The Contribution of Monocular and Binocular Depth Cues to Size Discrimination of 3D Objects

Iroshini Gunasekera

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements
for the Degree of Master of Arts

Graduate Program in Psychology
York University
Toronto, Ontario

August 2025

Abstract

The study of perceived size has often been linked to distance, with the role of depth cues, motion and interaction receiving less attention. Here, we evaluated the contribution of depth cues to perceived size during passive viewing and active interaction with 3D shapes in virtual reality. In Experiment 1, observers' precision was similar in both passive and active conditions, and across modality. In Experiment 2, we increased shape complexity and added an object motion condition to make the task more difficult. Results revealed that precision was high under binocular viewing in both conditions and but significantly reduced in the monocular active condition. Movement data indicated that observers moved laterally to obtain depth from motion parallax in the monocular active condition, but this strategy did not improve precision. These findings underscore the importance of binocular depth information in perceiving 3D object size, even when motion or interaction could theoretically enhance size judgements.

Table of Contents

List of Tables

List of Figures

Introduction

The visual system relies on multiple sources of depth information to help us navigate and interact with our environment. One of the most precise sources of depth information is binocular disparity, the difference in the images received by each eye due to their horizontal separation. The brain uses this disparity to compute the relative depth of an object, and its location in space, a process known as stereopsis. When viewing the world binocularly the visual system also has input from the vergence state of the eyes and how it changes with changing fixation which can signal the distance of objects. Depth can be perceived with one eye (monocularly) using cues such as relative size, motion parallax, texture gradients and occlusion. Together, these cues support our perception of 3D structure and spatial relationships (Howard & Rogers, 2012).

Depth cues not only inform spatial layout but also determine our perception of object size. For example, as an object moves further away, its retinal projection becomes smaller; conversely, as it moves closer, its projection enlarges (Gogel, 1969). Despite these changes in retinal size, over a range of distances, our visual system compensates to maintain a constant perceived size, a phenomenon known as size constancy (Holway & Boring, 1941; Mckee & Welch, 1992). Seminal work by Holway & Boring (1941) examined how distance influenced perceived size when visual angle was held constant. With binocular cues (e.g. vergence, stereopsis), apparent size increased linearly with distance. As more depth cues were removed, perceived size depended on visual angle in most conditions. These findings, later replicated by Lichten & Lurie (1950), support the size-distance invariance hypothesis (SDIH), which proposes that perceived size is a function of both visual angle and perceived distance. However, other studies have shown that contextual factors, conflicting depth cues, and perceptual distortions can disrupt the size-distance relationship (Kilpatrick & Ittelson, 1953). Gruber's (1956) size-distance paradox also highlighted that perceived size and distance do not always co-vary in predictable ways.

The conflicting evidence arising from studies of the SDIH raises the question of whether size is derived from distance information or is a distinct perceptual feature. Haber & Levin (2001) conducted an experiment to test this theory. Using both familiar (e.g. a bike, milk bottle) and unfamiliar (flat cutouts of shapes: ovals, rectangles and triangles) stimuli, observers were asked to judge the distance from themselves to the stimuli and estimate the height of each stimulus (to ascertain the size). The results revealed that the perceived size of these familiar objects was unaffected by available distance information. Further, they found that with strong distance information, object size had little impact on perceived distance. From this, they argued that perceived size is separable from perceived distance.

While Haber and Levin (2001) demonstrated that size perception can be separated from distance information, the underlying mechanisms were unclear. One possibility is that the visual system uses depth cues, such as binocular disparity, to infer object size without relying directly on distance estimates. In a foundational study, Gogel (1964) demonstrated that the perceived depth from a fixed binocular disparity can vary depending on the visual context. In this study, when the background made the scene appear farther away, the observers perceived more depth between objects, even though the actual disparity between the objects did not change. This suggests that the brain scales depth based on how large nearby objects appeared relative to their visual angle. Thus, Gogel's work provides early evidence that binocular cues support size perception even when distance cues are unreliable and ambiguous.

Vergence, the inward rotation of the eyes when fixating on a near object, provides an important source of absolute distance in near space. While binocular disparity signals the relative depth between objects, disparity does not inform the visual system of how far the object is relative to the observer. Vergence helps resolve this ambiguity by signalling absolute distance through triangulation. For example, a large vergence angle indicates that the object is closer to the observer; if the angle is small, the object is farther away. This distance estimate is then used to scale retinal image size and disparity to estimate the object's true size (Howard & Rogers, 2012; Mon-Williams & Tresilian, 1999).

Dynamic depth cues can also support size and depth perception. Motion parallax, for instance, arises when an observer moves, causing nearby objects to shift positions on the retina more rapidly than distant objects. This differential motion provides an important cue for estimating depth (Gibson et al., 1959). Although motion parallax is a monocular depth cue, under some conditions it is comparable to stereopsis. Seminal work by Rogers & Graham (1979) compared motion parallax and stereopsis and found that they each were sufficient cues to infer depth when presented in isolation. In this study, observers viewed a random dot pattern (where the movement of the dots was consistent with the presence of different 3D surface shapes) monocularly while moving their head laterally and were asked to match the perceived depth on a stereoscope. The results showed that even with small movements, observers were able to compute depth accurately, with the perceived depth from parallax being comparable to stereopsis. In a follow-up study, Rogers and Graham (1982) investigated depth thresholds for stereopsis and motion parallax, and confirmed that sensitivity to motion parallax closely matched sensitivity to stereopsis. However, stereopsis supports finer depth discrimination with lower thresholds overall. Thus, motion parallax is a reliable and precise depth cue, capable of supporting fine depth discrimination even in situations where stereopsis may be limited or unavailable.

While motion parallax can provide depth information based on the observer’s own movement, transformations in the object itself, such as changes in its motion, can also influence perceived size and shape. One such cue is structure from motion, the process in which the brain infers the three-dimensional structure of an object and its motion, from the two-dimensional motion of the object’s projection onto the retina, even without prior knowledge about the object (Tittle & Braunstein, 1991; Ullman & Brenner, 1997). This integration of motion cues and 3D structure becomes particularly important when static visual information is insufficient or ambiguous. For example, Johnston et al. (1994) explored how disparity and motion are integrated by the brain to perceive 3D shape. They found that when stereopsis and motion were presented together, observers were able to accurately perceive 3D shape. In contrast, when stereopsis and motion cues were presented individually, the perception of shape was distorted, revealing that motion cues in combination with stereopsis, assist with resolving perceptual ambiguities. This maintenance of perceptual stability is further supported by Langdon (1951) who demonstrated that when simple shapes undergo continuous rotational motion, shape perception is preserved even in the absence of supplementary depth cues. This finding highlights that object motion itself can reinforce the perception of a stable 3D structure, if the movement is regular and predictable.

Beyond motion cues, shape perception has also been assessed with stationary objects as a function of size and viewpoint. Norman et al. (2009) examined the accuracy of shape discrimination when objects varied in size and orientation in depth. Their results showed that shape discrimination performance declined as the orientation difference increased (e.g. from 25 degrees to 65 degrees), indicating a strong viewpoint dependency. However, observers were able to consistently judge shape, even when the object size doubled or halved, suggesting that shape perception is largely robust to size changes. This reinforces the idea that stable shape representation supports consistent size perception under varying viewing conditions.

Our interpretation of the 3D structure of a shape may also change based on viewpoint. The slant of the shape is an important part of its 3D structure that can affect our shape perception and in turn, our size estimations. Seminal work by Beck & Gibson (1955) introduced the shape-slant invariance hypothesis, which proposes that the retinal projection of a given object is determined by the relationship between apparent shape to apparent slant. For instance, from a slant of 0 degrees an object may appear to be trapezoidal, but if the object is slanted to 45 degrees, it may appear as a rectangle. To examine how the brain resolves this ambiguity, Beck and Gibson conducted a study where observers were asked to view glowing shapes (rectangular and trapezoidal in nature) in the dark and match them to their physical models. Observers were tested under reduced cue conditions (monocular, no motion, and untextured shapes) and with binocular disparity. The results indicated that with binocular disparity, observers were able to match the shape to the physical model accurately, but under reduced cue conditions, they showed a front-plane bias, consistent with assuming the shapes were not slanted in depth and were instead front-facing. Other studies have shown that errors in slant perception can lead to errors in shape perception. A study by Kaiser (1967) found that when slanted shapes were viewed monocularly, errors in both perceived shape and perceived slant were more likely, while binocular viewing significantly reduced these errors. These results demonstrate that shape perception is closely tied to accurate slant perception; errors in slant can systematically distort perceived shape, which may, in turn, distort perceived object size. Overall. These findings highlight that accurate slant perception is important for maintaining veridical shape and size perception.

Interacting with an object can supplement motion and shape cues by contributing relative information about its size. Harman et al. (1999) investigated whether active manual control of a 3D object's viewing angles enhances recognition compared to passive viewing. Observers were tested in two conditions; an active condition where the object could be manually rotated for 20 seconds, and a passive condition where observers viewed a pre-recorded sequence of rotations made by another observer. The results showed that actively controlling the object allowed for faster recognition and response speeds compared to passive viewing. The importance of active engagement in perception is further supported by Wexler & van Boxtel (2005) who demonstrated that actively manipulating objects (e.g. rotating them manually) enhances shape and size recognition, while active control of viewing angles helps observers maintain a consistent 3D representation. Making active head movements while viewing a scene also leads to more accurate depth judgements compared to passive head movements. These findings emphasize the potential importance of interaction in providing a dynamic understanding of an object's spatial properties and allowing the observer to view the object in different perspectives, all which are important for accurate object size estimation.

From the preceding research, binocular depth cues, motion and interaction all have the potential to influence the perceived size of an object. However, these studies have mostly used 2D stimuli and/or only examined a limited set of cues (e.g. disparity and motion parallax). In the study presented here, we take a more comprehensive approach by examining how depth from stereopsis, object motion, and interaction contribute to judging the size of 3D objects. We used virtual reality (VR) to present stereoscopic 3D stimuli in an interactive environment. In Experiment 1, we used simple 3D shapes and assessed size discrimination monocularly and binocularly when the shape was i) static and ii) when it was moved (by the observer). In

Experiment 2, we increased the complexity of the 3D shapes and evaluated how depth cues influence perceived size when the object is in motion, compared to the effect of self-controlled interaction, while also examining observers' positional data to gain more insight into the strategies they used to make size judgements.

General methods

Observers

Sixty observers (half female) with an age range between 18 to 30 years, participated in the experiments. All met the stereoacuity criteria of at least 40 arc seconds, assessed using the RANDOT$^{\text{TM}}$ Preschool Stereoacuity Test and the RANDOT$^{\text{TM}}$ Stereo Test, and self-reported normal or corrected-to-normal visual acuity. Interocular distance (IOD), which is the distance between the center of the pupils, was measured for each participant using a GR-4 digital pupillometer. The study was approved by the York University Ethics Board.

Stimuli

The stimuli were virtual renditions of a popular children's shape-posting toy, with the virtual container and shapes resembling their real-world counterparts, including wood texture and colour (Figure 1).

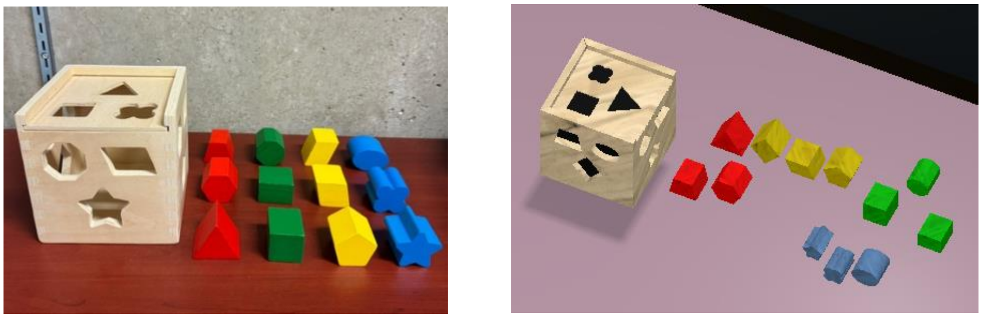


Figure 1. Real shape posting toy (left) and its virtual rendition (right)

The container and shapes were created in Blender (version 4.1). Four of the original twelve shapes; triangle, square, pentagon, and quatrefoil were chosen. The shapes were selected based on their symmetry and number of sides, with the quatrefoil representing a more 'complex' geometrical shape, i.e., rounded sides. The real shapes and container were measured to obtain their dimensions and the virtual reference shapes were rendered with matching dimensions (Table 1).

	Width (m)	Length (m)	Height (m)
Container	0.14m	0.14m	0.14m
Triangle	0.045m	0.039m	0.039m
Square	0.023m	0.023m	0.040m
Pentagon	0.040m	0.040m	0.038m
Quatrefoil	0.036m	0.039m	0.036m

Table 1. Dimensions of the reference shapes and the container used in the experiments.

We systematically manipulated the size of the reference shapes to create three smaller and three larger versions, resulting in seven different sizes per shape. Using the method of constant stimuli, the shape size was varied across trials. Step sizes ranged from ±1.5% (narrowest) to ±5.5% (largest). The sizes were randomised during testing, and each shape - size combination was presented four times (Figure 2).

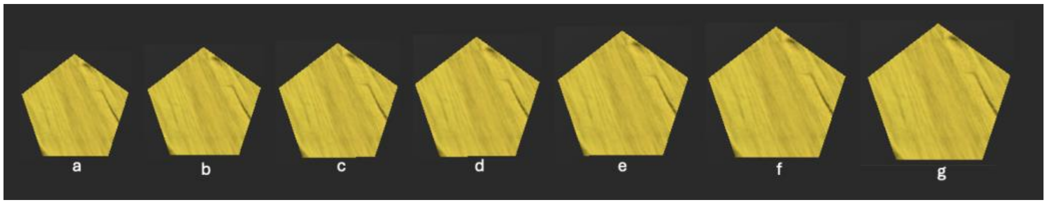


Figure 2. Sample test stimuli with a 4% change in size from the reference shape (d).

The test environment was a virtual room with a desk that appeared in front of the observer against a black background. The environment was world-locked, i.e., the head was tracked, and the virtual objects remained fixed in space while the observer moved relative to the world. Two wooden platforms were placed next to each other on the desk at a fixed distance from the observer. The heights of the platforms were set to the observer's eye level. The shapes appeared on the top of the right platform, and the front surface of the container changed on each trial to display the side with the slot that matched the shape presented (Figure 3).

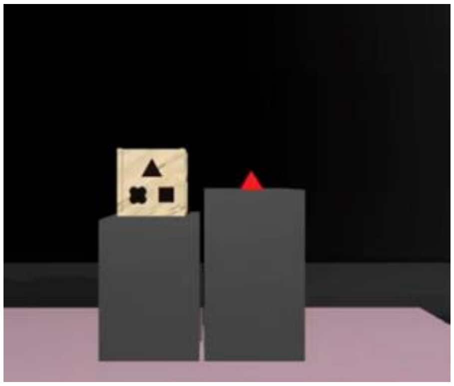


Figure 3. Observer's view of the virtual environment and the stimuli.

Apparatus

All 3D Blender models were exported to the Unity game engine (version 3.8). The experiments were designed in Unity and were conducted on a Meta Quest Pro standalone head-mounted display (HMD) with 1920 by 1800 pixels per eye resolution and a 90 Hz refresh rate.

General Procedure

After informed consent was obtained, observers were screened for stereoacuity, and their IODs were measured. They were then instructed on the experimental procedure and how to use the controllers. The experimenter manually adjusted the lenses on the headset to match the observer's IOD. Observers were asked to judge the size of the target shape relative to the matching slot on the container. They pressed the left trigger if the shape appeared smaller than the slot, and the right trigger if it appeared larger. Before the main experiment, each observer completed a preliminary session consisting of 28 trials. This gave them an opportunity to become accustomed to the virtual environment, use the controllers, and become comfortable with responding. The results of this preliminary session were also used to determine each observer's step size.

All experiments included two manipulation conditions (interaction vs no interaction), presented under two viewing modalities (binocular and monocular). Observer's responses and reaction times were recorded on every trial. To control for order effects, the study used a counterbalanced design. Observers were first assigned to complete either the no-interaction condition followed by the interaction condition, or the interaction condition followed by the no interaction condition. Within each task block, the order of viewing modality was also counterbalanced: Half the observers completed the binocular trials before the monocular trials, while the other half completed the monocular trials before the binocular trials (Table 2). Each manipulation condition consisted of 224 trials: 112 per viewing modality (7 shape sizes x 4 shapes x 4 repetitions per shape-size combination). For monocular viewing, observers were asked to choose their preferred eye, and the non-preferred eye was occluded with an eye patch. Each manipulation condition was tested on separate days, and after each block of 112 trials, observers were allowed to take a 5-minute break.

	Manipulation condition	Modality	
Group 1	No interaction first	Binocular/Monocular first	Monocular/Binocular second
	Interaction second	Binocular/Monocular first	Monocular/Binocular second
Group 2	Interaction first	Binocular/Monocular first	Monocular/Binocular second
	No interaction second	Binocular/Monocular first	Monocular/Binocular second

Table 2. Experimental conditions and viewing modality

Data Analysis

For each condition, the proportion of "larger" responses was plotted and fit using a cumulative normal distribution. Just-noticeable differences (JNDs) were computed using the individual data by taking the difference between the values corresponding to 75% and 50% "larger" responses. The point of subjective equality (PSE) was determined as well; that is, the size at which observers perceive the test shape as equal to the reference shape (the 50% point).

Experiment 1

Introduction

The first experiment assessed the precision and accuracy of size judgements for a set of 3D shapes presented in VR. We predicted that size judgements would be most precise when stereopsis was available in the binocular viewing conditions, as stereopsis provides accurate depth information that should stabilize size perception (Hypothesis 1.1). We also expected that performance would improve under monocular viewing when observers could interact with and move the shapes (Hypothesis 1.2).

Observers

Thirty observers (half female) participated in Experiment 1. All observers met the screening criteria described in the General Methods section.

Stimuli

As described in the General Methods, the stimuli were a triangle, square, pentagon, and quatrefoil (Figure 4).

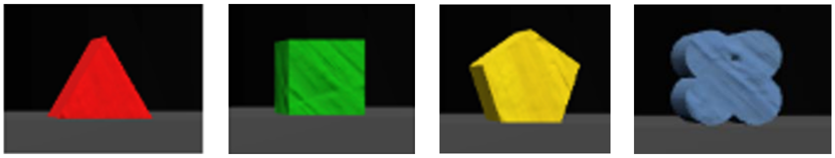


Figure 4. Shapes used in experiment 1. From left to right – triangle, square, pentagon, and quatrefoil.

Preliminary sessions were used to calculate the proportion of responses in which the observer responded “larger”, and a psychometric curve was fit. Based on the overall fit of the curve, an ideal ‘step size’ was assigned for each observer, and the test values were centered around the true size of the object. The narrowest range was ±2.5% the size of the reference shape, while the broadest range was ±4.5%. The shapes and the container appeared at a fixed distance of 0.8 m from the observer, and each shape appeared on a platform in front of the observer.

Procedure

All observers were tested in two manipulation conditions: passive (no interaction) and active (interaction) (Figure 5). In the passive condition, observers viewed the shape for 2.5 seconds, made a response, and received an auditory cue confirming that the response was recorded. An inter-trial interval of 2 seconds followed. In the active condition, observers picked up the shape via the grip button using a virtual pointer that extended from the controller. Once the shape was picked up, the observer could move it freely. To prevent direct matching to the slot, a 0.43 m invisible boundary was implemented. If the shape crossed this boundary, it disappeared and reappeared in its original position on the platform. Trial duration was 5 seconds, followed by a 2-second pause. For both manipulation conditions, the subsequent trial began after the observer responded. In the event that there was no response, an upper limit of 300 seconds was used, after which the experiment moved to the next trial.

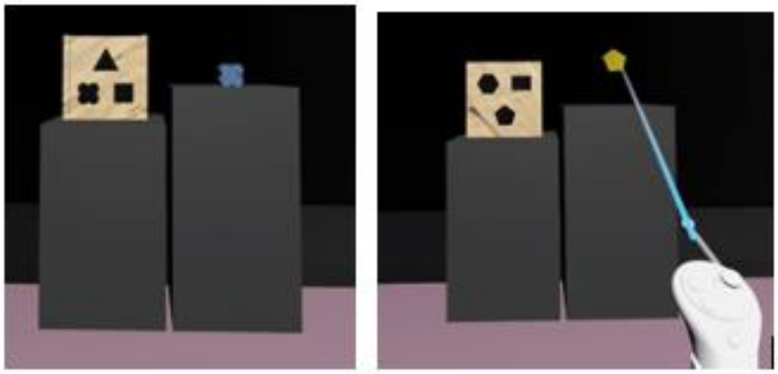


Figure 5. No-interaction (passive) condition (left) and interaction (active) condition (right)

Results

Twenty-eight observers' data were included in the analysis. Two observers were excluded as their data showed extreme response biases, and their psychometric functions were flat.

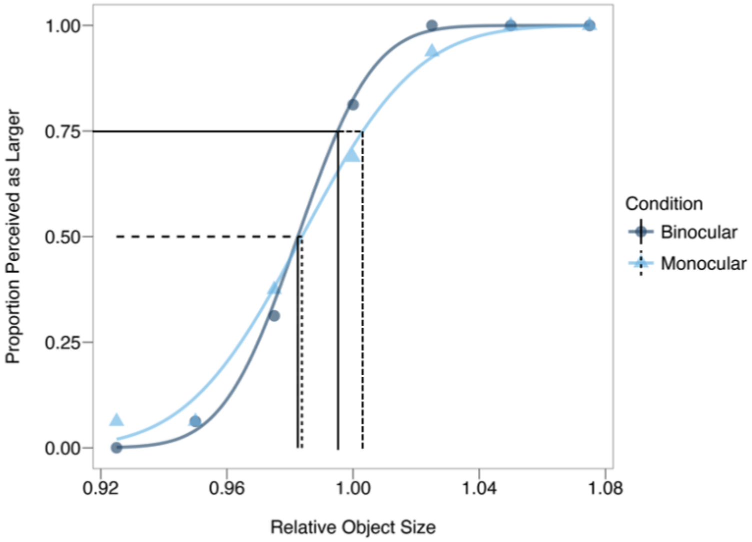


Figure 6. One observer's psychometric function from Experiment 1 for illustration. The x-axis represents the size of the comparison shape expressed as a proportion of the reference shape. A value of 1.0 indicates that the comparison and reference shapes were identical. Values below 1.0 indicate the comparison shape was smaller than the reference, and values above 1.0 indicate that

the comparison shape was physically larger. The y-axis shows the proportion of trials in which observers judged the comparison shape as ‘larger’ than the reference. The dark blue curve represents the psychometric function under binocular viewing, and the light blue curve represents monocular viewing. The vertical solid (binocular) and dashed (monocular) lines extending from the 75% point represents the observer’s JNDs. The vertical solid (binocular) and dashed (monocular) lines extending from the 50% point represents the observer’s PSEs.

The average JND and PSEs collapsed across shape are shown in Figure 7. JNDs were lower under binocular viewing compared to monocular viewing, indicating that observers were more precise in their size judgements when binocular depth cues were available.

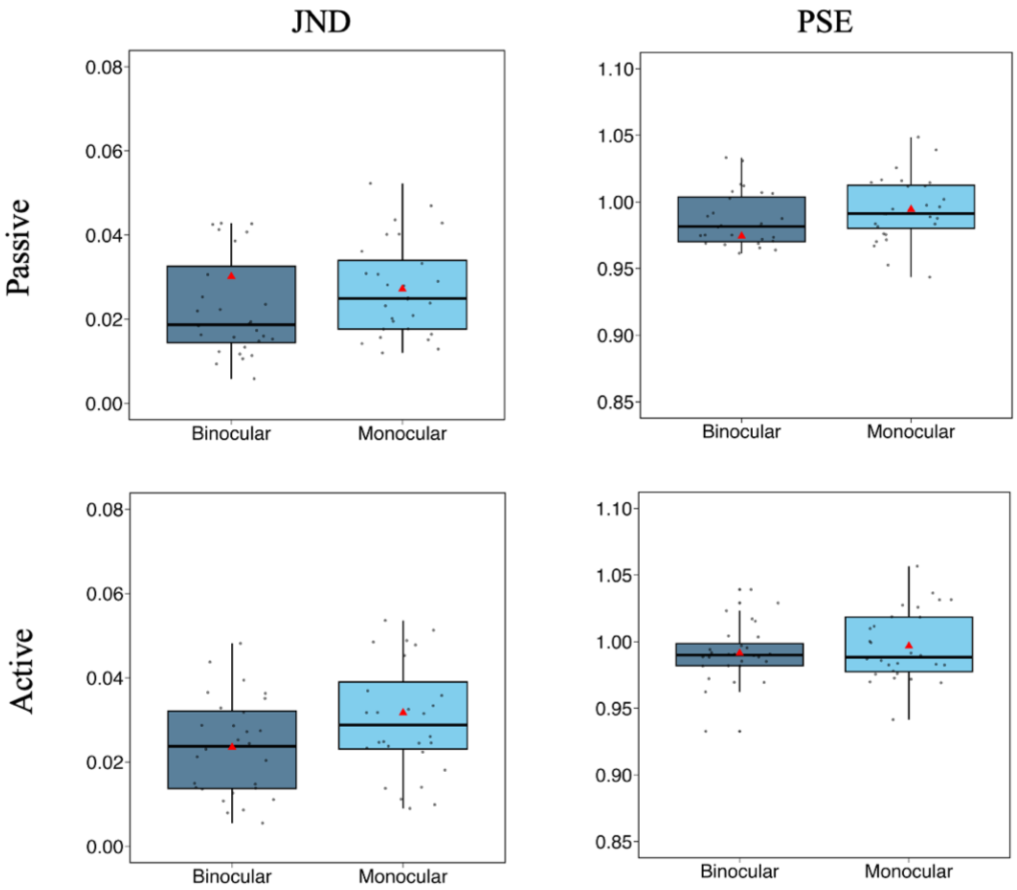


Figure 7. Boxplots depict the average JND (left), and PSE (right) collapsed across shape for the passive (top row) and active (bottom row) conditions. The y-axis for JND shows the relative size difference (expressed as a proportion of the reference size), and the y-axis for PSE shows the perceived size of the test stimulus relative to the reference size (1.0 = veridical). Dark blue represents binocular, and light blue represents monocular conditions. The boxes show the interquartile range, with the horizontal line indicating the median and the red triangle indicating the mean. Whiskers extend to the most extreme values within 1.5 x IQR from the first and third quartiles. Black dots signify individual observer means.

A three-way repeated-measures ANOVA was conducted on the JND values using the within-subjects' factors of Modality (binocular, monocular), Condition (passive, active) and

Shape (pentagon, quatrefoil, square, triangle). Mauchly's test indicated violations of sphericity for main effects and interactions involving Shape ($p < .001$), Modality x Shape ($p < .001$), Condition x Shape ($p < .001$), and Modality x Condition x Shape ($p < .001$). Therefore, Greenhouse-Geisser corrections were applied. There were no significant main effects of Modality (F (1, 27) = 0.77, $p = .389$, partial $\eta^2 = .03$), or Condition (F (1, 27) = 0.03, $p = .871$, partial $\eta^2 < .001$). There was also no main effect of Shape after the Greenhouse-Geisser correction (F (1.38, 37.2) = 1.99, $p = .161$, partial $\eta^2 = .07$). No significant two-way or three-way interactions were observed.

A separate three-way repeated-measures ANOVA was conducted on the PSE values using the same within-subjects factors. Mauchly's test indicated that the assumption of sphericity was violated for Shape ($p < .001$), Modality x Shape ($p < .001$), Condition x Shape ($p < .001$), and Modality x Condition x Shape ($p < .001$). Greenhouse-Geisser corrections were therefore applied to all the shape-related effects, and after the correction there remained a significant main effect of Shape (F (1.87, 50.5) = 9.88, $p < .001$). Modality approached significance (F (1, 27) = 3.66, p =.066, partial $\eta^2 = .12$). There was no significant main effect of Condition (F (1, 27) = 1.59, p =.219, partial $\eta^2 = .06$). There were also no significant two-way or three-way interactions. Follow-up pairwise comparisons using Tukey's-adjusted estimated marginal means revealed that under monocular viewing, the pentagon was perceived as significantly smaller under binocular ($\eta^2 = .25$) and monocular ($\eta^2 = .69$) viewing and in both passive ($\eta^2 = .23$) and active ($\eta^2 = .71$) conditions, relative to the other shapes (see Figure 8).

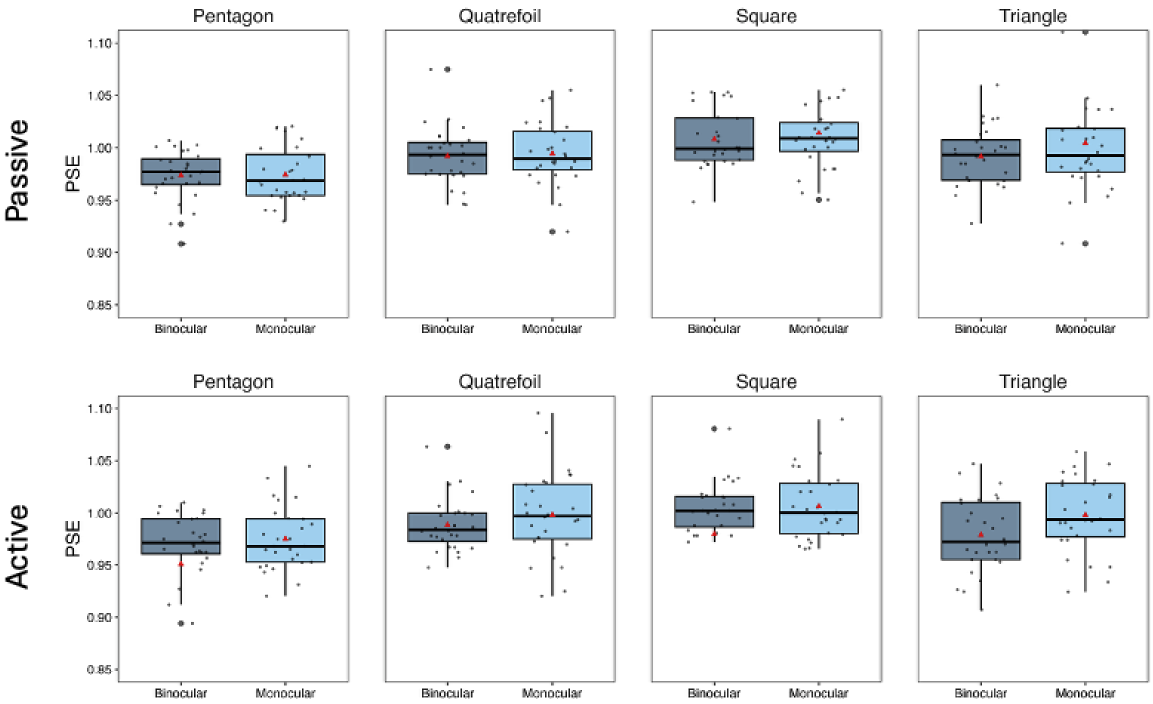


Figure 8. Boxplots depict the average PSEs for each shape (Pentagon, Quatrefoil, Square and Triangle) for the passive (top row) and active (bottom row) conditions. Dark blue represents binocular viewing, and light blue represents monocular viewing. The boxes show the interquartile range, with the horizontal line indicating the median and the red triangle indicating the mean. Whiskers extend to the most extreme values within 1.5 x IQR from the first and third quartiles. Black dots signify individual observer means.

Experiment 1 Discussion

Although we predicted greater precision under binocular viewing due to the availability of binocular depth information, precision did not differ between modalities (failing to support Hypothesis 1.1). Moreover, interaction with the stimuli did not improve precision (failing to support Hypothesis 1.2). The lack of an effect of viewing modality and its interaction in this study was somewhat surprising, as we initially hypothesized that precision would be higher when observers were allowed to interact with the stimuli. The fact that thresholds were quite low and very consistent, suggests that the task was too easy. In Experiment 2 we address this by increasing the complexity of the stimuli and task overall.

The significant effect of shape on PSE was driven by the pentagon results. Specifically, the pentagon was consistently perceived as smaller than the other shapes across modality and in both manipulation conditions. It is unclear why this was the case, perhaps certain geometric properties of the pentagon, such as the symmetry or having more sides than the other shapes, may have contributed to this effect. We return to this possibility in the General Discussion, where we consider potential perceptual factors that may underlie this bias.

Experiment 2

Introduction

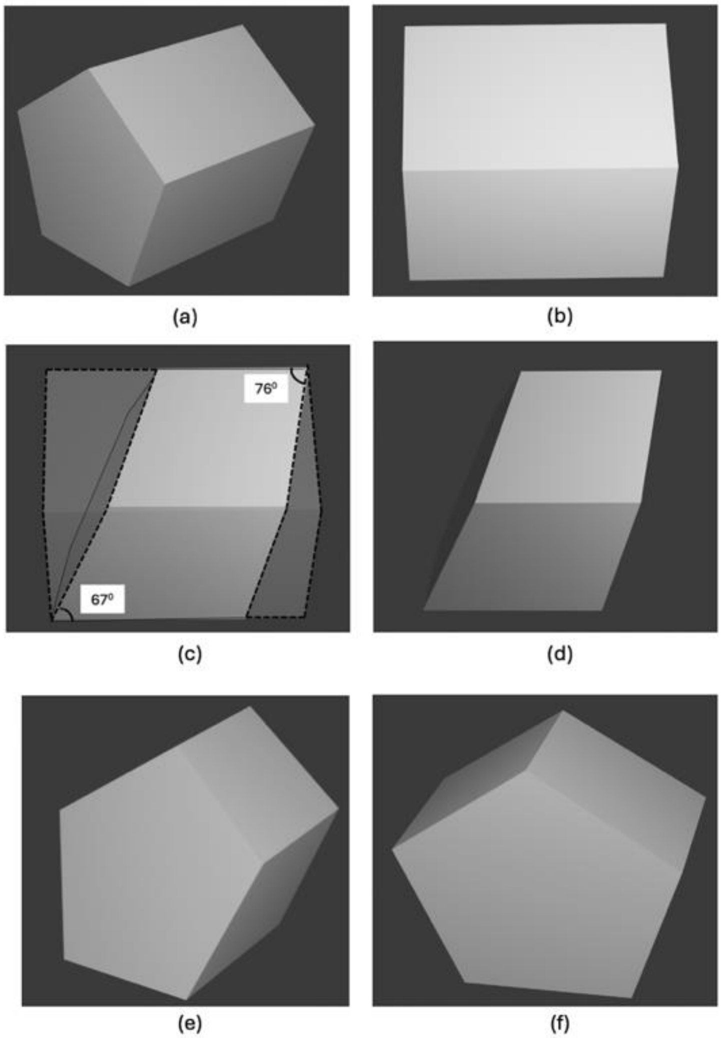
In Experiment 1, we found that observers were equally precise in their size judgements under binocular and monocular views; and the opportunity to interact with the stimuli had little effect. We hypothesized that this may have been because the task and stimuli were overly simple. Based on these findings, we increased task difficulty by altering the shapes to be more complex and ambiguous, and eliminated the different colors. Following these modifications, Experiment 1 was replicated with a new no-interaction condition: we introduced a rotating shape to evaluate the effect of structure from motion cues. The interaction condition closely resembled that used in Experiment 1, with minor modifications. We hypothesized that due to the increased difficulty of the task, the availability of stereopsis under binocular viewing would improve precision relative to monocular viewing (Hypothesis 2.1). Additionally, we predicted that interaction would enhance precision, particularly under monocular viewing, as interacting with a more complex and ambiguous shape should provide access to motion-based depth cues (Hypothesis 2.2).

Observers

Twenty-eight observers (nineteen female) participated in Experiment 2. All observers met the criteria described in the General Methods section.

Stimuli

The same shapes from Experiment 1 were used with several modifications designed to increase task difficulty and perceptual ambiguity. Each shape was sliced diagonally from both sides: one at an angle of 23 degrees and the other at 14 degrees – producing asymmetrical faces of different sizes (Figure 9). Additionally, the shapes were rendered with diffuse lighting to avoid shading and highlights that would lead to immediate recognition of the shape at the start of the trial.



(a) (b) (c) (d) (e) (f)

Figure 9. Image of the pentagon shape to illustrate the effect of slicing, prior to the addition of the texture. a) original stimulus, b) rotated version, side-view, c) angle of the slices, d) side-view of sliced shape, e) larger face, f) smaller face. All shapes were sliced using the same parameters.

As in Experiment 1, preliminary sessions were used to determine the step size for each observer. The narrowest range was ±2.5% the scale of the reference shape, while at the broadest, the range was ±4.5%. The shapes and the container appeared at a fixed distance of 0.6 m from the observer. We also positioned them closer as observers in Experiment 1 reported that they had to reach further than expected to interact with the shape. The shapes appeared on the platform with their narrow edge facing the observer to further increase task difficulty (Figure 10).

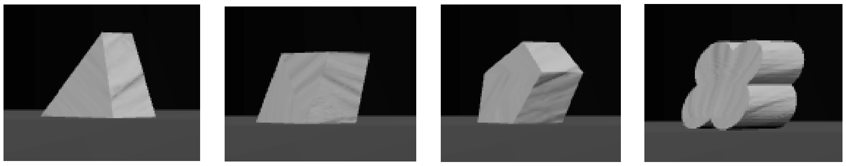


Figure 10. Shapes used in Experiment 2. From left to right: triangle, square, pentagon, and quatrefoil.

Procedure

Each observer was tested in two manipulation conditions: passive (no interaction) and active (interaction) (Figure 11). In the passive condition, at the beginning of each trial, the shape appeared with its side towards the observer (rotated at 90 degrees). After a 0.5-second delay, the shape began to oscillate, rotating 90 degrees clockwise to face the observer, then rotating 180 degrees counterclockwise at a speed of 45 degrees per second. Within a 10-second trial, the shape completed a maximum of 3.5 oscillation cycles (rotated back and forth for a total of three and a half turns). Observers could pause the rotation as often as they wanted during the trial. To pause and play the rotation, observers pressed “A” on the right controller. To prevent observers from speeding through the blocks without viewing the shapes, and to standardize viewing time with the interaction condition, the trial did not proceed even if a response was made before 5 seconds elapsed. There was an inter-trial interval of 2 seconds. The active condition was similar to that described in Experiment 1, with the only differences being a shorter virtual rod used to pick up the shape, and trial length was increased to 10 seconds. For both manipulation conditions, if a response was not recorded, the experiment paused for up to 300 seconds. The subsequent trial only began after a response was recorded or if 300 seconds had elapsed. Also, similar to Experiment 1, the manipulation conditions were counterbalanced (see Table 2).

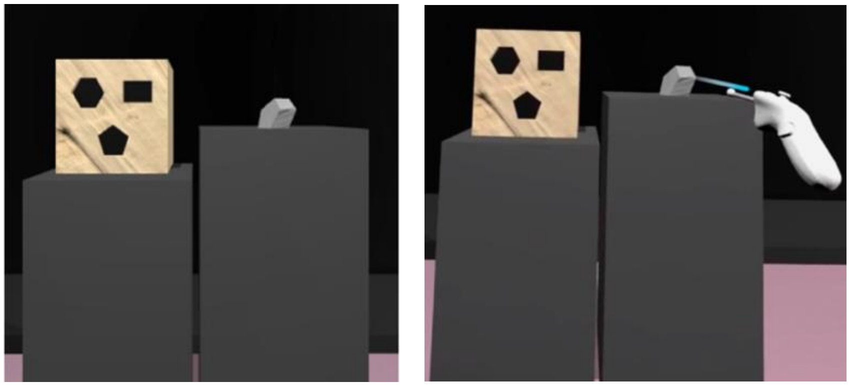


Figure 11: No-interaction (passive) condition (left) and interaction (active) condition (right).

Head motion tracking

Motion tracking data were collected to determine if observers used specific movement strategies while making their judgements. Due to technical issues, positional data from eleven of the twenty-eight observers were unavailable, resulting in seventeen sets of tracking data. Head and object tracking data was obtained from the HMD's built-in tracking system, which used sixteen cameras to track the head and controllers, with up to six degrees of freedom, and a field of view of 106^0 horizontal and 95.57^0 diagonal (Meta, n.d.). Head and object movement data was sampled at 0.5 second intervals.

Results

Experimental results

Twenty-six observer's data were included in the analysis. Two observers were excluded as their psychometric functions were flat and showed extreme response biases. JND and PSEs were calculated as described in the General methods section.

The average JND and PSEs collapsed across shape are shown in Figure 12. JNDs were lower under binocular viewing compared to monocular viewing which showed that observers were more precise in their size judgements when binocular depth cues were available. JNDs were also lower in the passive condition compared to the active condition when viewing monocularly.

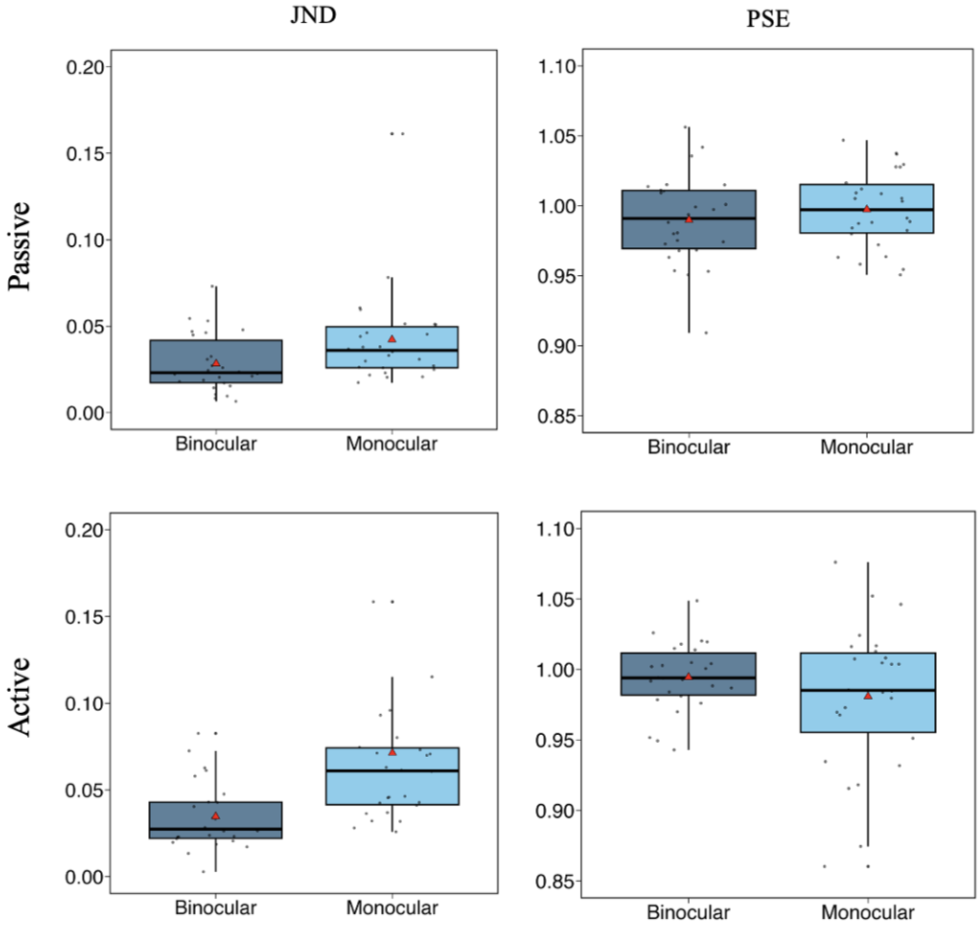


Figure 12. Boxplots depict the average JND (left), and PSE (right) collapsed across shape for the passive (top row) and active (bottom row) conditions. The y-axis for JND shows the relative size difference (expressed as a proportion of the reference size), and the y-axis for PSE shows the perceived size of the test stimulus relative to the reference size (1.0 = veridical). Dark blue represents binocular viewing, and light blue represents monocular viewing. The boxes show the interquartile range, with the horizontal line indicating the median and the red triangle indicating the mean. Whiskers extend to the most extreme values within 1.5 x IQR from the first and third quartiles. Black dots signify individual observer means.

A three-way repeated-measures ANOVA was conducted on the JND values using the within-subjects' factors of Modality (binocular, monocular), Condition (active, passive), and Shape (Pentagon, Quatrefoil, Square, Triangle). Mauchly's test indicated violations of sphericity for the main effect of Shape ($p < .001$), and all interactions involving Shape ($p < .001$) and Greenhouse-Geisser corrections were applied. There was a significant main effect of Modality (F (1, 25) = 25.33, $p < .001$, $\eta^2 = .50$), with lower JNDs (higher precision) under binocular view (M = 0.0323, SE = 0.0028) compared to the monocular view (M = 0.0561, SE = 0.0051). There was also a significant main effect of Condition (F (1, 25) = 5.69, $p = .025$, $\eta^2 = .19$), with lower JNDs in the passive trials (M = 0.0361, SE = 0.00346) than in the active trials (M = 0.0522, SE = 0.00538). There was no significant main effect of Shape (F (2.02, 50.48) = 0.94, $p = .40$, η^2 = .04). No significant two-way or three-way interactions were observed ($p > .05$).

A separate three-way repeated-measures ANOVA was conducted on PSE values using the same within-subjects factors. Mauchly's test indicated that the assumption of sphericity was violated for Shape and all shape-related interactions ($p < .001$). Therefore, Greenhouse-Geisser corrections were applied. There was a significant main effect of Shape on PSE (F (4, 100) = 15.34, p <.001, $\eta^2 = .38$). No significant main effects of Modality, Condition or any interactions were found ($p > .48$). Pairwise comparisons (Tukey adjusted) showed that the Square was perceived as significantly larger than other shapes under both viewing modalities and condition ($p < .001$) (Figure 13).

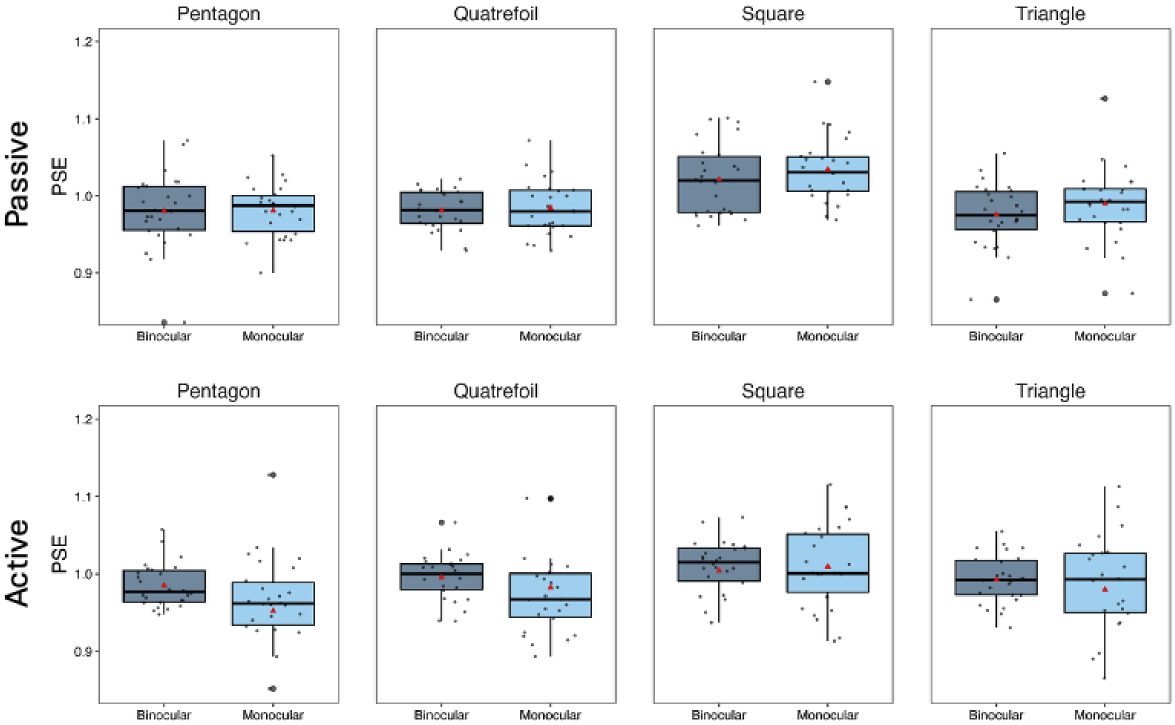
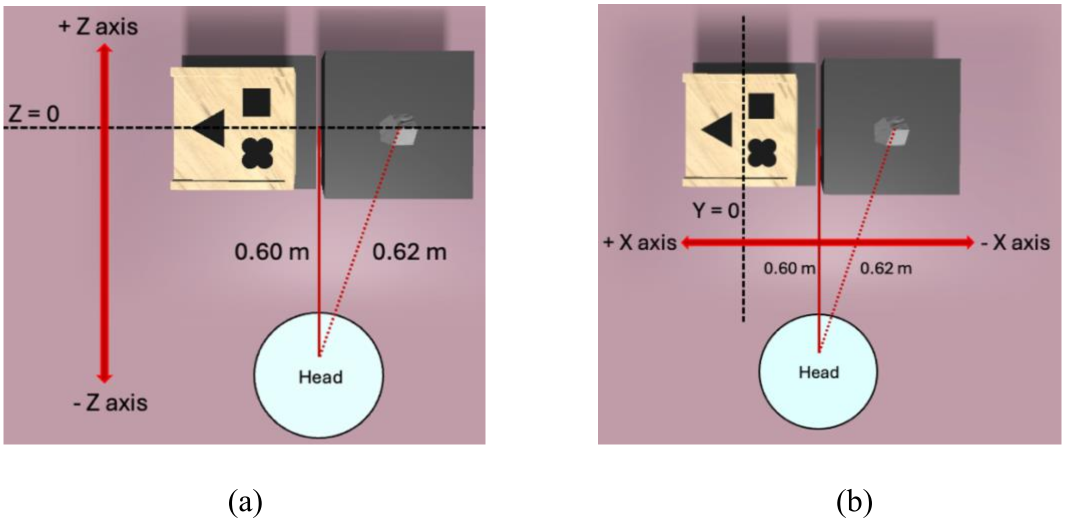


Figure 13. Boxplots showing average PSEs for each shape (Pentagon, Quatrefoil, Square and Triangle). Passive (top row) and active (bottom row) conditions. The graphing conventions are the same as Figure 8.

Head motion tracking results

We also examined the head motion data to gain insight into other sources of depth information, such as motion parallax, that observers may have relied on or attempted to use, particularly in the monocular active condition to compensate for the lack of binocular depth information. We computed movement in depth (z-axis) and lateral movement (x-axis) in 3D space using a cartesian coordinate system (Figure 14).



(a) (b)

Figure 14. Top-down view of the experimental setup (not to scale). The images show the position of the observer's head (white circle) relative to the container (left) and the shape (right). The platforms and the shape were positioned 0.6 m and 0.62 m respectively, from the observer along the z-axis. (a) image shows movement in depth on the z-axis. In the coordinate system, + and – indicate more distant and nearer location respectively. The more negative the z-axis value, the more the observer leaned back. (b) image shows lateral movement on the x-axis. In the coordinate system, + and – indicate more distant and nearer location respectively. The more negative the x-axis value, the more to the right the observer moved.

Head movement in depth

A linear mixed-effects model was conducted on total depth movement with Modality (binocular vs monocular), Condition (passive vs active), and their interaction entered as fixed effects, and random intercepts for observers. The model revealed a significant main effect of Condition, $b = -0.147$, $SE = 0.007$, $t(7581) = -20.45$, $p < .001$, indicating that movement in depth was reduced during the passive condition relative to the active condition. The main effect of Modality was not significant, $b = 0.009$, $SE = 0.007$, $t(7581) = 1.31$, $p = .188$. However, the Modality x Condition interaction was significant, , $b = 0.075$, $SE = 0.010$, $t(7581) = 7.32$, $p < .001$, confirming that observers moved more in the active condition (Figure 15).

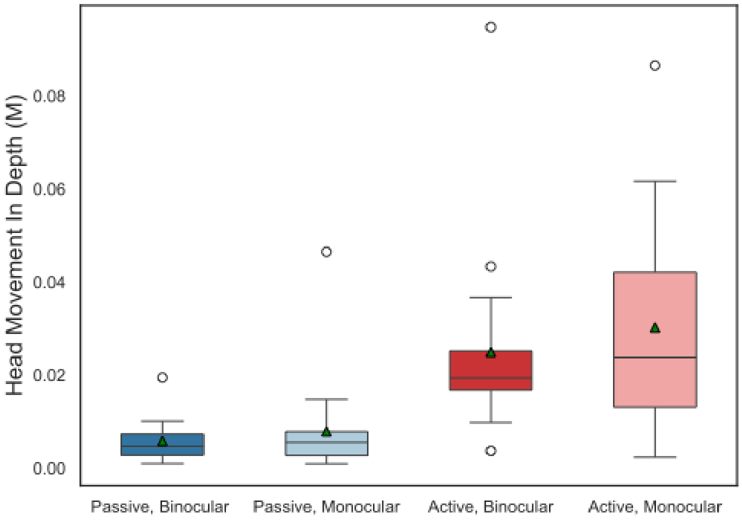


Figure 15. Total head movement in depth across all trials. The passive conditions are shown in blue and active in red. Darker and lighter colors indicate binocular and monocular viewing respectively. Green triangles indicate the group mean, horizontal black line shows the median and black circles represent outliers.

Lateral head movement

Increased lateral movement was expected to enhance access to motion-based depth cues, such as motion parallax, which should theoretically improve precision, particularly under monocular viewing (Rogers and Graham, 1979).

A linear mixed-effects model was conducted to examine the effects of Modality (binocular vs monocular) and Condition (active vs passive) on total lateral head movement. The model revealed significant main effects of Modality, $b = 0.06$, $SE = 0.02$, $t(7581) = 2.82$, $p = .0048$, and Condition, $b = -0.20$, $SE = 0.02$, $t(7581) = -10.04$, $p < .001$. Modality x Condition interaction was also significant, $b = 0.15$, $SE = 0.03$, $t(7581) = 5.18$, $p < .001$. Post hoc comparisons of estimated marginal means indicated that observers moved more laterally under monocular viewing compared to binocular viewing in both the active ($M = -0.056$, $SE = 0.02$, $z = -10.15$, $p = .005$) and passive ($M = -0.20$, $SE = 0.02$, $z = -10.15$, $p < .001$) conditions. In addition, lateral movement was greater in the active condition than the passive condition across modality (binocular ($M = 0.20$, $SE = 0.02$, $z = 10.041$, $p <.001$) and monocular ($M = 0.05$, $SE = 0.02$, $z = 2.71$, $p = 0.006$)) (Figure 16).

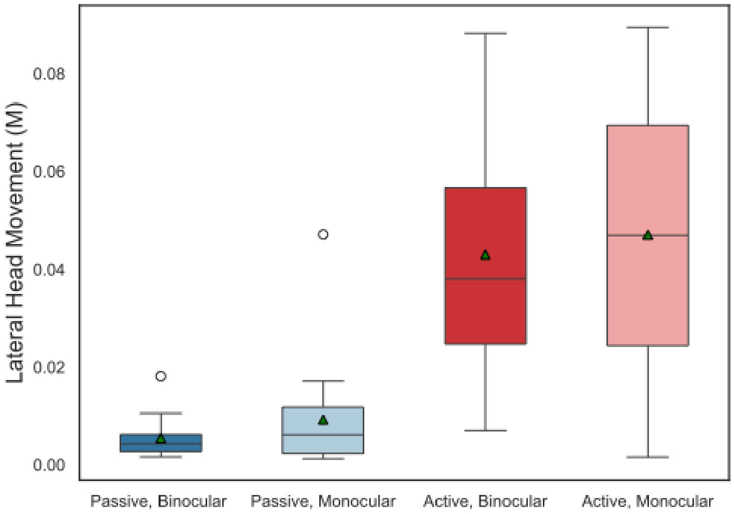


Figure 16. Total lateral head movement across all trials. Graph conventions are the same as Figure 15.

Face-choice data

As mentioned in the Stimuli section, the shapes used in this experiment featured two distinct surfaces: a larger face and a smaller face (Figure 17). Because observers may have made their responses while viewing either face, we examined whether responses tended to occur when the larger or smaller face was oriented towards them. Discerning the object size required that observers consider the degree of the slant. Although both surfaces were slanted in depth, the shallower slant of the large face would have made the object look larger if the 3D form was not perceived. Without binocular disparity, the visual system lacks reliable metric depth information, making slant is harder to interpret accurately. As a result, perceived surface area may be underestimated leading to systematic errors in judging object size in the monocular condition. Thus, if the surface slant played a role, we predicted better performance under binocular viewing, particularly when the larger face was visible.

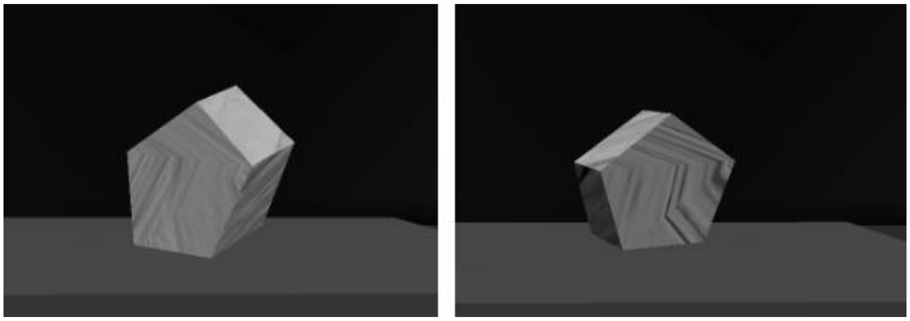


Figure 17. Larger sliced face (left) and smaller sliced face (right) of the pentagon.

The proportion of responses made when viewing the larger or smaller faces for both the passive and active tasks is shown in Figure 18. In the passive condition, observers responded significantly more often when the larger face was oriented toward them. This finding was supported by a binary logistic regression model, in which the intercept was significant β = 0.85, *SE* = 0.07, *p* < .001. There were no significant effects of Modality or Shape. In the active condition, the distribution of responses did not differ between the large and small faces. There were no significant differences in face orientation as a function of viewing Modality (*p* = .374) or Shape (*p* > .22), suggesting that the observers did not prefer a specific viewpoint, when they could interact with it.

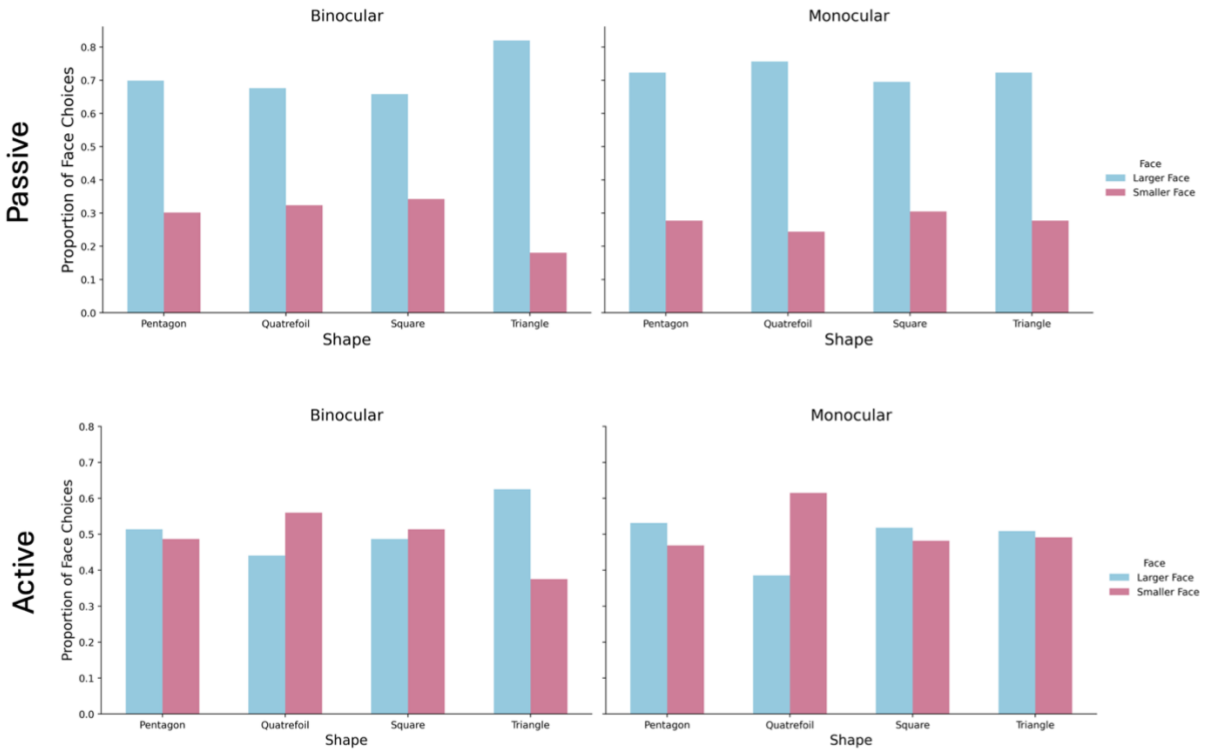


Figure 18. Proportion of responses made when the shape was oriented toward the larger (blue) or smaller face (pink) for the passive (top) and active (bottom) conditions.

Response time

Although observers were not told to make their responses rapidly, we evaluated the time taken to make a response for all conditions and modalities (passive vs active and binocular and monocular). The data showed that observers had shorter response times (< 5 seconds) under binocular viewing, irrespective of interaction (Figure 19). A three-way ANOVA was conducted to examine the effects of Modality, Condition, and Shape on response time. The results revealed significant main effects of Modality (F (1, 7600) = 82.47, p < .001), and Condition (F (1, 7600) = 514.99. p < .001), confirming that in the monocular active condition observers responded significantly slower, M = 5.27, SE = .07, p < .001.

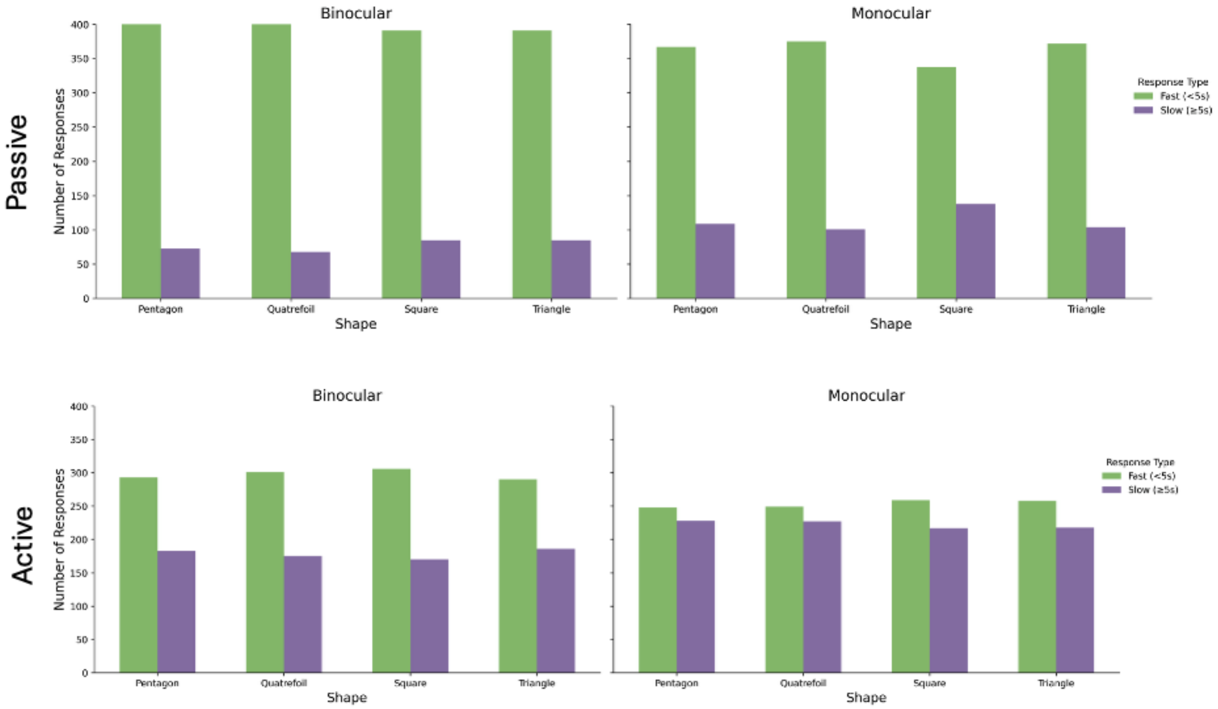


Figure 19. Number of responses made in the passive (top) and active (bottom) conditions. Responses made < 5 seconds are indicated in green, while responses beyond > 5 seconds are indicated in purple.

Experiment 2 Discussion

Observers' size judgements were more precise under binocular viewing in both the passive and active conditions, supporting Hypothesis 2.1. This is likely due to the availability of binocular depth cues in combination with motion-based depth cues, such as structure from motion (which was accessible under both viewing conditions). In contrast, precision was lower under monocular viewing, particularly in the active condition, consistent with the proposal that binocular depth information enhances precise size discrimination. Surprisingly, active interaction with the objects did not improve size estimation under monocular viewing but instead decreased precision. As discussed in the General Discussion, it is possible that this reflects aspects of the control interface, and/or observers' lack of experience.

The head movement tracking data are consistent with these findings: under monocular viewing, observers exhibited increased front-and-back head movement in the monocular active condition; although this is unsurprising as observers needed to lean forward to pick up the shape. Observers' lateral head movement also increased in the monocular active condition, possibly indicating a reliance on motion parallax to extract additional depth information for size judgements.

In the passive condition, observers often responded as soon as the shape rotated to reveal the larger face. Since all trials began with the larger face appearing first, this shows that observers did not wait to view the smaller face before responding. Response times support this observation, with most responses occurring within the first five seconds. This may partly explain the differences in precision between binocular and monocular viewing. That is, in both conditions observers responded when the larger surface was visible, a task that would be more difficult in the monocular condition (without binocular disparity). In contrast, in the active condition, the proportion of responses made when viewing the small and large surfaces was similar, indicating that there was no apparent preference.

As in Experiment 1, we found a shape bias in Experiment 2. In this case, it was the square that was perceived as larger. Subsequent review of the stimuli revealed that there was one lighting/orientation combination in which the larger face may have appeared to blend with the surrounding facets, enhancing its apparent size. Under these conditions the square would have appeared larger, biasing observers' size judgements.

General discussion

The goal of this thesis was to examine how different sources of depth information affect the perceived size of 3D shapes. Binocular depth cues have been shown to be a sufficient cue for informing accurate size discrimination (Gogel, 1964). However, previous studies have typically focussed on how perceived size correlates with perceived distance (Holway & Boring, 1941). Further, these studies often used 2D stimuli and did not consider the roles that stereopsis, motion, and interaction might play in the perceived size of 3D stimuli. In Experiment 1, we used 3D shapes to assess perceived size as a function of viewing modality along with the ability to interact with the stimuli. There was no difference in JND or PSEs between modalities or manipulation conditions, suggesting that the task and stimuli were too simple. Based on these results, in Experiment 2, we made the stimuli more ambiguous by slicing the shapes diagonally from both ends. We also introduced a structure from motion condition where the shape rotated in place.

These modifications were successful in increasing task difficulty as performance showed meaningful differences in JND and PSEs in both manipulation conditions and modalities for Experiment 2. Observers were more precise under binocular viewing compared to monocular viewing for both passive and active conditions. Interestingly, performance was poorest in the monocular active condition, where observers had to rely on motion-based cues in the absence of binocular depth information. When we examined the motion tracking data, it was clear that observers moved themselves more in the monocular active condition. The fact that performance remained poor despite the increased movement, highlights the limitations of motion cues alone for supporting fine size discriminations, and potentially points to increased task demands. In sum, these results suggest that stereopsis provides sufficient depth information to support precise size perception in our environment. In addition to binocular disparity, vergence may have also contributed to this advantage, as it provides information about the distance to objects in near space. This distance information is important to scale depth from binocular disparity.

The effect of shape

In Experiment 1, we found a significant effect of shape on PSEs with the pentagon being perceived as smaller than the quatrefoil, square, and triangle. Previous studies have found an effect of bilateral symmetry on visual perception and recognition of shapes, where symmetric shapes are judged more accurately and recognized faster than asymmetric shapes (Palmer, 1985). However, all the shapes in Experiment 1 were symmetric so this property cannot explain the bias. Zhang et al. (2024) found that shapes with multiple sides (such as a pentagon, or hexagon) that could approximate a circle, were perceived as larger than shapes with less sides (square and triangle). But this study only assessed 2D (flat) shape depictions. Thus, it is unclear if this result would translate to 3D shapes. Furthermore, this bias was not seen for the polygon in Experiment 2. Therefore, we interpret this outcome with caution.

In Experiment 2, the square was perceived as larger than the other shapes. This effect was consistent across manipulation conditions ($\eta^2 = .07$) and modality ($\eta^2 = .07$). Upon reviewing the appearance of the shapes, it was noted that, due to the lighting direction and orientation of the surfaces, there was a viewpoint in which the shading of the square stimulus became ambiguous. This caused the different faces of the square to blend together, making it look larger. This was not the case in Experiment 1 and there was no consistent over-estimation in the size of the square in that study. Based on the results from both Experiment 1 and Experiment 2, we conclude that the shape of an object can affect its perceived size, however, it does not affect precision or sensitivity in discriminating size differences.

The effect of interaction

Initially, we hypothesized that interaction with the stimuli would result in more precise size judgements. However, in Experiment 1 we did not find a significant effect of interaction on perceived size. Thus, we concluded that the stimuli were too simple, and since they appeared front facing to the observer, there was no need to interact with the shape to judge its size accurately. In Experiment 2, precision was similar for both passive and active conditions when the stimuli were viewed binocularly. However, precision was worse in the monocular active condition relative to the passive condition. This result was surprising, as we anticipated that the opportunity to engage with the stimuli would improve performance. Instead, it appears that interaction degraded precision. This result is inconsistent with previous research (Wexler & Boxtel, 2005).

However, unlike previous studies, this experiment did not involve direct contact with the object, which may have reduced the perceptual benefits of active engagement. For example, James et al., (2002) examined the effect of active exploration versus passive viewing on object recognition using VR. In their study, observers used their hands to manipulate a box with magnetic sensors that rotated the virtual object in real time. They concluded that manual exploration provided observers with proprioceptive and tactile feedback, allowing them to construct a coherent 3D model of the object. In contrast, our study used a virtual rod to pick up the shapes, avoiding tactile feedback that could directly inform observers of object size. Thus, although the observers still had access to proprioceptive and motor cues associated with their hand movements, they were forced to manipulate the object 'indirectly'. This, coupled with the lack of binocular depth cues in the monocular condition, may have added noise that degraded precision and reduced the benefit typically afforded by active engagement. Future work should examine if the addition of tactile feedback to this task would improve size judgements, particularly during monocular viewing, and compare the level of performance to that of binocular vision.

The role of experience

In Experiment 1, all observers had substantial VR experience, specifically in psychophysical tasks and extensive practice with VR headsets in experimental settings. We observed high precision across both manipulation conditions. In Experiment 2, in addition to the increased task difficulty, the observers had little to no experience with psychophysical tasks or VR displays. The results showed that under binocular viewing, precision was high, while precision was reduced under monocular viewing particularly in the active condition. While the increase in task difficulty contributed to this decline, the use of inexperienced observers may have also played a role. There is evidence in the literature that observers with little experience with 3D displays and psychophysics tend to make more errors in depth judgements (Hartle & Wilcox, 2016).

Experience with VR has also been shown to influence the extent to which observers rely on binocular depth information. In a study that investigated how observer experience influences depth cue use in VR, it was found that observers that were new to VR initially tended to show weak sensitivity to binocular cues, but sensitivity to both binocular and monocular cues increased with more time spent in the task (Fulvio et al., 2020). These findings suggest that some of the variability in size judgements in this study may be due to limited prior exposure to depth cues in virtual environments. Although a 28-trial preliminary session was completed prior to each manipulation condition, during which observers were instructed on how to use the controllers and pick up the shape, it is possible that this familiarization was insufficient to fully develop the necessary manipulation skills in an unfamiliar VR environment.

The use of motion-based depth information

One of the most interesting findings of Experiment 2 was that motion parallax did not effectively compensate for the absence of binocular depth cues during the monocular active condition. As outlined in the Introduction, motion parallax can, in principle, provide depth information comparable to stereopsis (Rogers & Graham, 1979). In Experiment 2, in both passive and active conditions, observers exhibited larger lateral head movements under monocular viewing compared to binocular viewing. Yet, the increase in movement did not translate to more precise size judgements, with precision being greatly reduced in the monocular active condition. Examination of the motion tracking data showed that the total range of movement did not exceed 5 cm (Figure 16). While this amount of movement seems small, prior research has shown that motion parallax can still provide reliable depth information with head movements as small as 5 – 8 mm (Aytekin & Rucci, 2012). It is possible that the limited head movements may reflect the increased task difficulty in the monocular active condition. As the observers were naïve, making size judgements in VR may have imposed substantial cognitive demands (particularly under monocular viewing). To manage this, observers may have prioritized visual clarity and stability over larger exploratory movements.

Naepflin & Menozzi (2001) found that extracting depth information from motion parallax takes more time compared to stereopsis. But our response time data showed that observers tended to respond within 5 seconds despite having up to 300 seconds if needed, suggesting that they had sufficient viewing time to use the depth information from motion parallax to make their size judgments. Other studies have shown that depth estimates from motion parallax can be generated faster than predicted by Nawrot & Stroyan (2012), but factors such as lack of experience, difficulty in moving the object, and judging object size, may have impaired performance. Future work should examine how greater lateral head movements, task timing, and observer training interact to determine the effectiveness of motion parallax in active conditions, particularly in VR environments where unfamiliarity with movement and interaction may impair performance.

Contribution of depth cues

In these experiments we systematically controlled which sources of depth information were available. As noted previously, in Experiment 1, the task was too easy and so there were no differences between conditions. However, in Experiment 2 where task demands were higher and judgements involved more complex 3D shapes, we found that observers tried to enhance motion-based cues in the absence of binocular cues to aid in their size judgements. Prior research has shown that stereopsis and motion cues are integrated to support accurate depth perception, and that such integration can often reflect an optimal strategy based on cue reliability (Johnston et al., 1994).

Under binocular viewing, observers had access to binocular disparity and vergence, in addition to motion-based cues such as structure from motion and motion parallax. In contrast, under monocular viewing, only motion-based cues were available. Previous studies have reported a stronger dependence on monocular depth information in VR (e.g. Fulvio et al., 2020), but these typically involved tasks such as object motion detection, which place different demands on depth processing. In contrast, this task required observers to make subtle size discriminations, judgements that may bias observers towards relying on the more precise depth information provided by binocular depth cues. As outlined in the Introduction, binocular depth cues in particular has been shown to support precise depth estimation in near space which is important for fine grained size comparisons (Gogel, 1969; Mon-Williams & Tresilian, 1999). This contrast underscores the importance of considering task demands when evaluating the contribution of different depth cues in VR. The effectiveness of a given cue may vary not only with its availability, but also with the nature and precision requirements of the perceptual task.

Considerations

Upon examination of the motion tracking data, we identified a possible limitation of how the stimuli were presented to the observers. In Experiment 2, the stimuli were initially presented with their side facing the observer. Response data showed that two observers made their responses within 2 seconds in both the passive and active conditions. With this rapid response rate there would have been little to no object rotation in the passive condition, or time to pick up the shape in the active condition. Although their response times were fast, it did not greatly reduce precision or accuracy of their size judgements. This suggests that these observers were able to use the vertical height of the stimulus to make precise judgements. Despite the presence of this strategy, we are confident that it was only used by two observers (based on tracking data). However, in future experiments we plan to vary size to avoid this confound.

Conclusion

The results of the experiments reported here highlight the role of binocular depth cues in determining object size. Performance on this task was most precise across all conditions in Experiment 2 when binocular cues were available. In the monocular conditions, despite increases in object and self-motion which should have aided performance, precision was consistently lower. More broadly, the findings contribute to our understanding of how the visual system uses available depth information to construct percepts of 3D size and highlight the importance of tailoring depth cue availability to the specific task demands in virtual environments. These results also have implications for the design and optimization of virtual reality environments. VR systems should prioritize the delivery of binocular depth cues though high-quality stereoscopic rendering to support fine perceptual judgements in immersive virtual environments.

References

Aytekin, M., & Rucci, M. (2012). Motion parallax from microscopic head movements during visual fixation. *Vision Research*, *70*, 7–17. https://doi.org/10.1016/j.visres.2012.07.017

Beck, J., & Gibson, J. J. (1955). The relation of apparent shape to apparent slant in the perception of objects. *Journal of Experimental Psychology*, *50*(2), 125–133. https://doi.org/10.1037/h0045219

Fulvio, J. M., Ji, M., Thompson, L., Rosenberg, A., & Rokers, B. (2020). Cue-dependent effects of VR experience on motion-in-depth sensitivity. *PLoS ONE*, *15*(3). https://doi.org/10.1371/journal.pone.0229929

Gibson, E. J., Gibson, J. J., Smith, O. W., & Flock, H. (1959). Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, *58*(1), 40–51. https://doi.org/10.1037/h0043883

Gogel, W.C. (1964). Perception of depth from binocular disparity. *Journal of Experimental Psychology*, *67*(4), 379–386. https://doi.org/10.1037/h0042913

Gogel, W. C. (1969). The sensing of retinal size. *Vision Research*, *9*(9), 1079–1094. https://doi.org/10.1016/0042-6989(69)90049-2

Gruber, H. E. (1956). The Size-Distance Paradox: A Reply to Gilinsky. *The American Journal of Psychology*, *69*(3), 469–476. https://doi.org/10.2307/1419056

Haber, R. N., & Levin, C. A. (2001). The independence of size perception and distance perception. *Perception & Psychophysics*, *63*(7), 1140–1152. https://doi.org/10.3758/BF03194530

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*(22), 1315–1318. https://doi.org/10.1016/S0960-9822(00)80053-6

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

Holway, A. H., & Boring, E. G. (1941). Determinants of Apparent Visual Size with Distance Variant. *The American Journal of Psychology*, *54*(1), 21–37. https://doi.org/10.2307/1417790

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

James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 383–390. https://doi.org/10.3758/BF03195466

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

Kaiser, P. K. (1967). Perceived shape and its dependency on perceived slant. *Journal of Experimental Psychology*, *75*(3), 345–353. https://doi.org/10.1037/h0025039

Kilpatrick, F. P., & Ittelson, W. H. (1953). The size-distance invariance hypothesis. *Psychological Review*, *60*(4), 223–231. https://doi.org/10.1037/h0060882

Langdon, J. (1951). The perception of a changing shape. *Quarterly Journal of Experimental Psychology*, *3*(4), 157–165. https://doi.org/10.1080/17470215108416791

Lichten, W., & Lurie, S. (1950). A New Technique for the Study of Perceived Size. *The American Journal of Psychology*, *63*(2), 280–282. https://doi.org/10.2307/1418936

Mckee, S. P., & Welch, L. (1992). The precision of size constancy. *Vision Research*, *32*(8), 1447–1460. https://doi.org/10.1016/0042-6989(92)90201-S

Mon-Williams, M., & Tresilian, J. R. (1999). Some Recent Studies on the Extraretinal Contribution to Distance Perception. *Perception*, *28*(2), 167–181. https://doi.org/10.1068/p2737

Meta. (n.d.). Tracking with Meta Quest: How the Meta Quest headset tracks movement. Meta. https://www.meta.com/help/quest/598701621088668/

Naepflin, U., & Menozzi, M. (2001). Can movement parallax compensate lacking stereopsis in spatial explorative search tasks? *Displays*, *22*(5), 157–164. https://doi.org/10.1016/S0141-9382(01)00067-1

Nawrot, M., & Stroyan, K. (2012). Integration time for the perception of depth from motion parallax. *Vision Research*, *59*, 64–71. https://doi.org/10.1016/j.visres.2012.02.007

Norman, J. F., Swindle, J. M., Jennings, L. R., Mullins, E. M., & Beers, A. M. (2009). Stereoscopic shape discrimination is well preserved across changes in object size. *Acta Psychologica*, *131*(2), 129–135. https://doi.org/10.1016/j.actpsy.2009.03.009

Palmer, S. E. (1985). The role of symmetry in shape perception. *Acta Psychologica*, *59*(1), 67–90. https://doi.org/10.1016/0001-6918(85)90042-3

Rogers, B., & Graham, M. (1979). Motion Parallax as an Independent Cue for Depth Perception. *Perception*, *8*(2), 125–134. https://doi.org/10.1068/p080125

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, *22*(2), 261–270. https://doi.org/10.1016/0042-6989(82)90126-2

Tittle, J. S., & Braunstein, M. L. (1991). Shape perception from binocular disparity and structure-from-motion. *Sensor Fusion III: 3D Perception and Recognition*, *1383*, 225–234. https://doi.org/10.1117/12.25258

Ullman, S., & Brenner, S. (1997). The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *203*(1153), 405–426. https://doi.org/10.1098/rspb.1979.0006

Wexler, M., & van Boxtel, J. J. A. (2005). Depth perception by the active observer. *Trends in Cognitive Sciences*, *9*(9), 431–438. https://doi.org/10.1016/j.tics.2005.06.018

Zhang, X., Li, S., & Yin, K. (2024). The effect of shape on visual size perception. *I-Perception*, *15*(4). https://doi.org/10.1177/20416695241270301