided viewing conditions. *Visual Cognition*, *4*(2), 133–142.

Jakobson, L. S., & Goodale, M. A. (1991). Factors affecting higher-order movement planning: A kinematic analysis of human prehension. *Experimental Brain Research*, *86*(1), 199–208.

Jay, C., Glencross, M., & Hubbold, R. (2007). Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment. *ACM Transactions on Computer-Human Interaction*, *14*(2). https://doi.org/10.1145/1275511.1275514

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*(7–8), 1351–1360. https://doi.org/10.1016/0042-6989(91)90056-B

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, *34*(17), 2259–2275. https://doi.org/10.1016/0042-6989(94)90106-6

Julesz, B. (1971). *Foundations of cyclopean perception*. University of Chicago Press.

Keefe, B. D., & Watt, S. J. (2009). The role of binocular vision in grasping: A small stimulus-set distorts results. *Experimental Brain Research*, *194*(3), 435–444. https://doi.org/10.1007/s00221-009-1718-4

Kerrigan, I. S., & Adams, W. J. (2013). Learning different light prior distributions for different contexts. *Cognition*, *127*(1), 99–104. https://doi.org/10.1016/j.cognition.2012.12.011

Knill, D. C. (1998). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research*, *38*(11), 1683–1711. https://doi.org/10.1016/S0042-6989(97)00325-8

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, *7*(7), 1–24. https://doi.org/10.1167/7.7.5

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. https://doi.org/10.1016/S0042-6989(03)00458-9

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, *8*(2), 377–385.

Komoda, M. K., & Ono, H. (1974). Oculomotor adjustments and size-distance perception. *Perception & Psychophysics*, *15*(2), 353–360. https://doi.org/10.3758/BF03213958

Krakauer, J. W., Pine, Z. M., Ghilardi, M. F., & Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, *20*(23), 8916–8924. https://doi.org/10.1523/jneurosci.20-23-08916.2000

Landy, M. S., & Brenner, E. (2001). Motion-Disparity Interaction and the Scaling of Stereoscopic Disparity. *Vision and Attention*, 131–152. https://doi.org/10.1007/978-0-387-21591-4_7

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*(3). https://doi.org/10.1016/0042-6989(94)00176-M

Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: From retinal images to perceived shapes. *Psychological Review*, *107*(1), 6–38.

Linton, P. (2020). Does vision extract absolute distance from vergence? *Attention, Perception, and Psychophysics*, *82*(6), 3176–3195.
Loomis, J. M., & Philbeck, J. W. (1999). Is the anisotrophy of perceived 3-D shape invariant across scale? *Perception and Psychophysics*, *61*, 397–402.
Maloney, L. T., & Landy, M. S. (1989). A Statistical Framework for Robust Fusion of Depth Information. *Visual Communications and Image Processing IV*, *1199*. https://doi.org/10.1117/12.970125
Marotta, J. J., & Goodale, M. A. (1998). The role of learned pictorial cues in the programming and control of grasping. *Experimental Brain Research*, *121*(4), 465–470. https://doi.org/10.1007/s002210050482
Marotta, J. J., Perrot, T. S., Nicolle, D., Servos, P., & Goodale, M. A. (1995). Adapting to monocular vision: Grasping with one eye. *Experimental Brain Research*, *104*(1), 107–114.
McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research*, *23*(2), 191–198. https://doi.org/10.1016/0042-6989(83)90142-6
McKee, S. P., & Taylor, D. G. (2010). The precision of binocular and monocular depth judgments in natural settings. *Journal of Vision*, *10*(10), 5–5. https://doi.org/10.1167/10.10.5
Melmoth, D. R., Storoni, M., Todd, G., Finlay, A. L., & Grant, S. (2007). Dissociation between vergence and binocular disparity cues in the control of prehension. *Experimental Brain Research*, *183*(3), 283–298. https://doi.org/10.1007/s00221-007-1041-x
Mon-Williams, M., & Tresilian, J. R. (2000). Ordinal depth information from accommodation? *Ergonomics*, *43*(3), 391–404. https://doi.org/10.1080/001401300184486
Mon-Williams, M., Warm, J. P., & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics*, *13*(4), 387–391. https://doi.org/10.1111/j.1475-1313.1993.tb00496.x
Mon-Williams, Mark, & Dijkerman, H. C. (1999). The use of vergence information in the programming of prehension. *Experimental Brain Research*, *128*(4), 578–582. https://doi.org/10.1007/s002210050885
Mon-Williams, Mark, Tresilian, J. R., & Roberts, A. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research*, *133*(3), 407–413. https://doi.org/10.1007/s002210000410
Murdoch, J. R., McGhee, C. N. G., & Glover, V. (1991). The relationship between stereopsis and fine manual dexterity: Pilot study of a new instrument. *Eye (Basingstoke)*, *5*(5), 642–643. https://doi.org/10.1038/eye.1991.112
Nakayama, K., Silverman, G. H., Macleod, D. I. A., & Mulligan, J. (1985). Sensitivity to shearing and compressive motion in random dots. *Perception*, *14*, 225–238.
Nawrot, M., Ratzlaff, M., Leonard, Z., & Stroyan, K. (2014). Modeling depth from motion parallax with the motion/pursuit ratio. *Frontiers in Psychology*, *5*, 1–14. https://doi.org/10.3389/fpsyg.2014.01103
Newman, E. B. (1933). The validity of the just noticeable difference as a unit of psychological magnitude. *Transactions of the Kansas Academy of Science*, *36*, 172–175.
Norman, J. F., Lappin, J. S., & Zucker, S. W. (1991). The discriminability of smooth stereoscopic surfaces. *Perception*, *20*(6), 789–807.
Norman, J. Farley, & Todd, J. T. (1995). The perception of 3-D structure from contradictory optical patterns. *Perception & Psychophysics*, *57*(6), 826–834. https://doi.org/10.3758/BF03206798
Norman, J. Farley, Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The Visual Perception of Three-Dimensional Length. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1). https://doi.org/10.1037/0096-1523.22.1.173
Ogle, K. N. (1952a). Distortion of the image by ophthalmic prisms. *A.M.A. Archives of Ophthalmology*, *47*(2), 121–131. https://doi.org/https://doi.org/10.1001/archopht.1952.01700030126001
Ogle, K. N. (1952b). On the limits of stereoscopic vision. *Journal of Experimental Psychology*, *44*(4), 253–259.

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America*, *43*(10), 907–913. https://doi.org/10.1364/josa.43.000906
Ono, H., & Comerford, J. (1977). Stereoscopic depth constancy. In *Stability and constancy in visual perception: Mechanisms and processes* (Epstein, W). Wiley.
Ono, H., & Steinbach, M. J. (1990). Monocular without stereopsis with and head movement. *Perception & Psychophysics*, *48*(2). https://doi.org/10.3758/BF03207085
Ono, M. E., Rivest, J., & Ono, H. (1986). Depth Perception as a Function of Motion Parallax and Absolute-Distance Information. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 331–337. https://doi.org/10.1037/0096-1523.12.3.331
Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature*, *414*(6860), 197–200.
Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23). https://doi.org/10.1016/S0042-6989(03)00435-8
Ozana, A., Berman, S., & Ganel, T. (2020). Grasping Weber's Law in a Virtual Environment: The Effect of Haptic Feedback. *Frontiers in Psychology*, *11*(November), 1–15. https://doi.org/10.3389/fpsyg.2020.573352
Palmisano, S., Gillam, B., Govan, D. G., Allison, R. S., & Harris, J. M. (2010). Stereoscopic perception of real depths at large distances. *Journal of Vision*, *10*(6), 1–16. https://doi.org/10.1167/10.6.19
Parker, A. J., Harris, J. M., Cumming, B. G., & Sumnall, J. H. (1996). Binocular correspondence in stereoscopic vision. *Eye*, *10*(2), 177–181. https://doi.org/10.1038/eye.1996.44
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4). https://doi.org/10.1163/156856897X00366
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, C. (2015). *nlme: Linear and nonlinear mixed effects models* (R package version 3.1-120). http://cran.rproject.org/package=nlme
Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience*, *7*, 379–412. https://doi.org/10.1146/annurev.ne.07.030184.002115
Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*. https://doi.org/10.2307/271063
Read, J. C. A., Begum, S. F., McDonald, A., & Trowbridge, J. (2013). The binocular advantage in visuomotor tasks involving tools. *I-Perception*, *4*(2), 101–110. https://doi.org/10.1068/i0565
Redding, G. M., & Wallace, B. (1996). Adaptive spatial alignment and strategic perceptual-motor control. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 379–394.
Redding, G. M., & Wallace, B. (2006). Generalization of prism adaptation. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1006–1022.
Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A*, *2*(2), 343–349. https://doi.org/10.1364/josaa.2.000343
Richards, W. (2009). Configuration stereopsis: A new look at the depth-disparity relation. *Spatial Vision*, *22*(1). https://doi.org/10.1163/156856809786618493
Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics*, *22*(4), 400–407. https://doi.org/10.3758/BF03199707
Rogers, B. (2022). Cues, clues and cognitivisation of perception: Do words matter? *Perception*, *51*(5), 295–299.
Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, *24*(2), 155–179. https://doi.org/10.1068/p240155
Rogers, Brian, & Cagenello, R. (1989). Disparity curvature and the perception of three-dimensional surfaces. *Nature*, *339*(6220), 135–137. https://doi.org/10.1038/339135a0
Rogers, Brian, & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*, 125–134.

Rogers, Brian, & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, *22*(2). https://doi.org/10.1016/0042-6989(82)90126-2
Rogers, Brian J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, *361*(6409), 253–255. https://doi.org/10.1038/361253a0
Rogers, Brian J., & Collett, T. S. (1989). The Appearance of Surfaces Specified by Motion Parallax and Binocular Disparity. *The Quarterly Journal of Experimental Psychology Section A*, *41*(4). https://doi.org/10.1080/14640748908402390
Rogers, Brian J., & Graham, M. E. (1983). Anisotropies in the perception of three-dimensional surfaces. *Science*, *221*(4618). https://doi.org/10.1126/science.6612351
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 231–244). Russel Sage Foundation.
Sato, M., & Howard, I. P. (2001). Effects of disparity-perspective cue conflict on depth contrast. *Vision Research*, *41*(4), 415–426.
Scarfe, P., & Glennerster, A. (2019). The science behind virtual reality displays. *Annual Review of Vision Science*, *5*, 529–547. https://doi.org/10.1146/annurev-vision-091718-014942
Scarfe, P., & Hibbard, P. B. (2006). Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy. *Vision Research*, *46*(10), 1599–1610. https://doi.org/10.1016/j.visres.2005.11.002
Scarfe, Peter, & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision*, *11*(7), 1–17. https://doi.org/10.1167/11.7.1
Sedgwick, H. A. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–57). Wiley-Interscience.
Servos, P., & Goodale, M. A. (1994). Binocular vision and the on-line control of human prehension. *Experimental Brain Research*, *98*(1), 119–127.
Servos, P., Goodale, M. A., & Jakobson, L. S. (1992). The role of binocular vision in prehension: A kinematic analysis. *Vision Research*, *32*(8), 1513–1521.
Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, *33*, 89–108. https://doi.org/10.1146/annurev-neuro-060909-153135
Sherman, A., Papathomas, T. V., Jain, A., & Keane, B. P. (2012). The role of stereopsis, motion parallax, perspective and angle polarity in perceiving 3-D shape. *Seeing and Perceiving*, *25*(3–4), 263–285. https://doi.org/10.1163/187847511X576802
Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, *11*(8), 11–11. https://doi.org/10.1167/11.8.11
Shum, L. C., Valdés, B. A., & van der Loos, H. M. (2019). Determining the accuracy of oculus touch controllers for motor rehabilitation applications using quantifiable upper limb kinematics: Validation study. *JMIR Biomedical Engineering*, *4*(1), e12291.
Snijders, H. J., Holmes, N. P., & Spence, C. (2007). Direction-dependent integration of vision and proprioception in reaching under the influence of the mirror illusion. *Neuropsychologia*, *45*(3), 496–505. https://doi.org/10.1016/j.neuropsychologia.2006.01.003
Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, *28*(3), 371–386.
Stevens, S. S. (1961). To Honor Fechner and Repeal His Law: A power function, not a log function, describes the operating characteristic of a sensory system. *Science*, *133*, 80–86.
Tassinari, H., Domini, F., & Caudek, C. (2008). The intrinsic constraint model for stereo-motion integration. *Perception*, *37*(1), 79–95. https://doi.org/10.1068/p5501
Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics*, *54*(2). https://doi.org/10.3758/BF03211751

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic Distorition of Perceived Three-Dimensional Structure From Motion and Binocular Stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 633–678.

Todd, J. T., Oomes, A. H. J., Koenderink, J. J., & Kappers, A. M. L. (2001). On the affine structure of perceptual space. *Psychological Science*, *12*(3), 191–196.

Todd, James T. (1985). The perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance*, *11*(6), 689–710. https://doi.org/10.1037/0096-1523.11.6.689

Todd, James T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, *8*(3), 115–121. https://doi.org/10.1016/j.tics.2004.01.006

Todd, James T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception and Psychophysics*, *65*(1), 31–47. https://doi.org/10.3758/BF03194781

Todd, James T., Thaler, L., Dijkstra, T. M. H., Koenderink, J. J., & Kappers, A. M. L. (2007). The effects of viewing angle, camera angle, and sign of surface curvature on the perception of three-dimensional shape from texture. *Journal of Vision*, *7*(12). https://doi.org/10.1167/7.12.9

Tresilian, J. R., Mon-Williams, M., & Kelly, B. M. (1999). Increasing confidence in vergence as a cue to distance. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1414), 39–44. https://doi.org/10.1098/rspb.1999.0601

Tsirlin, I., Wilcox, L. M., & Allison, R. S. (2012). Effect of crosstalk on depth magnitude in thin structures. *Journal of Electronic Imaging*, *21*(1), 011003.

Van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology*, *12*(10), 834–837. https://doi.org/10.1016/S0960-9822(02)00836-9

van Boxtel, J. J. A., Wexler, M., & Droulez, J. (2003). Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision*, *3*(5). https://doi.org/10.1167/3.5.1

Volcic, R., Fantoni, C., Caudek, C., Assad, J. A., & Domini, F. (2013). Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination. *Journal of Neuroscience*, *33*(43), 17081–17088. https://doi.org/10.1523/JNEUROSCI.2936-13.2013

Vreven, D. (2006). 3D shape discrimination using relative disparity derivatives. *Vision Research*, *46*(25), 4181–4192. https://doi.org/https://doi.org/10.1016/j.visres.2006.08.014.

Wade, N. J. (2021). On the Origins of Terms in Binocular Vision. *I-Perception*, *12*(1). https://doi.org/10.1177/2041669521992381

Wallach, H., & Zuckerman, C. (1963). The Constancy of Stereoscopic Depth. *The American Journal of Psychology*, *76*(3), 404–412.

Walsh, G., & Charman, W. N. (1988). Visual sensitivity to temporal change in focus and its relevance to the accommodation response. *Vision Research*, *28*(11), 1207–1221. https://doi.org/10.1016/0042-6989(88)90037-5

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, *35*(19), 2731–2736. https://doi.org/10.1016/0042-6989(95)00018-U

Watt, S. J., Akeley, K., & Banks, M. S. (2003). Focus cues to display distance affect perceived depth from disparity. *Journal of Vision*, *3*(9). https://doi.org/10.1167/3.9.66

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10), 7. https://doi.org/10.1167/5.10.7

Wei, K., & Kording, K. (2009). Relevance of error: What drives motor adaptation? *Journal of Neurophysiology*, *101*(2), 655–664.

Welch, R. B. (1969). Adaptation to prism-displaced vision: The importance of target-pointing. *Perception*

& Psychophysics, *5*(5), 305–309.
Westheimer, G. (1979). Cooperative neural process involved in stereoscopic acuity. *Experimental Brain Research*, *36*(3), 585–597.
Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research*, *17*(8), 941–947. https://doi.org/10.1016/0042-6989(77)90069-4
Wheatstone, C. (1838). 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*, *128*, 371–394.
Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*, *63*(8), 1293–1313. https://doi.org/10.3758/BF03194544
Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, *63*(8). https://doi.org/10.3758/BF03194545
Willemsen, P., Gooch, A. A., Thompson, W. B., & Creem-Regehr, S. H. (2008). Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, *17*(1), 91–101. https://doi.org/10.1162/pres.17.1.91
Witmer, B. G., & Kline, P. B. (1998). Judging Perceived and Traversed Distance in Virtual Environments. *Presence*, *7*(2), 144–167.
Witmer, B. G., & Sadowski, W. J. (1998). Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors*, *40*(3), 478–488. https://doi.org/10.1518/001872098779591340
Yoonessi, A., & Baker, C. L. (2011). Contribution of motion parallax to segmentation and depth perception. *Journal of Vision*, *11*(9), 13–13.
Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combination of texture and motion cues. *Vision Research*, *33*(18), 2685–2696.
Zhao, J., Allison, R. S., Vinnikov, M., & Jennings, S. (2017). Estimating the motion-to-photon latency in head mounted displays. *Proceedings - IEEE Virtual Reality*. https://doi.org/10.1109/VR.2017.7892302
Ziemer, C. J., Plumert, J. M., Cremer, J. F., & Kearney, J. K. (2009). Estimating distance in real and virtual environments: Does order make a difference? *Attention, Perception, and Psychophysics*, *71*(5), 1095–1106. https://doi.org/10.3758/APP

APPENDICES

Appendix 2.A: Summary of the Data and Analysis Independent of Condition Order

Table 2.A1: *The Linear Mixed-Effects Analysis Independent of Condition Order*

	Estimate	DF	*t*	*p*	*r*
Monocular					
Depth x Apparatus: PTE vs Stereoscope	-0.13	346	-2.61	0.01	0.14
Depth x Apparatus: PTE vs HMD	-0.19	346	-3.78	<0.001	0.20
Depth x Apparatus: HMD vs Stereoscope	0.06	346	1.26	0.21	0.07
Depth x Viewing Distance	-0.09	346	-2.40	0.02	0.13
Depth x Viewing Distance x Apparatus: PTE vs Stereoscope	-0.01	346	-0.32	0.75	0.02
Depth x Viewing Distance x Apparatus: PTE vs HMD	-0.03	346	-0.62	0.53	0.03
Depth x Viewing Distance x Apparatus: HMD vs Stereoscope	0.01	346	0.37	0.72	0.02
Binocular					
Depth x Apparatus: PTE vs Stereoscope	-0.10	346	-1.44	0.15	0.08
Depth x Apparatus: PTE vs HMD	-0.20	346	-2.73	0.01	0.15
Depth x Apparatus: HMD vs Stereoscope	0.09	346	1.34	0.18	0.07
Depth x Viewing Distance	-0.10	346	-2.47	0.01	0.13
Depth x Viewing Distance x Apparatus: PTE vs Stereoscope	-0.06	346	-1.15	0.25	0.06
Depth x Viewing Distance x Apparatus: PTE vs HMD	-0.04	346	-0.82	0.41	0.04
Depth x Viewing Distance x Apparatus: HMD vs Stereoscope	-0.02	346	-0.38	0.70	0.02

Note. Depth refers to the slope of estimates as a function of predicted depth.

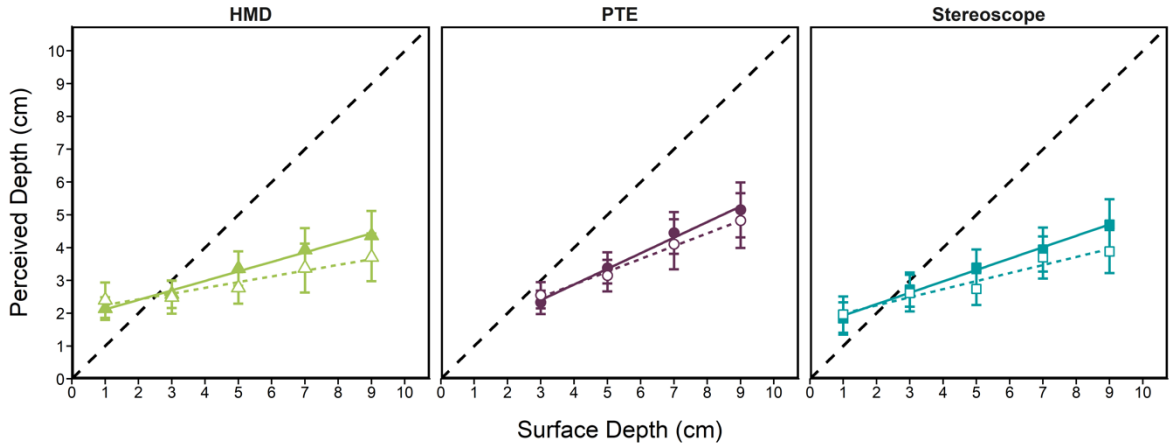


Figure 2.A1. *Mean perceived depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under monocular viewing conditions. The dashed line represents the accurate depth estimates and error bars represent the standard error of the mean.*

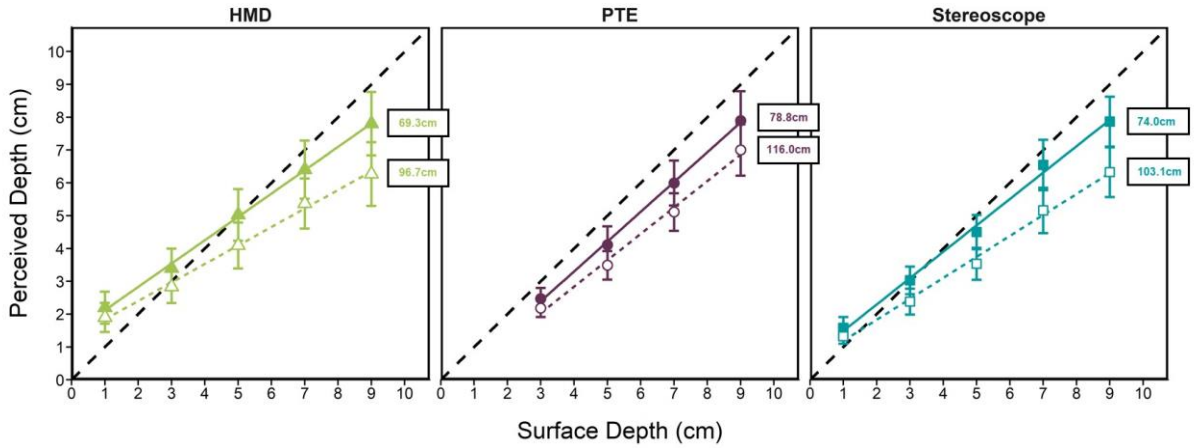


Figure 2.A2. *Mean perceived depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open points, respectively) under binocular viewing conditions. The inferred viewing distance is annotated for each condition (in cm). The dashed line represents accurate depth estimates and error bars represent the standard error of the mean.*

Table 2.A2: *The Accuracy and Stereoscopic Depth Constancy Analyses Independent of Condition Order*

	Estimate	DF	*t*	*p*	*r*
Accuracy Relative to Theoretical					
PTE					
Near Viewing Distance					
Intercept	-0.33	15	-0.46	0.65	0.12
Slope	-0.09	94	-0.63	0.53	0.06
Far Viewing Distance					
Intercept	-0.38	15	-0.60	0.56	0.15
Slope	-0.20	94	-1.53	0.13	0.16
Stereoscope					
Near Viewing Distance					
Intercept	0.69	15	1.81	0.09	0.42
Slope	-0.20	126	-2.06	0.04	0.18
Far Viewing Distance					
Intercept	0.55	15	1.60	0.13	0.38
Slope	-0.36	126	-4.70	<0.0001	0.39
HMD					
Near Viewing Distance					
Intercept	1.41	15	2.61	0.02	0.56
Slope	-0.29	126	-2.68	0.01	0.23
Far Viewing Distance					
Intercept	1.28	15	2.98	0.01	0.61

Slope	-0.44	126	-4.64	<0.0001	0.38
Stereoscopic Depth Constancy					
Viewing Distance x Apparatus: PTE					
Intercept	-0.05	15	-0.20	0.84	0.05
Slope	-0.10	94	-2.18	0.03	0.19
Viewing Distance x Apparatus: Stereoscope					
Intercept	-0.14	15	-0.89	0.39	0.22
Slope	-0.16	126	-4.26	<0.0001	0.35
Viewing Distance x Apparatus: HMD					
Intercept	-0.14	15	-0.80	0.44	0.20
Slope	-0.15	126	-5.79	<0.0001	0.46

Notes. The accuracy relative to theoretical analysis is comparing the intercept and slope of depth estimates in each apparatus and viewing distance relative to a theoretical observer with perfect accuracy. The stereoscopic depth constancy analysis is comparing the relative intercepts and slopes for the two viewing distances for each apparatus.

Appendix 3.A: Individual PSE and JND Estimates

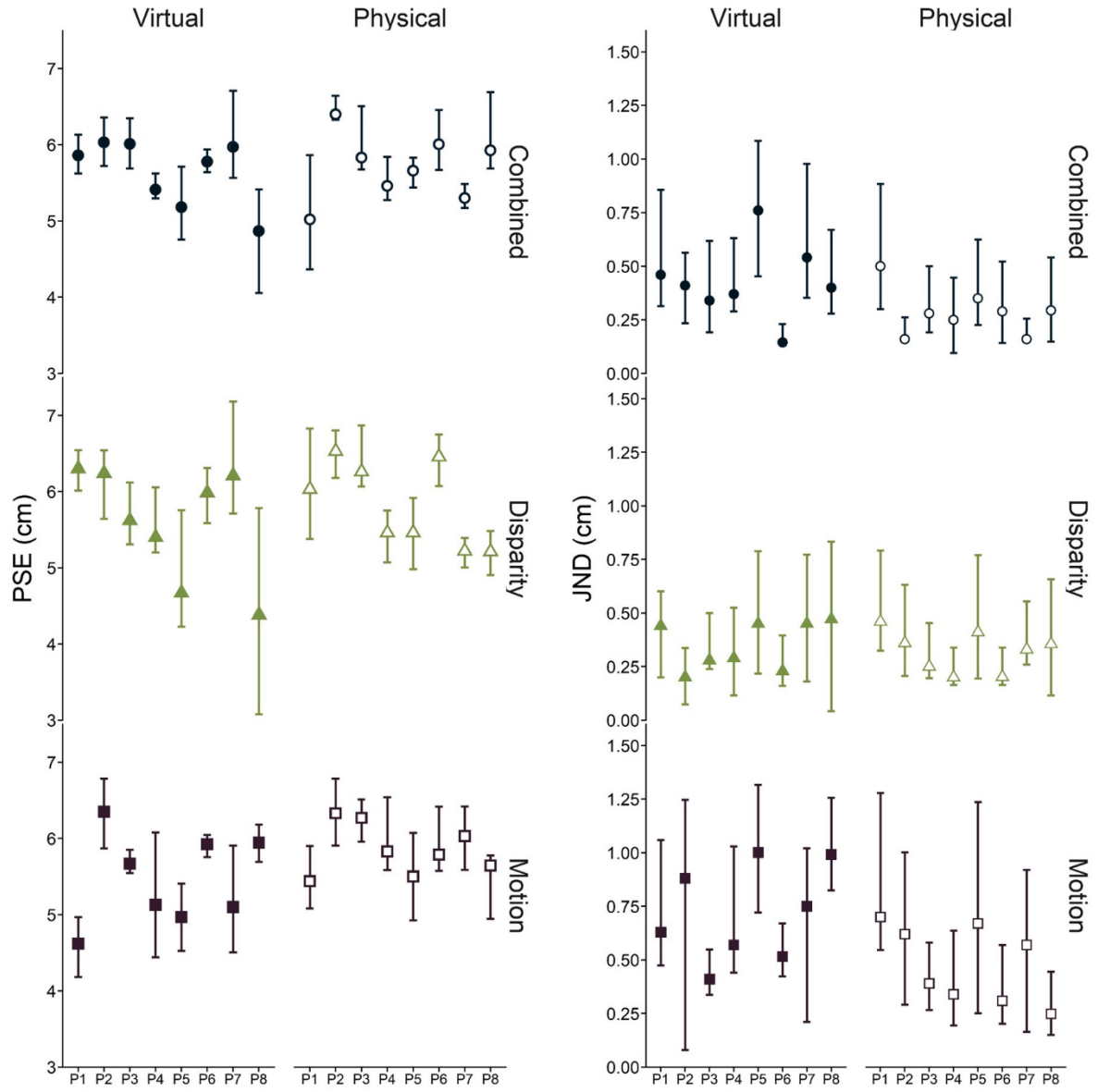


Figure 3.A1. *The PSEs and JNDs for each of the three cue conditions: binocular disparity only (green triangles), motion parallax (purple squares), and their combination (blue circles) for each observer (n = 8) in the virtual and physical viewing conditions. Error bars represent 95% confidence intervals.*

Appendix 3.B: Bayesian Model Fits

Table 3.B1: *Bayesian Model Fits*

Model Comparison							
Virtual Environment				Physical Environment			
Observer	**Model**	**BIC**	**Difference**	**Observer**	**Model**	**BIC**	**Difference**
P1	Linear	7093.0	0.0	P1	Linear	2853.6	0.0
	Correlated	1512.8	-5580.2		**Correlated**	1450.7	-1402.9

	Veto	58.6	-7034.3		Veto	1476.9	-1376.7
	Linear	5834.9	0.0		Linear	190.7	0.0
P2	Correlated	4501.6	-1333.3	P2	Correlated	3068.9	2878.2
	Veto	4392.1	-1442.8		**Veto**	66.1	-124.6
	Linear	4368.6	0.0		**Linear**	1443.1	0.0
P3	Correlated	2942.7	-1425.9	P3	Correlated	2921.9	1478.8
	Veto	1501.2	-2867.4		Veto	2904.7	1461.6
	Linear	3075.1	0.0		Linear	1551.9	0.0
P4	Correlated	1441.0	-1634.1	P4	**Correlated**	36.7	-1515.2
	Veto	26.4	-3048.7		**Veto**	30.7	-1521.2
	Linear	4320.4	0.0		Linear	4403.8	0.0
P5	Correlated	4291.8	-28.6	P5	Correlated	3017.3	-1386.5
	Veto	2884.9	-1435.5		**Veto**	1514.2	-2889.6
	Linear	1531.1	0.0		Linear	2859.6	0.0
P6	Correlated	2908.1	1377.0	P6	Correlated	1523.7	-1335.9
	Veto	105.3	-1425.8		**Veto**	1500.6	-1359.0
	Linear	2923.3	0.0		Linear	101.9	0.0
P7	**Correlated**	1445.5	-1477.8	P7	Correlated	2933.0	2831.1
	Veto	1440.1	-1483.2		**Veto**	26.2	-75.7
	Linear	1469.1	0.0		Linear	2988.4	0.0
P8	Correlated	1467.1	2.0	P8	Correlated	3029.7	41.3
	Veto	1441.8	-27.3		**Veto**	205.3	-2783.1

Note. Bold model names indicate the best-fitting model to the observed data. If two models are bolded, then the difference between the BIC difference values is less than 10 and both models are equally valid.

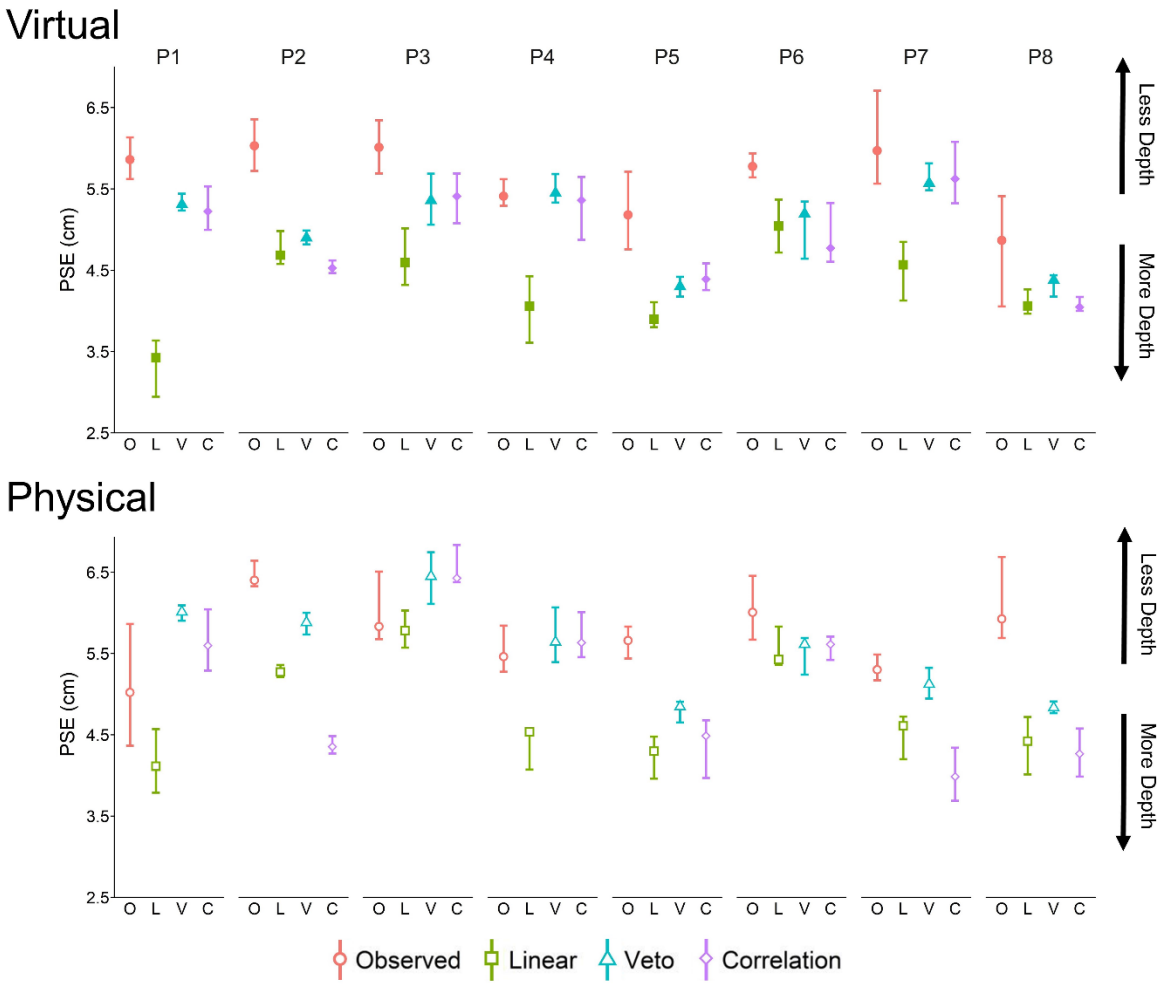


Figure 3.B1. *The measured PSEs for the combination of binocular disparity and motion parallax, and the predicted PSEs for the linear, veto, and correlated models for each observer (n = 8) in the virtual and physical viewing conditions. Error bars represent 95% confidence intervals.*

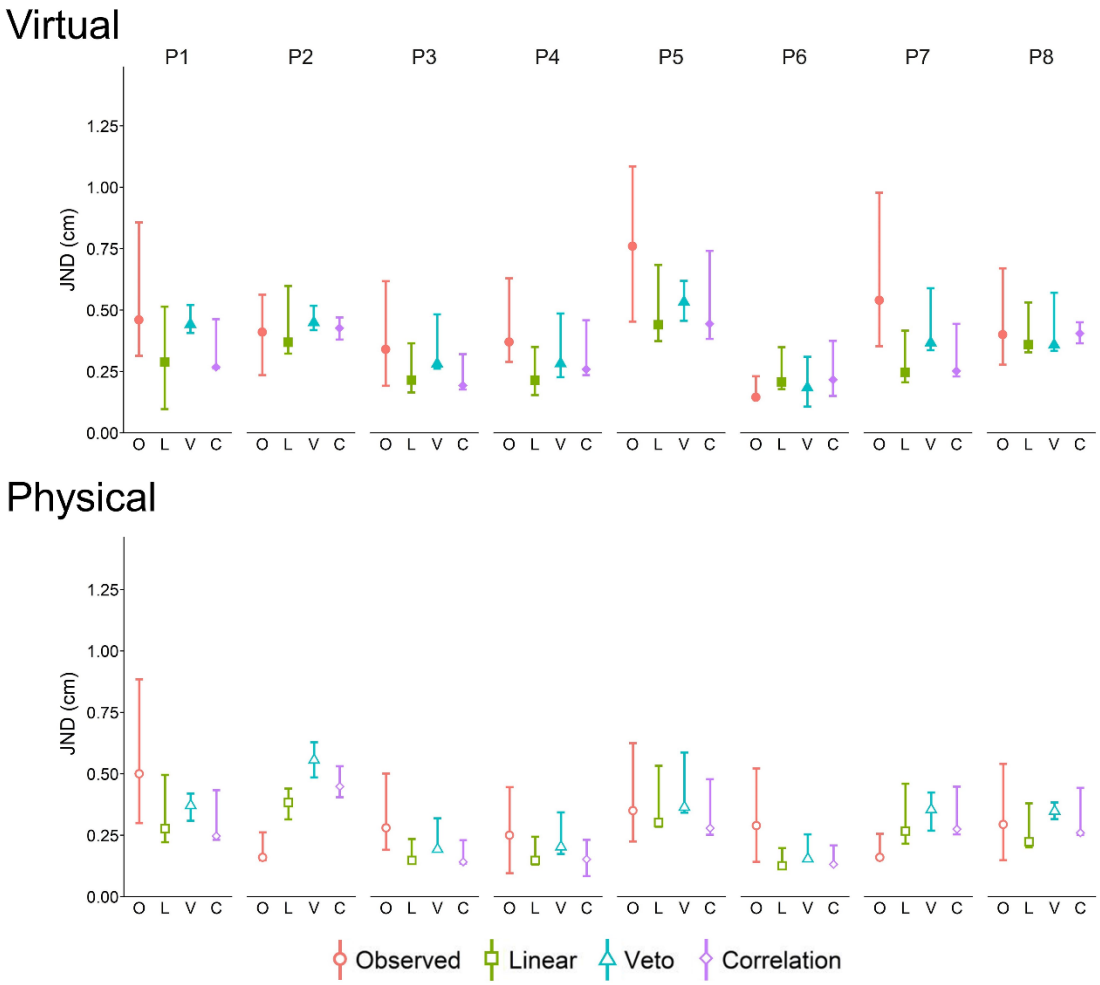


Figure 3.B2. *The measured JNDs for the combination of binocular disparity and motion parallax, and the predicted JNDs for the linear, veto, and correlated models for each observer (n = 8) in the virtual and physical viewing conditions. Error bars represent 95% confidence intervals.*

Appendix 4.A: Proprioceptive Data for the Reach Task in Experiment 2

Figure 4.A1 shows the blind reach estimates for the proprioceptive assessment task for both the right and left hand reaches. Figure 4.A1.A shows that the distance of the reaches was underestimated, and too close to the midline, similar to the results of Experiment 1. The reach errors were less accurate on average for the left hand (10.34cm) than the right hand (6.05cm), which is expected given most observers were right-handed (Figure 4.A1.B). The large overestimates in the right hand distribution are the result of a single observer that showed large errors throughout the proprioceptive assessment task. The direction of the reach errors made during the proprioceptive assessment in Experiment 2 were consistent with Experiment 1; however, this group of observers made larger errors (especially with their left hand) than the observers in Experiment 1.

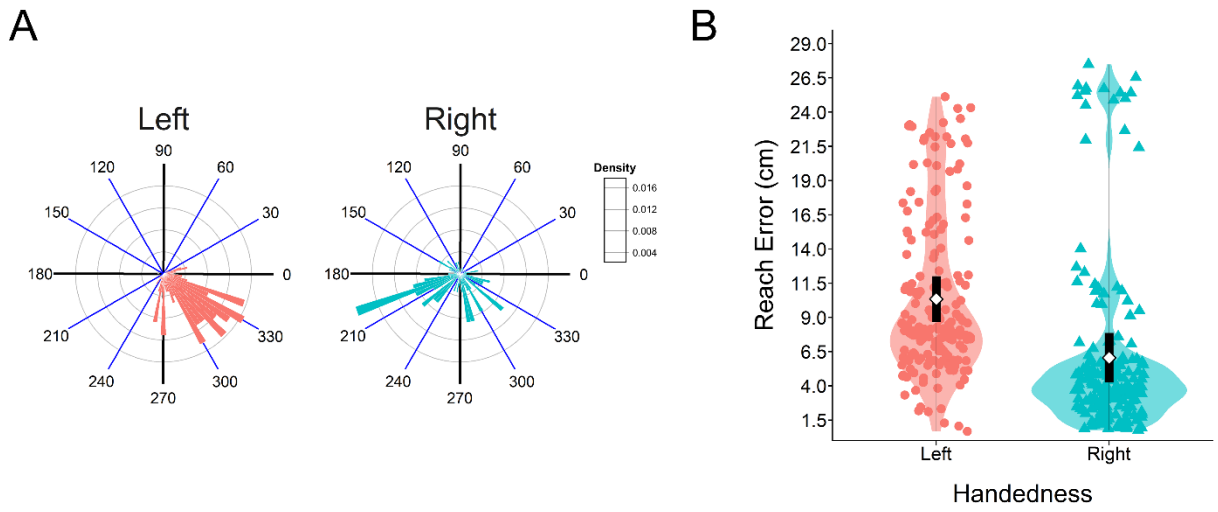


Figure 4.A1. *Graph A shows a radial histogram of the blind reach estimates for each hand for all observers in the proprioceptive assessment task. A value of 90 deg is an error further than the target peg, and 270 deg is an error too far in front of the target peg. Graph B shows the magnitude of all reach errors made in the proprioceptive assessment task in cm. The white diamond represents the mean of the distribution, and the black box represents the standard error of the mean.*

Appendix 5.A: PSE and JND Data for the Follow-up Motion Parallax and Binocular Disparity Experiment

Table 5.A1: *PSE and JND Analyses for Follow-up Motion Parallax Experiment*

	Estimate	DF	*t*	*p*	*r*
PSE					
Combined vs. Binocular Disparity	-0.35	12	-1.64	0.13	0.43
Combined vs. Motion Parallax	-0.21	12	-1.00	0.34	0.28
Motion Parallax vs. Binocular Disparity	-0.56	12	-2.64	0.02	0.61
JND					
Combined vs. Binocular Disparity	-0.02	12	-0.29	0.78	0.08
Combined vs. Motion Parallax	0.16	12	2.17	0.05	0.53

Motion Parallax vs. Binocular Disparity	0.14	12	1.88	0.08	0.48