Stereoscopic depth constancy for physical objects and their virtual counterparts

Brittney Hartle

Department of Psychology and Centre for Vision Research, York University, Toronto, Ontario, Canada ✉

Laurie M. Wilcox

Department of Psychology and Centre for Vision Research, York University, Toronto, Ontario, Canada ✉

Stereopsis plays an important role in depth perception; if so, disparity-defined depth should not vary with distance. However, studies of stereoscopic depth constancy often report systematic distortions in depth judgments over distance, particularly for virtual stimuli. Our aim was to understand how depth estimation is impacted by viewing distance and display-based cue conflicts by replicating physical objects in virtual counterparts. To this end, we measured perceived depth using virtual textured half-cylinders and identical three-dimensional (3D) printed versions at two viewing distances under monocular and binocular conditions. Virtual stimuli were viewed using a mirror stereoscope and an Oculus Rift head-mounted display (HMD), while physical stimuli were viewed in a controlled test environment. Depth judgments were similar in both virtual apparatuses, which suggests that variations in the viewing geometry and optics of the HMD have little impact on perceived depth. When viewing physical stimuli binocularly, judgments were accurate and exhibited stereoscopic depth constancy. However, in all cases, depth was underestimated for virtual stimuli and failed to achieve depth constancy. It is clear that depth constancy is only complete for cue-rich physical stimuli and that the failure of constancy in virtual stimuli is due to the presence of the vergence-accommodation conflict. Further, our post hoc analysis revealed that prior experience with virtual and physical environments had a strong effect on depth judgments. That is, performance in virtual environments was enhanced by limited exposure to a related task using physical objects.

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 2 m) along the midline (for a 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, Gooch, Thompson, & Creem-Regehr, 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, 1985; Foley & Richards, 1972; Rogers & Bradshaw, 1993; Wallach & Zuckerman, 1963). Unless the content fills a wide area of the visual field vertical disparity signals are very weak (Backus, Banks, van Ee, & Crowell, 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; Gogel, 1977; Johnston, 1991; Komoda & Ono, 1974; Linton, 2020), except at distances less than 30 cm (Mon-Williams, Tresilian, & Roberts, 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 judgments and increasing discomfort (Hoffman, Girshick, Akeley, & Banks, 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, Proffitt, Olson, & Reinke, 1995; Ritter, 1977). Although 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, Buckley, & Horsman, 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 judgments. 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 judgments. In other experiments, simultaneous matching tasks have shown that the relative depth from binocular disparity is accurate at viewing distances under 2 m (Glennerster, Rogers, & Bradshaw, 1996), whereas 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, Applebaum, & Richards, 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, 1952a; Ogle, 1952b; Ogle, 1953), ruler adjustment (Tsirlin, Wilcox, & Allison, 2012), or haptic matching tasks (Brenner & van Damme, 1999; Hornsey, Hibbard, & Scarfe, 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).

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 judgments[1]. 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 (83 cm and 130 cm) 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.

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 the Procedure section) based on the slope of depth judgments 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 judgments 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 tends 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, but be more accurate at the near relative to the far condition. If observers achieve depth constancy in any viewing condition, then the magnitude of depth judgments 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 judgments) 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, Warm, & Rushton, 1993; Ogle, 1952a). Although 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 200 cm (HMD) versus 74 cm (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 decreases the reliability 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.

Methods

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 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.

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 14 cm, which subtended 9.6 degrees and 6.2 degrees 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 9 cm. However, the 1 cm 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
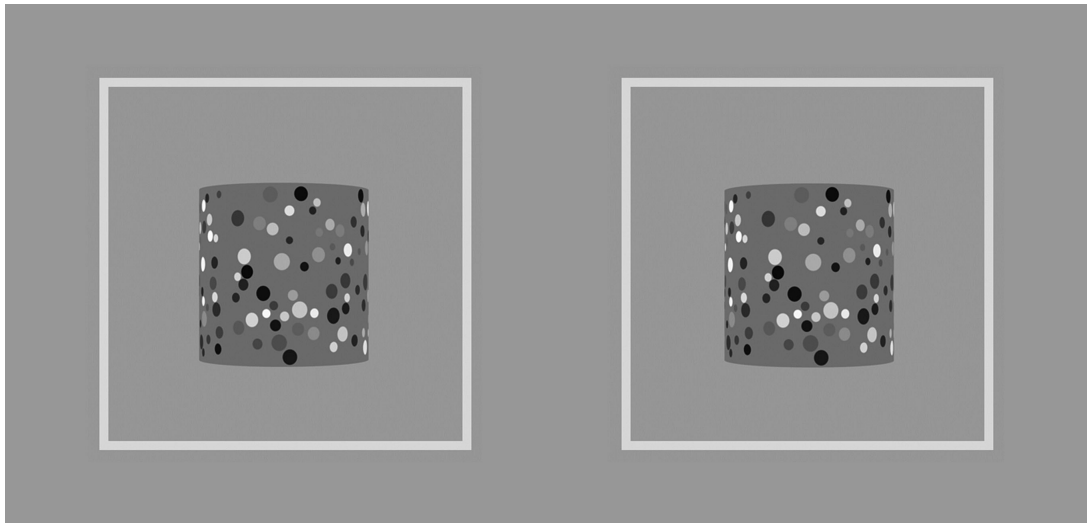
Figure 1. A stereopair of a textured half-cylinder stimulus. The stereopair is arranged for crossed fusion. The cylinder and reference frame are not to scale.

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 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, Bülthoff, & Sheinberg, 1993; Cumming, Johnston, & Parker, 1993). The textures were generated in MATLAB. The radius of the circular elements ranged from 0.38 to 0.95 degrees at the near viewing distance, and 0.24 degrees to 0.60 degrees 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 180 degrees between blocks to make the fixed position of the texture elements an uninformative reference for depth judgments.

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 judgments. The frame subtended 21.8 degrees and 14.0 degrees at the near and far viewing distances, respectively. The distance between the edge of the cylinder and the inner edge of the frame was 5.5 degrees and 3.5 degrees 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 83 cm and 130 cm 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. Although focal blur is available and could aid estimates in the physical environment, the two viewing distances (83 cm and 130 cm) were chosen so any focal differences between the edge and peak of the surface should be perceptually indistinguishable (Hoffman & Banks, 2010; Watt, Akeley, Ernst, & Banks, 2005). That is, the difference in focal blur between the reference frame and the peak of the surface at the largest depth of 9 cm would be approximately 0.15 D and 0.06 D at the near and far viewing distances, respectively. Given the eyes' depth of focus under

typical viewing scenarios is approximately 0.33 D (Campbell, 1957; 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. Whereas 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. Although horizontal disparities provide information regarding surface curvature, vertical disparities provide information regarding the absolute distance to the object, although this information has been shown to be available primarily for stimuli that fill a large visual field (e.g. 70 degrees) at viewing distances below 50 cm (Backus et al., 1999; Bradshaw, Glennerster, & Rogers, 1996; Rogers & Bradshaw, 1995). Thus, vertical disparities are unlikely to play a significant role in our study.

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 75 Hz. Each monitor had a viewing distance of 74 cm from the observer and a horizontal field-of-view of 25 degrees. 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.
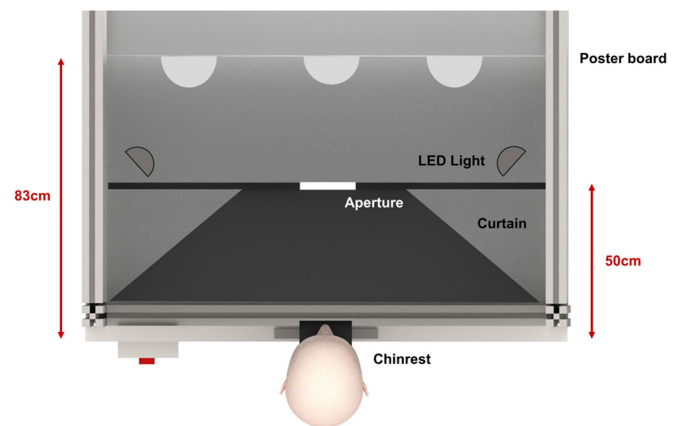

Figure 2. A top-down illustration of the PTE apparatus at the near viewing distance. The poster board is shown 83 cm from the observer. An aperture made from a black poster board was positioned 50 cm from the observer between the LED light fixtures and the enclosure curtain. The black curtains framed the apparatus, blocking residual light and the observer's view of the inside of the enclosure.

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 90 Hz and focal distance of approximately 200 cm. At a horizontal field-of-view of 94 degrees, 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 millimeter). 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; 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 0.4 mm per meter 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 3.8 cm thick polystyrene board (122 cm by 61 cm) 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 50 cm in front of the observer to limit the horizontal field of view to 29 degrees and 19 degrees 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.

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, whereas the other half completed the two virtual conditions first.

The data were analyzed using a linear mixed-effects model using the nlme package in R (Pinheiro, Bates, DebRoy, Sarkar, & Core Team, 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, Miles, & Field, 2012). The analysis focused on the comparison of the slope of the functions (estimated versus 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 judgments (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.

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 A, Tables A1 and A2 along

with summary plots of the monocular and binocular data (Figures A1, 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 four-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 judgments. The analysis was conducted separately for the monocular and binocular viewing conditions.

Monocular viewing

In the monocular viewing condition, the analysis revealed a lack of significant four-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, $\chi^2(28) = 0.33$, $p = 0.85$. However, the analysis did show a significant three-way interaction that suggested the order of test conditions impacted the slope of depth estimates for the different types of apparatuses, $\chi^2(23) = 9.75$, $p < 0.01$. No significant three-way interaction was found between the slope of depth estimates, viewing distance, and the order of conditions, $\chi^2(24) = 1.75$, $p = 0.19$. 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.

Virtual-first observers

Figure 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, $b = -0.07$, $t(170) = -1.32$, $p = 0.19$, $r = 0.10$.

Physical-first observers

Figure 4 depicts estimated depth as a function of predicted depth for each apparatus and viewing distance
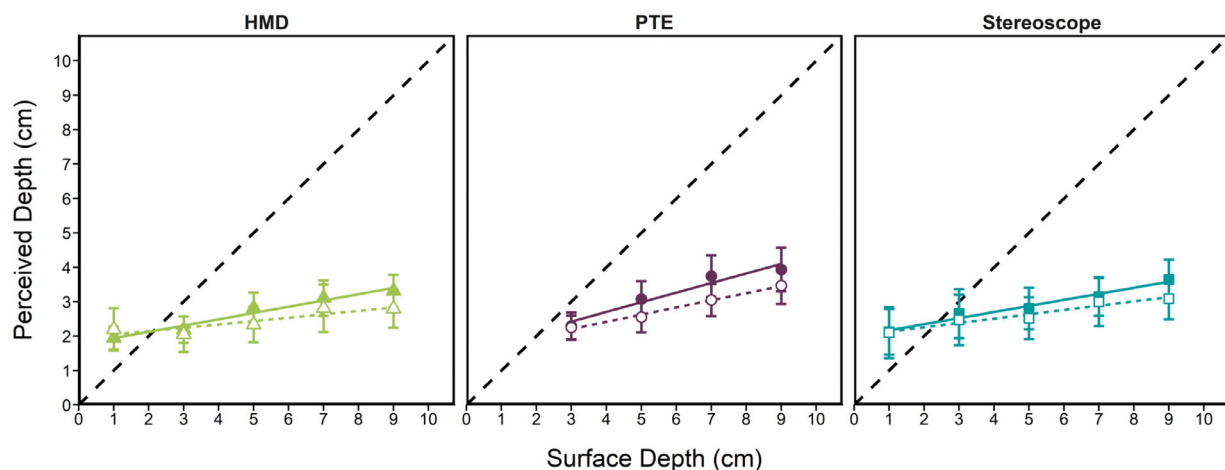

Figure 3. Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under monocular viewing conditions for virtual-first observers ($n = 8$). The dashed line represents accurate depth estimates and error bars represent the standard error of the mean.
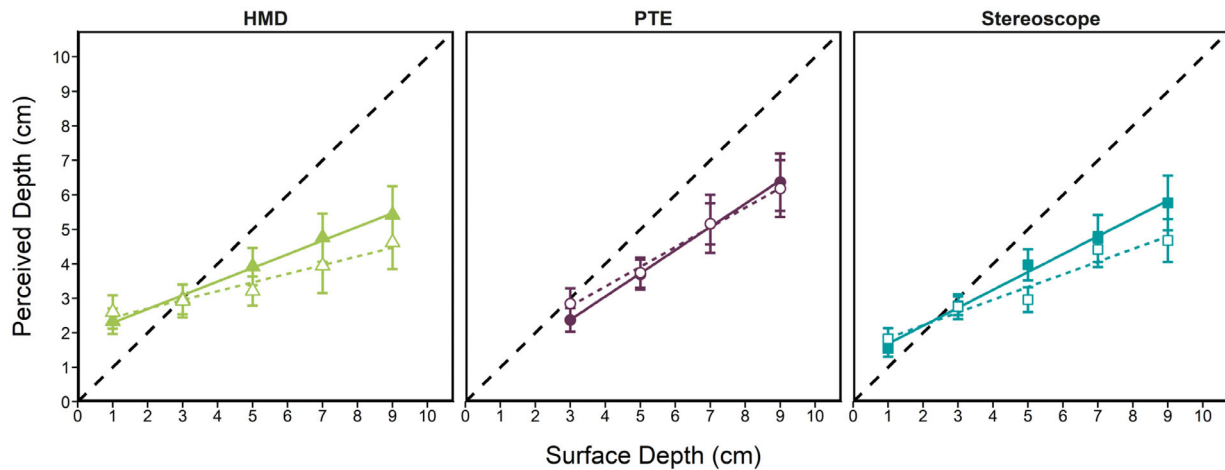
Figure 4. Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under monocular viewing conditions for physical-first observers ($n = 8$). The dashed line represents accurate depth estimates and error bars represent the standard error of the mean.

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.

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), whereas 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 3 and 4 show 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 viewing distance, 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.

Binocular viewing

Unsurprisingly, compared to the monocular conditions, when binocular cues were available perceived depth estimates were more accurat (see Figures 5, 6). Unlike the monocular analyses, the analysis of the binocular data revealed a significant four-way interaction showed that the relationship between the predicted and perceived surface depth depended on the type of apparatus, viewing distance, and condition order, $\chi^2 (28) = 10.94$, $p < 0.01$. 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.
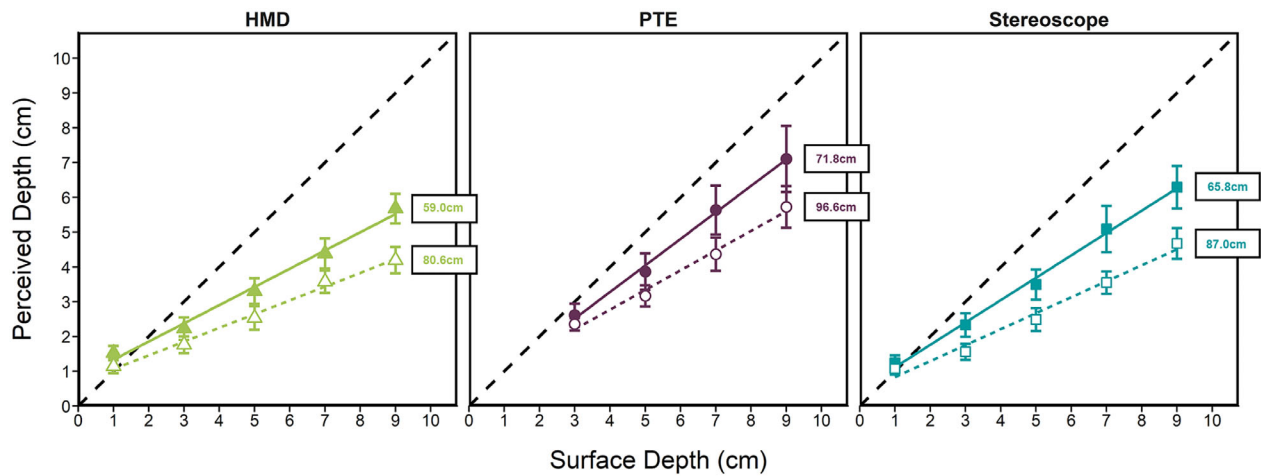
Figure 5. Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under binocular viewing conditions for virtual-first observers ($n = 8$). 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.
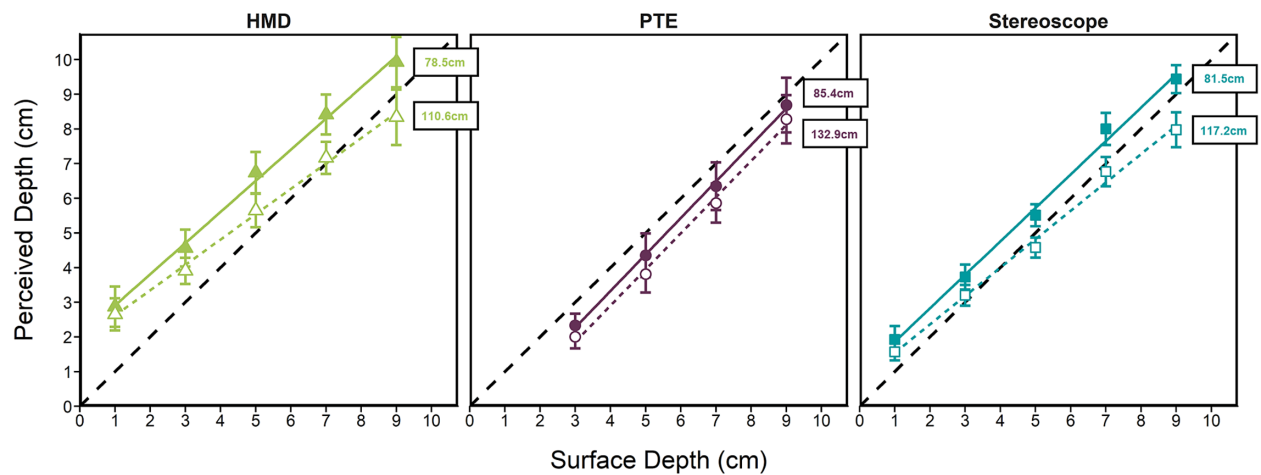

Figure 6. Mean depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open symbols, respectively) under binocular viewing conditions for physical-first observers ($n = 8$). 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.

Virtual-first observers

Figure 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, $b = -0.20$, $t(170) = -2.63$, $p = 0.01$, $r = 0.20$.

Physical-first observers

Figure 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 with observers that completed the task for virtual stimuli first. There was no significant difference in slopes obtained using physical stimuli versus the stereoscope, b = −0.09, t(170) = −1.03, $p = 0.30$, r = 0.08, or versus 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.

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.

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 among 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 nonlinear, given the narrow range of disparities presented in our study (0.5 degrees 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 5 and 6 show the inferred viewing distance in centimeters for each linear fit, and Figure 7 shows the individual inferred viewing distance estimates for each observer group in all viewing conditions.

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 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 1.

Table 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. Although observers that completed the physical test condition first showed slopes consistent
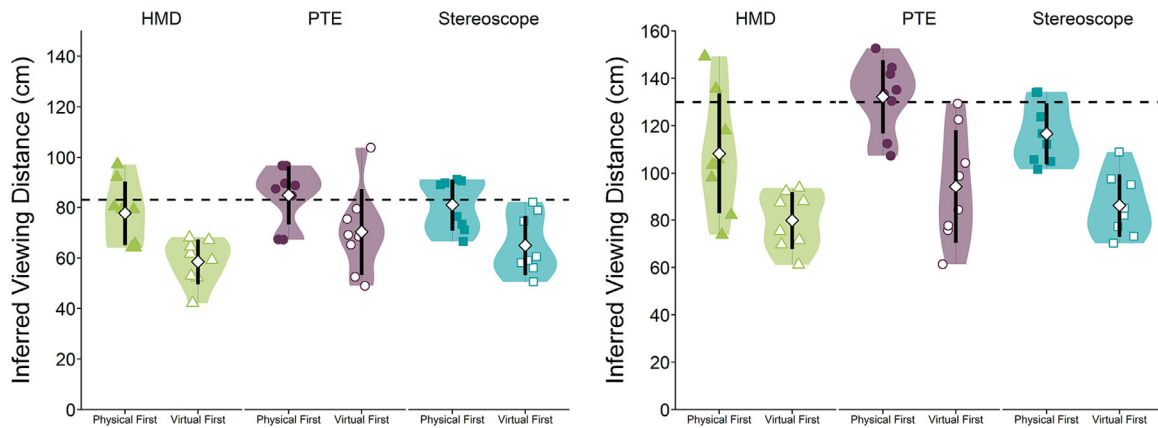
Figure 7. Inferred viewing distance estimates for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near (left plot) and far viewing distances (right plot), for the physical-first and virtual-first observers (filled and open symbols, respectively). The dashed line represents the viewing distance to the reference frame in the near and far conditions (83 cm and 130 cm, 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.

	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

Table 1. Accuracy of depth scaling relative to ideal observer model. Note: Each test compares the slope of the ideal observer model to the slope of the data.

with theoretical predictions for virtual stimuli at the near viewing distance, but not for the far viewing distance.

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.

Discussion

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 judgments 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 judgments 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.

HMD versus 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, 1952a). Our comparison of depth judgments for virtual stimuli rendered in the stereoscope versus 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 5 – HMD and Stereoscope). The overestimation at the smallest

cylinder depth was also seen in the combined analysis in Appendix A (see Figure 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 1 cm half-cylinder was 2.7 and 1.1 arcmin for the 83 cm and 130 cm 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.

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 the Stimuli section), 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, Gray, Carlin, & Mon-Williams, 2000; Hoffman & Banks, 2010; Shibata, Kim, Hoffman, & Banks, 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 judgments of virtual objects at two viewing distances in apparatuses with different focal distances (i.e. 74 cm and 200 cm for the stereoscope and HMD, respectively). At the near test distance of 83 cm 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 130 cm. 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, Cumming, & Landy, 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 (see Figure 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, Buckley, & Duke, 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 (Gårding, Porrill, Mayhew, & Frisby, 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, Schwarz, & Sobel, 1991; Foley, 1980; Glennerster, Rogers, & Bradshaw, 1998; Johnston, 1991; Tittle, Todd, Perotti, & Norman, 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 50 cm (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.

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 (see Figures 3, 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 judgments under binocular viewing, the virtual depth judgements for physical-first observers (see Figure 6 – HMD and Stereoscope) were compared to virtual judgments for virtual-first observers (see Figure 5 – HMD and Stereoscope). Observers with experience with the physical stimuli showed accurate depth scaling for all virtual depth judgments at near viewing distances, but observers without this experience did not (see Figure 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 judgments (compare Figures 4 and 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, Plumert, Cremer, & Kearney, 2009). Further, when the virtual and physical environments are made more similar, distance judgments were equivalent in both environments when the physical space was experienced first (Interrante, Ries, & Anderson, 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 judgments 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 (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, Kölsch, & Konrad, 2020). Although the explanation is consistent with some aspects of our data, it does not explain the substantial reduction in accuracy for physical judgments 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 judgments with distance in a virtual version.

Conclusions

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 judgments 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 judgments. Thus, it is important to consider, and perhaps control, participants' familiarity with test environments in judgments 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.

Keywords: binocular disparity, depth, absolute distance, virtual reality, real-world environment

Acknowledgments

Footnote

[1] 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).

References

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.

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

Blake, A., Bülthoff, H. H., & Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research, 33*, 1723–1737.

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.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Brenner, E., & Landy, M. S. (1999). Interaction between the perceived shape of two objects. *Vision Research, 39*(23), 3834–3848.

Brenner, E., & van Damme, W. J. M. (1999). Perceived distance, shape and size. *Vision Research, 39*(5), 975–986.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research, 33*, 919–933.

Campbell, F. W. (1957). The depth of field of the human eye. *Journal of Modern Optics, 4*, 157–164.

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

Coutant, B.E., & Westheimer, G. (1993). Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics, 13*, 3–7.

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.

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.

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.

Erkelens, C. J. (2000). Perceived direction during monocular viewing is based on signals of the viewing eye only. *Vision Research, 40*(18), 2411–2419.

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.

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks, CA: Sage.

Foley, J. M. (1977). Effect of distance information and range on two indices of visually perceived distance. *Perception, 6*, 449–460.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review, 87*(5), 411–434.

Foley, J. M. (1985). Binocular distance perception: Egocentric distance tasks. *Journal of Experiment Psychology: Human Perception and Performance, 11*(2), 133–149.

Foley, J. M., Applebaum, T. H., & Richards, W. A. (1975). Stereopsis with large disparities: discrimination and depth magnitude. *Vision Research, 15*(3), 417–421.

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.

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.

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.

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*, 129–154.

Gårding, J., Porrill, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research, 35*, 703–722.

Gibaldi, A., & Banks, M. S. (2019). Binocular eye movements are adapted to the natural environment. *Journal of Neuroscience, 39*(15), 2877–2888.

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, A., Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception, 27*(11), 1357–1365.

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, W. C. (1977). An indirect measure of perceived distance from oculomotor cues. *Perception & Psychophysics, 21*(1), 3–11.

Gogel, W. C., & Tietz, J. D. (1979). A comparison of oculomotor and motion parallax cues of egocentric distance. *Vision Research, 19*(10), 1161–1170.

Hartle, B., & Wilcox, L. M. (2016). Depth magnitude form stereopsis: Assessment techniques and the role of experience. *Vision Research, 125*, 64–75.

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

Hoffman, D. M., & Banks, M. S. (2010). Focus information is used to interpret binocular images. *Journal of Vision, 10*(5), 13.

Hoffman, D. 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), 33.

Hornsey, R. L., Hibbard, P. B., & Scarfe, P. (2020). Size and shape constancy in consumer virtual reality. *Behavior Research Methods, 52*(4), 1587–1598.

Howard, I. P., & Rogers, B. J. (2012). *Perceiving in depth, Volume 2: Stereoscopic vision*. Oxford, England, UK: Oxford University Press.

Interrante, V., Ries, B., & Anderson, L. (2006, March). Distance perception in immersive virtual environments, revisited. In *IEEE virtual reality conference (VR 2006)* (pp. 3–10). IEEE.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34*(17), 2259–2275.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31*(7–8), 1351–1360.

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.

Kerrigan, I. S., & Adams, W. J. (2013). Learning different light prior distributions for different contexts. *Cognition, 127*(1), 99–104.

Knill, D. C. (1998). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research, 38*, 1683–1711.

Komoda, M. K., & Ono, H. (1974). Oculomotor adjustments and size-distance perception. *Perception & Psychophysics, 15*(2), 353–360.

Linton, P. (2020). Does vision extract absolute distance from vergence? *Attention, Perception, & Psychophysics, 82*(6), 3176–3195.

McKee, S.P. (1983). The spatial requirements for fine stereoacuity. *Vision Research, 23*(2), 191–198.

Mon-Williams, M., & Tresilian, J. R. (2000). Ordinal depth information from accommodation? *Ergonomics, 43*, 391–404.

Mon-Williams, M., Tresilian, J. R., & Roberts, A. (2000). Vergence provides veridical depth

perception from horizontal retinal image disparities. *Experimental Brain Research, 133*(3), 407–413.

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*, 387–391.

Ogle, K. N. (1952a). Distortion of the image by ophthalmic prisms. *AMA Archives of Ophthalmology, 47*(2), 121–131.

Ogle, K. N. (1952b). On the limits of stereoscopic vision. *Journal of Experimental Psychology, 44*, 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), 906–913.

Ono, H., & Comerford, J. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes*. Hoboken, NJ: Wiley & Sons.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. R., & Team, Core. (2015). nlme: Linear and nonlinear mixed effects models (R package version 3.1-120). Retrieved from http://CRAN.Rproject.org/package=nlme.

Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics, 22*(4), 400–407.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature, 361*(6409), 253–255.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception, 24*, 155–179.

Scarfe, P., & Glennerster, A. (2019). The science behind virtual reality displays. *Annual Review of Vision Science, 5,* 529–547.

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.

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.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 663.

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

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.

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*, 1207–1221.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5*(10), 7.

Westheimer, G., & McKee, S.P. (1977). Spatial configurations for visual hyperacuity. *Vision Research, 17*, 941–947.

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.

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. J. (1998). Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors, 40*, 478–488.

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, & Psychophysics, 71*(5), 1095–1106.

Appendix

Appendix A: Summary of the data and analysis independent of condition order

Figure 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.
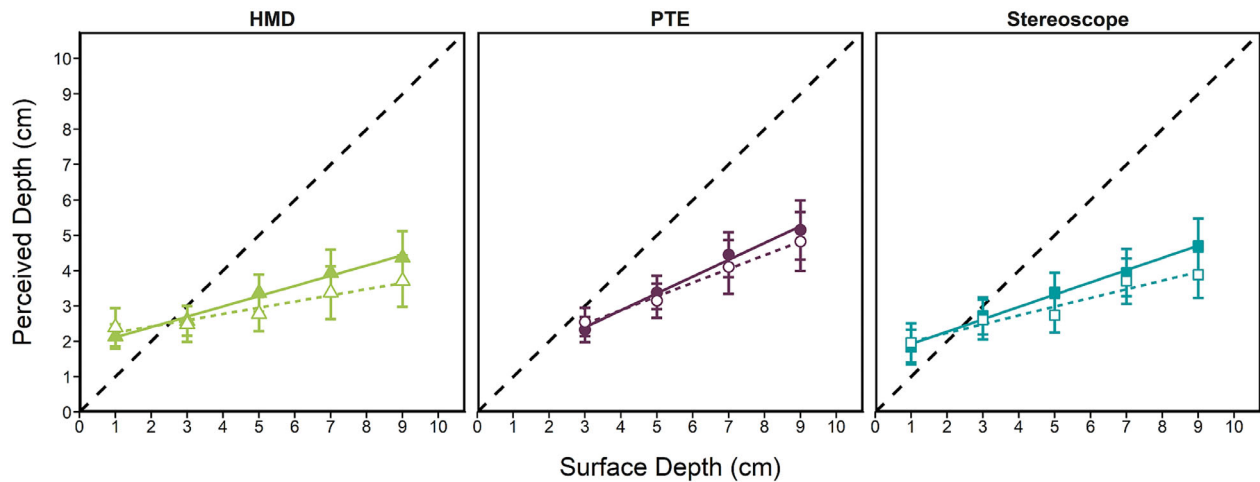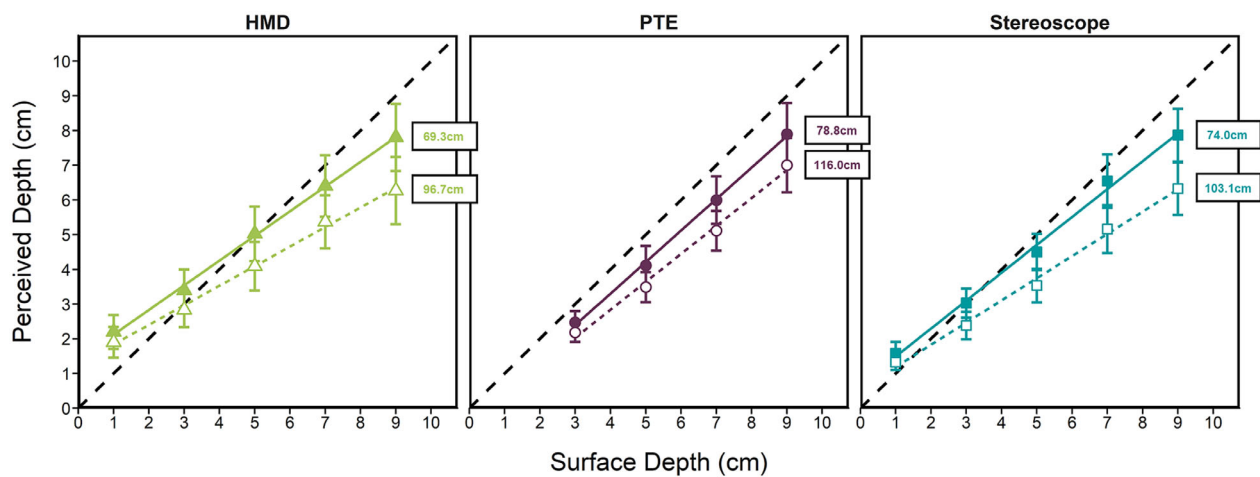

Figure A2. Mean perceived depth estimates as a function of surface depth (in cm) for each apparatus: HMD (triangles), PTE (circles), and stereoscope (squares), for the near and far viewing distances (filled and open points, respectively) under binocular viewing conditions. The inferred viewing distance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

|	Estimate	DF	*t*	*p*	*R*
Monocular					
Depth × apparatus: PTE vs. stereoscope	−0.13	346	−2.61	0.01	0.14
Depth × apparatus: PTE vs. HMD	−0.19	346	−3.78	<0.001	0.20
Depth × apparatus: HMD vs. stereoscope	0.06	346	1.26	0.21	0.07
Depth × viewing distance	−0.09	346	−2.40	0.02	0.13
Depth × viewing distance × apparatus: PTE vs. stereoscope	−0.01	346	−0.32	0.75	0.02
Depth × viewing distance × apparatus: PTE vs. HMD	−0.03	346	−0.62	0.53	0.03
Depth × viewing distance × apparatus: HMD vs. stereoscope	0.01	346	0.37	0.72	0.02
Binocular					
Depth × apparatus: PTE vs. stereoscope	−0.10	346	−1.44	0.15	0.08
Depth × apparatus: PTE vs. HMD	−0.20	346	−2.73	0.01	0.15
Depth × apparatus: HMD vs. stereoscope	0.09	346	1.34	0.18	0.07
Depth × viewing distance	−0.10	346	−2.47	0.01	0.13
Depth × viewing distance × apparatus: PTE vs. stereoscope	−0.06	346	−1.15	0.25	0.06
Depth × viewing distance × apparatus: PTE vs. HMD	−0.04	346	−0.82	0.41	0.04
Depth × viewing distance × apparatus: HMD vs. stereoscope	−0.02	346	−0.38	0.70	0.02

Table A1. The linear mixed-effects analysis independent of condition order. Note: Depth refers to the slope of estimates as a function of predicted depth.

	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 × apparatus: PTE					
Intercept	−0.05	15	−0.20	0.84	0.05
Slope	−0.10	94	−2.18	0.03	0.19
Viewing distance × apparatus: stereoscope					
Intercept	−0.14	15	−0.89	0.39	0.22
Slope	−0.16	126	−4.26	<0.0001	0.35
Viewing distance × apparatus: HMD					
Intercept	−0.14	15	−0.80	0.44	0.20
Slope	−0.15	126	−5.79	<0.0001	0.46

Table A2. The Accuracy and stereoscopic depth constancy analyses independent of condition order. 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.