

VISUAL PERCEPTION OF POSITION AND SIZE BY MOVING AND STATIONARY
OBSERVERS

JOHN JONG-JIN KIM

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

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

June 2025

© John Jong-Jin Kim, 2025

Abstract

Visual spatial perception, including perception of objects in their surrounding environment while moving or stationary, is fundamental for any humans to navigate the world. Numerous visual cues provide a moving observer information about the space they navigate and correctly interpreting such information is vital to effectively interact with, or avoid, the elements around them. In Chapter 2, I found that people generally thought an object was more eccentric from straight-ahead than it actually was when judging its position. After being moved laterally, they made additional updating errors when they thought their movement could not be real (i.e., when they were sitting on a fixed chair that they knew could not be moved) compared to when the real movement is possible (i.e., when they were sitting on a moving platform that they knew could move). These updating errors depended on how much they were simulated to move. In Chapter 3, I provide evidence that familiar objects displayed in a 2D scene can be used to judge an object's size, but not its position (i.e., elevation in the scene). However, it is not enough to derive the correct distance simulated on a screen. In Chapter 4, I demonstrate that when an object displayed in a semi-immersive set up of a 2D scene (with full-size screen, a ground plane continuous with the experimental room, and familiar objects) people can correctly judge its size and absolute distance but not its position. Combined with the results from Chapter 3, people appear to rely on different visual cues when judging size and distance which suggest that the two processes are independent from each other. Incorrectly judged eye height and the height of the horizon in the scene may also influence people's perception of object size and position differently, but my results were not conclusive and further research is needed. Overall, my research has demonstrated that visual spatial perception is a complex process where people rely not only on different cues for different purposes but are also influenced by whether what they experience could be real or not.

Acknowledgements

First, I would like to thank my advisor Dr. Laurence R. Harris for his guidance and support. The opportunity he gave me to be part of his lab allowed me to learn and improve as a student and a researcher. His vast knowledge of the field greatly helped me in this challenging experience for which I am forever grateful.

I would also like to thank my committee members, Dr. Robert Allison and Dr. Taylor Cleworth for guiding me throughout the project.

Thanks to all the Harris Lab members past and present, especially Dr. Meaghan McManus, Dr. Björn Jörges, Ambika Bansal, Fatemeh Ghasemi, Anita Pandey, and Ahmed Nadeem for generously helping me throughout the project. Special thanks to Bob Hou for helping with any technical issues. It was a pleasure working with you all. You made the entire process smooth and enjoyable.

Thank you NSERC for funding the equipment used for the experiment which made the whole project possible.

Finally, I thank all the participants who volunteered their time and who made this experiment possible. Without your help, this study would have never reached its goal.

Thank you all

John Jong-Jin Kim

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
1.0. General Introduction	1
1.1. Perceiving the Position of an Object.....	2
1.1.1. Distance Perception	2
1.1.2. Direction Perception	5
1.1.3. Updating Object Position After a Self-motion.....	8
1.2. Perceiving Self-motion	10
1.2.1. Visual Motion Cues	11
1.2.2. Non-visual (Physical) Motion Cues.....	14
1.2.3. Integrating and Reweighing Self-Motion Cues	16
1.3. Cognitive Factor in Human Perception.....	17
1.4. Aims of this Dissertation	18
1.4.1. Spatial Updating with a Lateral Translation	19
1.4.2. Perception of Size and Distance in a 2D Scene Displayed on Computer Screens	19
1.4.3. Perception of Size and Distance in a 2D Scene Displayed on a Large Screen.....	20
2.0. The effect of knowing if you can really move on spatial updating	22
2.1. Abstract.....	22
2.2. Introduction.....	23
2.2.1. Hypotheses	27
2.3. Materials and Methods.....	27
2.3.1. Overview	27
2.3.2. Participants.....	28
2.3.3. Apparatus	28
2.3.4. Stimuli.....	28
2.3.5. Training.....	29
2.3.6. Procedure	30
2.3.7. Data analysis	31
2.4. Results.....	32
2.5. Discussion.....	34

2.5.1.	Comparing the results to the previous experiment	35
2.6.	Conclusion	37
3.0.	Can People Infer Distance in a 2D Scene Using the Visual Size and Position of an Object? 39	
3.1.	Abstract	39
3.2.	Introduction.....	40
3.2.1.	Representing depth in a 2D scene.....	41
3.2.2.	Quantifying depth in a 2D scene.....	43
3.2.2.	Hypotheses.....	45
3.3.	Materials and Methods.....	46
3.3.1.	Apparatus	46
3.3.2.	Tasks	50
3.3.3.	Methods for experiment 1	52
3.3.4.	Methods for experiment 2.....	53
3.3.5.	Methods for experiment 3	54
3.3.6.	Data analysis	56
3.3.7.	Post-test data cleaning.....	57
3.4.	Results.....	57
3.4.1.	Experiment 1: the position-to-size task and then the size-to-position task.....	57
3.4.2.	Experiment 2: size-to-position then position-to-size	61
3.4.3.	Experiment 3: familiar objects in the 2D scene	62
3.4.4.	Between experiments analysis	64
3.5.	Discussion.....	65
3.5.1.	Why is there a task order effect?.....	69
3.5.2.	Can familiar objects improve object distance judgements in a 2D scene?	71
3.5.3.	Limitations and future studies.....	73
3.6.	Conclusion	74
4.0.	Can people determine object distance from its visual size and position in a correctly scaled 2D scene displayed on a large screen with aligned ground plane?.....	75
4.1.	Abstract	75
4.2.	Introduction.....	76
4.2.1.	Judging Distance in an Immersive Environment.....	77
4.2.2.	Task Dependent Effect of Depth Cues: Using Size or Position.....	78
4.2.3.	Judging Size and Distance in a 2D Scene Monocularly or Binocularly	79
4.2.4.	Hypotheses.....	80

4.3.	General Methods.....	81
4.3.1.	Materials and Methods.....	81
4.4.	Experiment 1.....	88
4.4.1.	Materials and Methods.....	88
4.4.2.	Results.....	93
4.4.3.	Discussion of Experiment 1.....	98
4.5.	Experiment 2.....	102
4.5.1.	Materials and Methods.....	103
4.5.2.	Results.....	108
4.5.3.	Discussion of Experiment 2.....	113
4.6.	General Discussion.....	118
5.0.	General Discussions.....	121
5.1.	Misperceived Self-motion Leads to Updating Errors.....	121
5.1.1.	Lack of physical motion cues.....	121
5.1.2.	Possibility of Physical Motion.....	122
5.2.	Visual Cues Influencing Size and Distance Perception.....	123
5.2.1.	Familiar Objects in the Scene and Ground Information.....	123
5.2.2.	The Horizon in the Scene and Eye-height.....	124
5.2.3.	Judging Size and Distance may use Different Visual Cues.....	126
5.3.	Immersive and Familiar Test Setting Improves Spatial Perception.....	126
5.4.	Limitations and Future Directions.....	127
5.5.	General Conclusions.....	128
6.0.	References.....	131
	Appendix A.....	140
	Appendix B.....	143

List of Tables

<i>Table 3.1:</i> Average initial and final sizes/positions for all tasks.	59
<i>Table A.1:</i> Exp 1 – Repeated Measures ANOVA – comparing target sizes.	140
<i>Table A.2:</i> Exp 1 – Repeated Measures ANOVA – comparing target positions.	140
<i>Table A.3:</i> Exp 2 – Repeated Measures ANOVA – comparing target positions.	140
<i>Table A.4:</i> Exp 2 – Repeated Measures ANOVA – comparing target sizes.	141
<i>Table A.5:</i> Exp 3: Group 1 – Repeated Measures ANOVA – comparing target positions.	141
<i>Table A.6:</i> Exp 3: Group 1 – Repeated Measures ANOVA – comparing target sizes.	141
<i>Table A.7:</i> Exp 3: Group 2 – Repeated Measures ANOVA – comparing target sizes.	142
<i>Table A.8:</i> Exp 3: Group 2 – Repeated Measures ANOVA – comparing target positions.	142

List of Figures

<i>Figure 1.1:</i> Sensitivity to depth cues. The difference in depth (Δdd) required for observers to distinguish two distances as a function of absolute distance (from Cutting & Vishton, 1995). The orange shaded area in the graph, between 1.75 m and 25 m, represents the distance covered in this dissertation.	3
<i>Figure 1.2:</i> Visual size and ground texture as distance cues. A) Visual size (red lines) of an object at different distances projected on retina. B) Ground texture in perspective to distance.....	5
<i>Figure 1.3:</i> Representations of the direction of an object from the observer. A) Directions in three-dimensional orthogonal, Cartesian axes: vertical, lateral, and longitudinal, from the observer's head. B) The visual direction of an object is obtained from the retinal images of the two eyes. X indicates the position of the 'cyclopean eye' which represents the origin of direction of the object perceived.	6
<i>Figure 1.4:</i> Diagram of the visual pathway. The visual pathway comprises of the neurons on the retina (retinal ganglion cells) that get the light signals from the photoreceptors on them where the right visual field is projected on the left side of the retina and the left visual field on the right side. These signals are sent out through the optic nerve, pass through the optic chiasm where the signals from the left visual field are crossed to the right side of the brain and the signals from the right visual field to the left side. Most of these signals synapse at the lateral geniculate nucleus (LGN) on the corresponding, then sent to the primary visual cortex (V1) where the visual signals from the both visual fields meet. (from De Moraes, 2013)	7
<i>Figure 1.5:</i> Change in object position relative to the observer before and after moving. These diagrams show the change in object position, originally in front of the observer, after different types of movement. A) Before the observer moves, the object is in front of them at a distance. B) After the observer rotates, in the yaw axis around the cyclopean eye, the direction of the object changes, but not the distance of the object. C) After the observer translates towards the object, the distance of the object changes, but not the direction of the object. D) After the observer translates laterally to the right, both the direction and the distance of the object changes relative to the observer.....	10
<i>Figure 1.6:</i> Examples of optic flow for different types of self-motion. A) When moving straight forward, optic flow radiates from the focus of expansion in the optic flow which indicates the direction of the movement. B) When moving to the right, optic flow moves to the left in the observer view which is the opposite direction of self-motion. In both diagrams, the black arrows represent the direction of self-motion, the red lines represent the direction of the objects' movement relative to the observer during self-motion, and the red dotted arrows represent the direction of the optic flow.....	12
<i>Figure 1.7:</i> Diagram of the vestibular system. The semicircular canals: anterior, posterior, and horizontal canals detect rotational acceleration and the otolith organs: utricle, macula, and saccule, detect linear acceleration of the head. (from Day & Fitzpatrick, 2005)	15
<i>Figure 2.1:</i> The Experimental setup. (A) Photo of the office-like test room with a chair used for the new experiment. (B) Photo of the MOOG test room with a seat on the motion platform used to produce physical motion cues in the old experiment. (C) Example image of the virtual rendition of the MOOG test room used in both experiments with a virtual seat and a screen.	26
<i>Figure 2.2:</i> Experimental procedures. (A) Participants fixated on the fixation cross for 500 ms. (B) Visual target presented for 500 ms. (C) Randomly generated blinking dots appeared on the screen after the visual target disappeared. (D) Participants were either stationary or passively	

- moved laterally for 5 seconds. (E) The screen turned dark. Participants then moved the indicator to the remembered location of the visual target as fast as they could..... 31
- Figure 2.3:* The positions at which participants remembered the targets. Black - Stationary, Purple - Short (.46 m), and Red - Long (1 m) travel conditions. The data from the rightward movement directions have been flipped to superimpose on those to the left where positive (+) values represent errors in the direction of observer's movement. The lighter colored dots represent remembered positions from each participant (N = 23) and the darker colored dots and lines represent the means. The dashed line represents true target positions, therefore the deviation from the dashed line to the remembered target position is the error made. For both x- and y-axes, positive (+) represents in the direction of the movement and negative (-), the opposite direction. The orange and green shaded area contains the targets within the start and the end positions (orange from 0 m to 0.46 m, and green from 0m to 1m) between which the observer traveled in the moving conditions. 33
- Figure 2.4:* Comparison between old and new. Mean errors participants made for each target for Stationary (Black) and Visual Motion (Red) conditions for old and the equivalent targets from new, where 0 on the vertical axis is the start position and collapsed across observer positions: Left and Right. For both x- and y-axis, positive (+) represents in the direction of the movement, and negative (-) the opposite direction. The faded area contains the errors for the targets from the start and the end position (from 0 to .46m) where the observer traveled in the Visual Motion conditions. 36
- Figure 2.5:* Impact of distance compression on updating position of an object after a lateral translation. Underestimated absolute distance of a target results in the viewer overestimating the shift in target position after being translated laterally, remembering it to be more eccentric than it is after moving. Therefore, updating error will be in the opposite direction of the viewer's translation..... 37
- Figure 3.1:* Example of a 2D scene with objects at various distances. Distances are represented with the objects' (1) size —objects of smaller visual size (e.g., the running person and the tree on the right side) are perceived as further away compared to the same objects of larger visual size (e.g., the standing person and the tree on the left of the figure); and their (2) distance below the horizon —objects placed closer to the horizon (the running person and the tree on the right side) are perceived as further away compared to objects further below the horizon (e.g., the standing person and the tree on the left side). The ground plane and the horizon play big roles in depth perception in this example. The running person, the dog, the airplane, and the bird in this scene are all the same visual size. When comparing the person and the dog, it appears obvious the person is further away. However, when comparing them to the bird or the plane, it is ambiguous whether the bird/airplane is: (i) flying at a distance close to the viewer and is smaller in physical size or (ii) flying at a far distance but is larger in size. 42
- Figure 3.2:* Geometry of an object represented on a 2D platform (the screen). The elevation of the object on the screen can be calculated with: $d_{bottom} = D_{screen} \times H_{eye} D_{object}$, and the height of the object on the screen can be expressed as h_{object} or: $d_{bottom} - d_{top}$, where $d_{top} = D_{screen} \times (H_{eye} - H_{object}) D_{target}$. Hence, $h_{object} = d_{bottom} - d_{top} = H_{object} \times D_{screen} D_{object}$ 44
- Figure 3.3:* Entering reference size. (A) Example photo of the phone size setting stage at the start of the experiment. The width (W_{phone}) and the height (H_{phone}) of the phones were recorded as a fraction of the height of the screen. (B) Example photo of the reference length ($H_{reference}$) setting stage. 47

<i>Figure 3.4: Visual stimuli and tasks.</i>	50
<i>Figure 3.5: Task orders for Exp 1 and 2.</i>	53
<i>Figure 3.6: The pathway scene with familiar objects.</i>	55
<i>Figure 3.7: Participants' settings for all tasks.</i>	60
<i>Figure 4.1: Test apparatus and setup.</i>	82
<i>Figure 4.2: Test stimuli.</i>	83
<i>Figure 4.3: Target distance simulated in a projection.</i>	84
<i>Figure 4.4: Relationship between stimulus size, target base position, and simulated distance presented graphically.</i>	85
<i>Figure 4.5: Diagram of the horizon in a visual scene and eye height of the observers.</i>	87
<i>Figure 4.6: Experimental conditions and orders.</i>	91
<i>Figure 4.7: Exp 1 – Target size and position from each task.</i>	95
<i>Figure 4.8: Exp 1 – Intercepts and Slopes.</i>	97
<i>Figure 4.9: Participants' responses in the past and the present experiments.</i>	99
<i>Figure 4.10: Example of cross-correlation analysis and participant responses.</i>	108
<i>Figure 4.11: Exp 2 – Target size and position from each task.</i>	110
<i>Figure 4.12: Exp 2 - Intercepts and Slopes.</i>	112

Visual Perception of Position and Size by Moving and Stationary Observers

1.0. General Introduction

In our daily lives, we must navigate through environments filled with objects. To do this effectively and safely, it is important to accurately gauge the spatial relationships between those objects in the environment and the self. This is called spatial perception and it is a key ability for any living animal to navigate the world. We can perceive space using various sensory systems including auditory (Voss, 2016), haptic (Kappers & Koenderink, 1999), and visual systems (Sedgwick, 2005). The visual system is the most reliable in spatial perception, however, people still make errors in distance (Foley, 1980; Z. Li et al., 2011) and position (Kim, 2019; Prablanc et al., 1979; Vercher et al., 1994). In this dissertation, I measured people's visuospatial ability to judge an object's position relative to them when viewing it in 2D and 3D rendition of a space. By varying visual cues and the presentation methods of the scenes I investigated how different visual cues impact people's perception of object positions in space.

In this introductory chapter, I review how people perceive the position of an object in space and the contributing factors that can influence its perceived position and size. In Chapter 2, I investigate how well the perceived position of an object is updated after a lateral self-motion in a 3D virtual reality. Chapter 3, I look at the accuracy of people's distance perception when viewing an object on a 2D rendition of a scene on a computer screen using the object's size and its elevation on the screen. In Chapter 4, the perceived distance using the object's size and its elevation on the screen is measured again with a full-size projection of the 2D scene onto a wall. Overall, this dissertation focuses on the contribution of self-motion and spatial features in the human perception of an object in their visual field.

1.1. Perceiving the Position of an Object

The key parameters that describe an object's position relative to the observer (i.e., egocentric representation of object position) are its distance and direction. The distance of an object from the observer (i.e., absolute or egocentric distance) is the amount of space between them. Knowing the correct distance to an object is important for various actions such as reaching, walking or driving to gauge how far one must move to interact with it. Knowing the direction of an object from the observer allows the person to plan out the direction of their movement to either avoid or interact with it when navigating.

It is also important to consider the relative distance (i.e., depth) and position between objects when there is more than one object (i.e., allocentric representation of object's position relative to other objects) as people use both egocentric and allocentric representations of space in their spatial perception and memory (Burgess, 2006). In this dissertation, I only measured the observer's perception of distance and position relative to themselves (i.e., absolute distance).

1.1.1. Distance Perception

There are two classes of visual distance cues involved in judging the absolute distance of an object: binocular and monocular cues. Binocular cues, such as stereoscopic disparity (Mather & Smith, 2000) - difference in the images projected on the two retinas - and vergence (Tresilian et al., 1999) - oculomotor cue from the two eyes' movements when looking at the object, are from the sensory information from the right and left eyes. Monocular cues, such as accommodation (Takeda et al., 1999) - oculomotor cues from the eyes' lenses adjusted to focus on the object, visual size (Ittelson, 1951; Kunnapas, 1968), and ground texture (Brenner & Smeets, 2018; Sedgwick, 2021) are some of the sources of visual information that can be detected with one eye. Motion (or monocular) parallax - angular velocity of an object inversely

proportional to its absolute distance, also provide depth information (Rogers & Graham, 1979). The sensitivity to each of these distance cues varies depending on the distance of the object, as shown in Figure 1.1. In this dissertation, all the distances were between 1.75 m and 25 m, which is considered within action space.

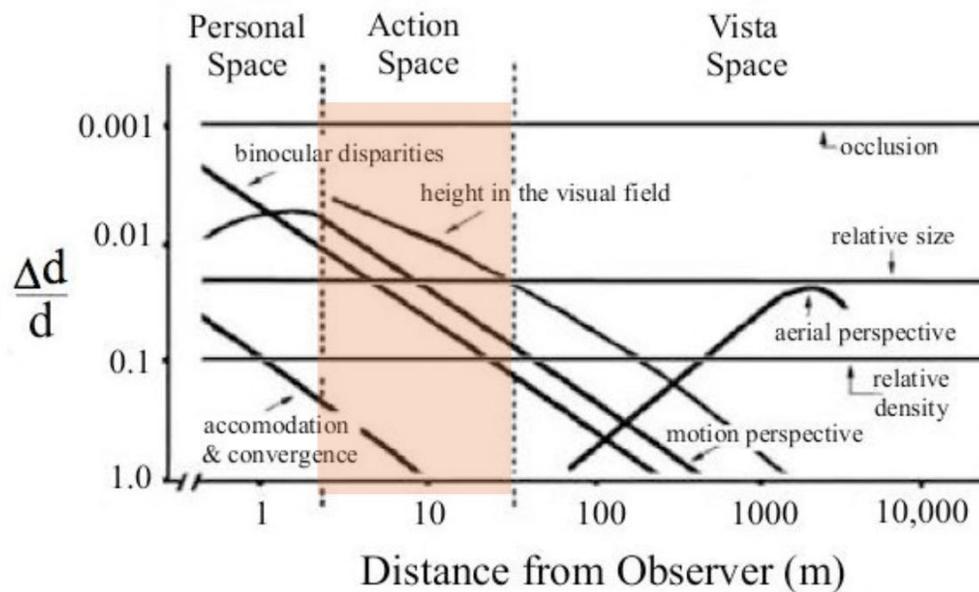


Figure 1.1: Sensitivity to depth cues. The difference in depth ($\frac{\Delta d}{d}$) required for observers to distinguish two distances as a function of absolute distance (from Cutting & Vishton, 1995). The orange shaded area in the graph, between 1.75 m and 25 m, represents the distance covered in this dissertation.

Past studies have shown that binocular depth cues become less reliable at a long distance as the disparity between the two eyes decreases (Harwerth et al., 1998; Schor & Flom, 1969). Although it does not become obsolete and still helps in judging relative distances between light sources (Allison et al., 2009; Palmisano et al., 2010), its effectiveness decreases beyond a few meters resulting in larger errors for more distant objects (Foley, 1980). Binocular cues can work against the viewer when working to interpret a 3-dimensional (3D) scene portrayed on a 2-dimensional (2D) space, e.g., a painting or a computer screen, since they tell the viewer about the

distance of the canvas or screen and compete with the distance portrayed in the picture. Therefore, we must rely more on monocular cues when looking at an object at far distances or at a scene projected on a 2D display. In the real world, monocular cues such as accommodation is typically accompanied by a vergence appropriate for the distance of the object as well as a change in pupil size. However, when looking at objects in a 2D scene, represented as being at different distance from the representation, the eyes can accommodate on the distance appropriate for the *apparent* distance of the object without a change in vergence (Takeda et al., 1999). This would be counterproductive and tend to blur the image confirming that binocular cues are not useful and can even impede judging distance in a 2D scene.

Visual size (or retinal size) of an object is an important cue when judging distance. The projection of an object on the retina becomes larger as the object moves closer and smaller as it moves further away (Figure 1.2A). In the natural world, the actual size of an object generally does not change (size constancy), and people perceive differences in visual size as difference in distance rather than size. When judging relative distance between objects, people can rely on this difference in visual size of the objects. To judge absolute distance, however, the correct size of the object must be known to the observer. Determining distance from retinal size is therefore influenced by prior experience with an object's size (Carlson & Tassone, 1971; Maltz et al., 2021; Predebon, 1993; Sousa et al., 2011). This 'familiar size effect' has been shown to lead to incorrect judgments of an object's size and distance if it is made to be different from its typical or familiar size (Maltz et al., 2021; Predebon, 1993), for example by being misled that a miniature Rubik's cube is actually a real-sized one.

The texture of the ground plane also provides important distance information which helps in judging the distance of an object on, or near, the ground. Similar to visual size, the texture on

the ground surface typically becomes finer, i.e., the texture elements (such as bricks, stones, or carpet pattern) become closer together and smaller, as distance increases (Brenner & Smeets, 2018; Sedgwick, 2021). See Figure 1.2B.

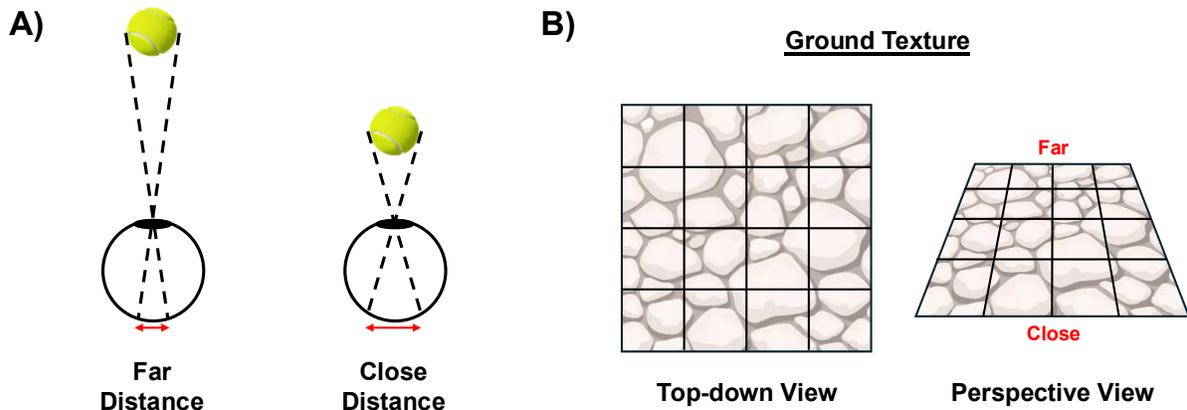


Figure 1.2: Visual size and ground texture as distance cues. A) Visual size (red lines) of an object at different distances projected on retina. B) Ground texture in perspective to distance.

As explained in this section, there are various distance cues people need to consider to correctly judge distance to an object. In this dissertation, I will address the effectiveness of monocular cues such as object height in the visual field and relative size, in the distance judgment of an object in a 2D scene. Now I will explain how people perceive the other key parameter of an object's position, the direction of an object.

1.1.2. Direction Perception

The position of an object from an observer can be represented in the 3D Euclidean space: vertical (up/down), lateral (left/right), and longitudinal (front/back), with the observer in the center (see Figure 1.3A) or using a polar coordinate system (using angular direction and distance from a reference, e.g., straight forward) originating from the observer or indeed any other of an infinite number of coordinate systems. Typically the visual direction, the direction from the eyes, of an object is believed to be judged from the 'cyclopean eye', a location in the head that serves

as the origin for directional judgement – in the center between the two eyes (Ono et al., 2002). Although it has been argued that cyclopean eye may not be required in directional judgement (Erkelens & van Ee, 2002) but for the purpose of this dissertation all directions of objects will be considered using the cyclopean eye as the origin.

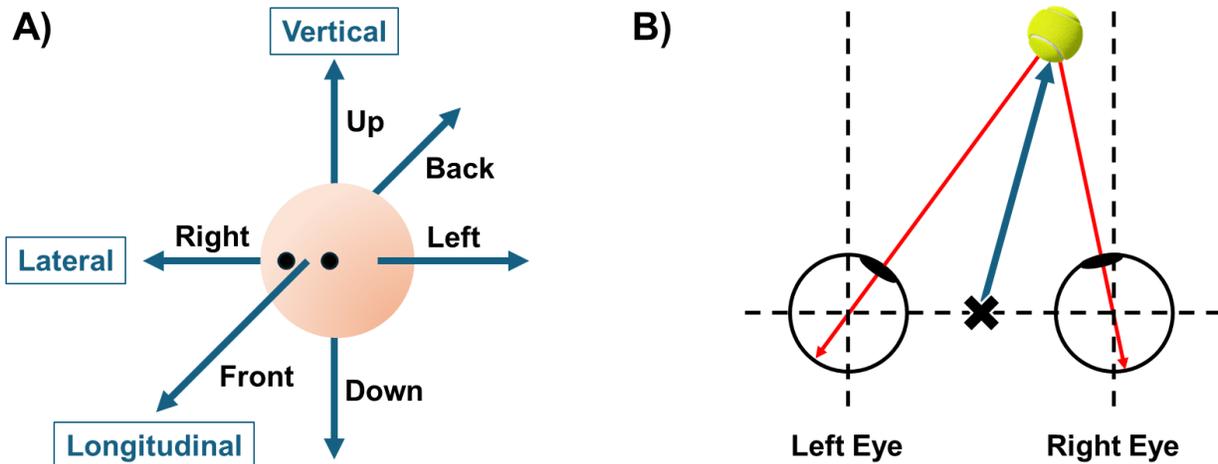


Figure 1.3: Representations of the direction of an object from the observer. A) Directions in three-dimensional orthogonal, Cartesian axes: vertical, lateral, and longitudinal, from the observer's head. B) The visual direction of an object is obtained from the retinal images of the two eyes. X indicates the position of the 'cyclopean eye' which represents the origin of direction of the object perceived.

The visual direction is obtained by the observer from the location of the stimuli projected onto the retina of the eyes where the image of the object is projected on different locations on the retinas (See Figure 3B). These visual signals from the retinas are sent to the visual cortex of the brain following the visual pathway as shown in Figure 1.4 (De Moraes, 2013), then combined in V1 (Hubel & Wiesel, 1972). However, there is evidence that suggest some binocular integration of these signals may start earlier in the LGN (Maier et al., 2022). These combined retinal signals provide information on where the visual stimulus is located. The locations stimulated on the retina is related to where the eyes are pointed, therefore the eye positions information is also important in judging target direction which can be obtained from the proprioceptive signals from

the eye muscles and tendons (Steinbach, 1987). Then how accurate are people at judging object direction from this information?

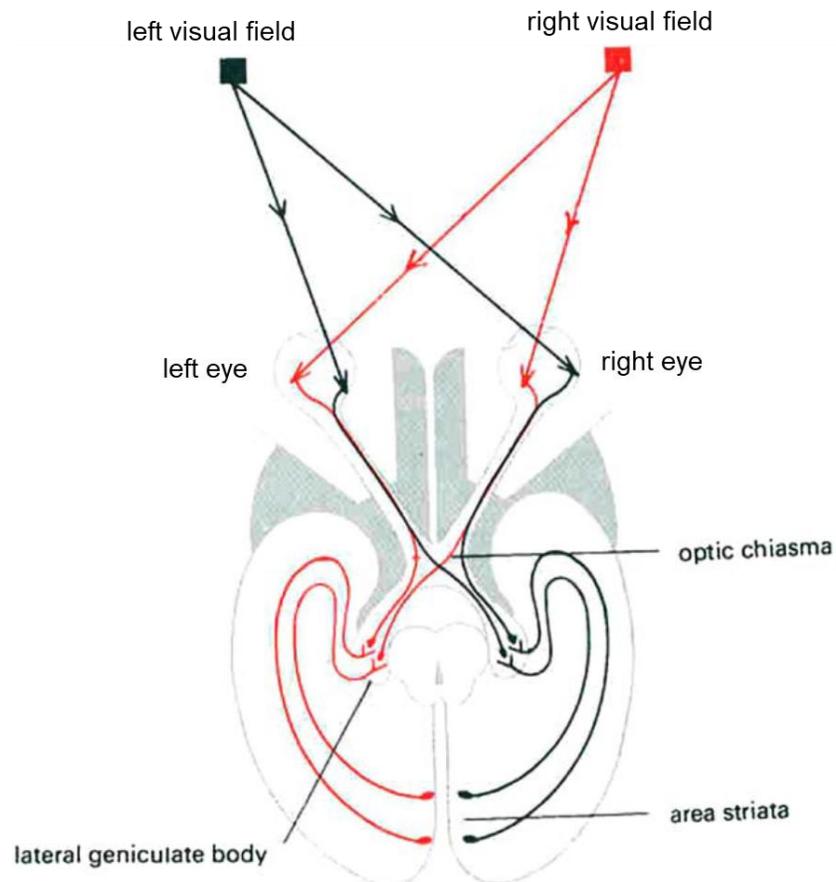


Figure 1.4: Diagram of the visual pathway. The visual pathway comprises of the neurons on the retina (retinal ganglion cells) that get the light signals from the photoreceptors on them where the right visual field is projected on the left side of the retina and the left visual field on the right side. These signals are sent out through the optic nerve, pass through the optic chiasm where the signals from the left visual field are crossed to the right side of the brain and the signals from the right visual field to the left side. Most of these signals synapse at the lateral geniculate nucleus (LGN) on the corresponding, then sent to the primary visual cortex (V1) where the visual signals from the both visual fields meet. (from De Moraes, 2013)

Studies have shown that people make systematic errors in judging an object's direction where they show a bias towards straight ahead (i.e., they judge objects as being less eccentric than they really are) when asked to point at a target while fixating it, but they show opposite bias when fixating straight ahead (Lewald & Ehrenstein, 2000). When people are asked to remember

the position of a target that was visible for just a brief moment, people show a bias towards the straight ahead (Kim, 2019; Prablanc et al., 1979; Vercher et al., 1994). These directional errors appear to be from the misjudged internal representation of space and head position with respect to the stable internal reference of the trunk position (Lewald & Ehrenstein, 2000), but may also be due to misregistering eye position or to misidentifying which retinal location was simulated (Morgan, 1978).

In this dissertation, I will use virtual reality (VR) to evaluate the effect of object eccentricity on the errors people make when remembering its position. However, the relative position of an object (distance and/or direction from the observer) changes when the observer moves and its position needs to be updated in accordance with the observer's movement. In the next section, I will explain this process of updating consequent on self-motion.

1.1.3. Updating Object Position After a Self-motion

The position of an object in the environment changes relative to an observer every time the observer moves, and so, if the observer wants to successfully interact with the object, its perceived position must be updated. Spatial updating, the cognitive ability to keep track of objects after self-movement, is an important skill when we are navigating around the world (Riecke et al., 2005; Wang, 2017; Wolbers et al., 2008). Although spatial updating in the brain occurs following all sorts of movement including eye (e.g., saccade) and head movement (Klier & Angelaki, 2008; Wolbers et al., 2008), in this dissertation I will focus on the updating consequent on whole-body self-motion.

Spatial updating is a complex process that relies heavily on our perception of space, i.e., the relative positions of the objects in the environment (Burgess, 2006; Wang & Spelke, 2000), as well as our knowledge of our own self-motion (Burgess, 2006). In the brain, spatial updating

has been shown to occur in the parietal cortex (Merriam et al., 2003). Spatial constancy, an ability to perceive the world as stable as we view it from different positions, is important in determining that the change in the relative position of objects is due to self-motion rather than movement in the environment. Past studies using neural recordings from macaque brain (Heiser & Colby, 2006) and neural network models (White III & Snyder, 2007) have suggested that this allocentric representation of a static space is encoded in lateral intraparietal cortex which is known to be involved in spatial working memory (Silk et al., 2010). In contrast, the egocentric representation of space changes with self-motion and this representation seems to be stored and updated in the precuneus, a region in the medial parietal cortex (Wolbers et al., 2008).

The change in the spatial relationship between an object and the observer depends on the type of observer movement (see Figure 1.5). After a rotation (e.g., a yaw around the cyclopean eye) only the azimuth direction of the object needs to be updated (Figure 1.5B) whereas in comparison after a linear movement, its distance (Figure 1.5C) or both direction and distance (Figure 1.5D) may change. Along with these different demands placed by the different movements, the accuracy in updating performance differs between the rotation and linear translation. People are quite accurate at updating when passively rotated (Riecke et al., 2005) but make systematic errors following passive linear translation - underestimating the angular change of the target after a lateral (Gutteling & Medendorp, 2016; Kim, 2019; Klier et al., 2008) and downward translation (Klier et al., 2008), and overestimating it after a upward, forward and backward translation (Klier et al., 2008).

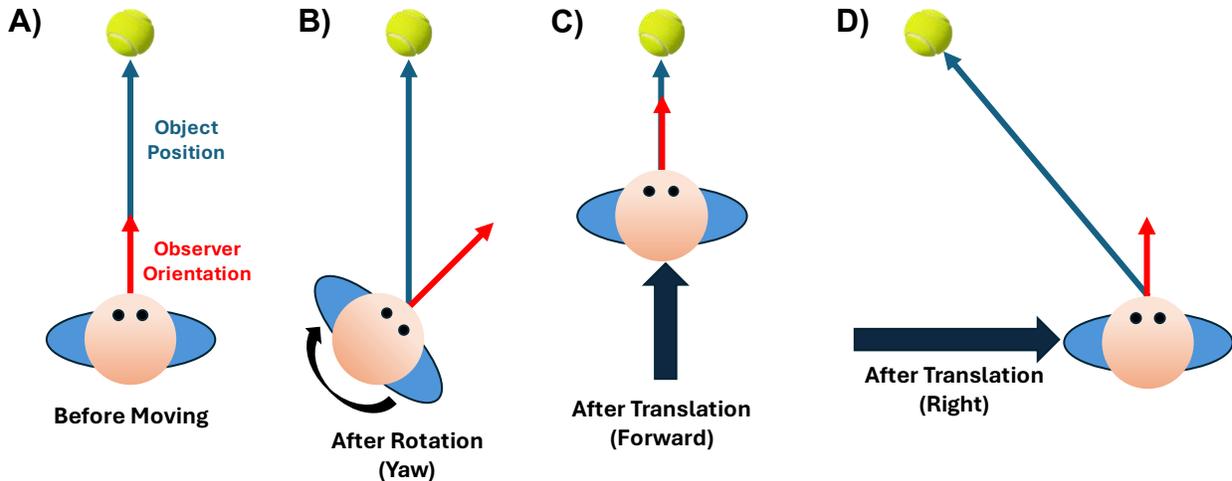


Figure 1.5: Change in object position relative to the observer before and after moving. These diagrams show the change in object position, originally in front of the observer, after different types of movement. A) Before the observer moves, the object is in front of them at a distance. B) After the observer rotates, in the yaw axis around the cyclopean eye, the direction of the object changes, but not the distance of the object. C) After the observer translates towards the object, the distance of the object changes, but not the direction of the object. D) After the observer translates laterally to the right, both the direction and the distance of the object changes relative to the observer.

In summary, spatial updating requires the representation of the static allocentric space (assuming nothing is independently moving in the world), as well as dynamic egocentric space which needs to be updated after every self-motion. To accurately compute the change in object position after moving, i.e., to update it, the observer must first accurately judge their self-motion. People receive various pieces of sensory information while moving. The next section will review these motion cues used to perceive self-motion.

1.2. Perceiving Self-motion

Knowing one's own self-motion is important for many reasons such as navigation and wayfinding. It is also required to correctly update the positions of objects in the environment while navigating as explained in the previous section. In an active motion, such as walking or cycling, motor commands are sent from the brain controlling our physical movement (Pearson,

1976). During this process, a copy of the motor command signal, i.e., an efferent copy, is provided to the rest of the brain to allow an anticipatory estimation of the sensory consequences of the action (Cullen, 2011; Harris et al., 2002) and the future state of the body, i.e., where a moving limb is going to be (Ebner & Pasalar, 2008). Our body also senses the self-motion from various self-motion cues registered by the visual (Britten, 2008; Lappe et al., 1999), vestibular (Cheng & Gu, 2018; Cullen, 2019; Harris et al., 2000) and somatosensory systems (Murata et al., 2014; Probst et al., 1985). During a passive motion where we are moved without any intent for action, e.g., sitting on a moving chair or in a vehicle, we must rely on these sensory self-motion cues to know we are moving. I will now outline what these cues are and how the brain is able to use them to generate a perception of self-motion.

1.2.1. Visual Motion Cues

Optic flow, a pattern of visual motion caused by relative motion between the observer and the visual scene (Gibson, 1950), is a visual cue that is a strong indicator of self-motion for any living creatures with ‘eyes’. From this optic flow, the observer can extract their moving direction (Warren Jr & Hannon, 1988) which is in the opposite direction of the pattern of motion. The nature of flow patterns differs between types of movements. For example, when moving straight forward the flow of pattern radiates from the point the observer is moving towards (the focus of expansion) whereas when moving to the right (laterally) the pattern generally moves to the left in the observer view as shown in Figure 1.6. Typically, optic flow accompanies self-motion, but in some cases, we observe flow of pattern from movement of a visual scene that is consistent with self-motion but is not caused by actual self-motion (e.g., the train illusion - looking at a moving train through the window of one’s stationary train). This visual experience

can be interpreted, erroneously, as a self-motion and this phenomenon is called vection, an illusory perception of self-motion induced by optic flow (Howard & Howard, 1994).

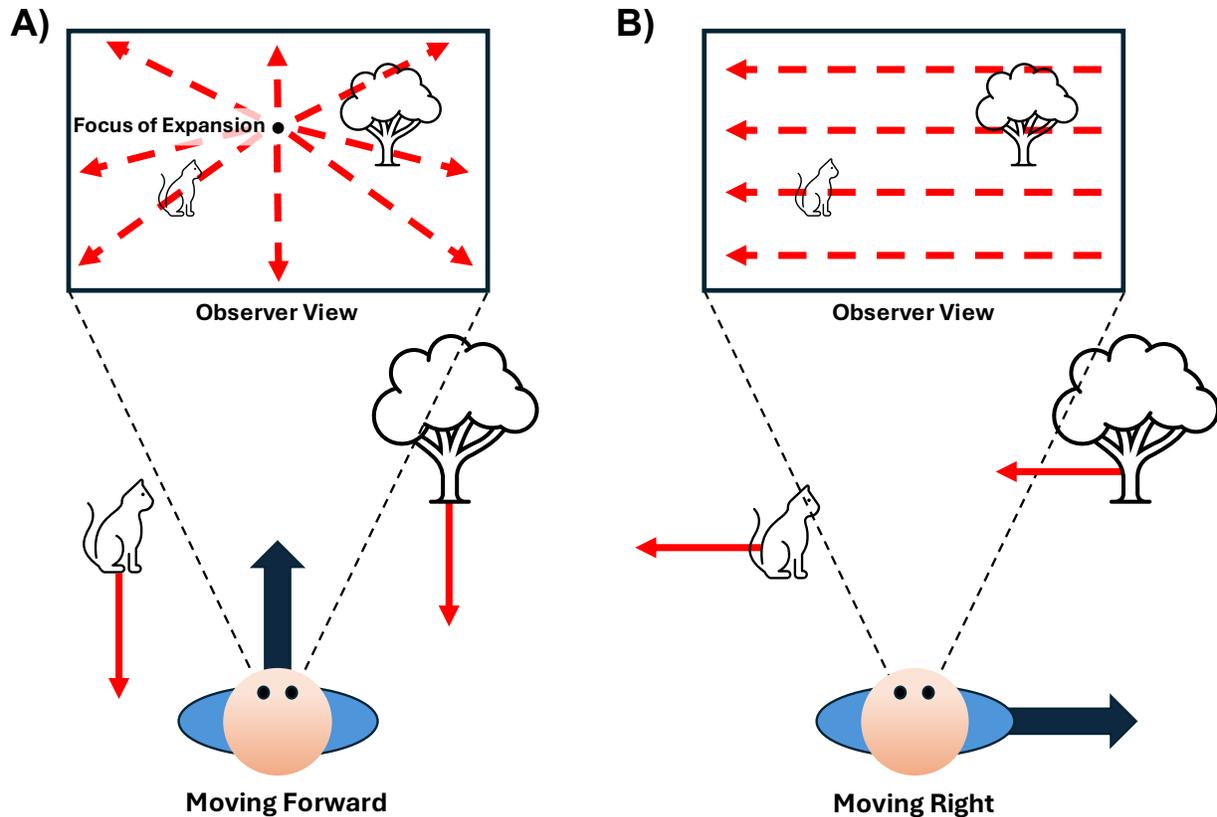


Figure 1.6: Examples of optic flow for different types of self-motion. A) When moving straight forward, optic flow radiates from the focus of expansion in the optic flow which indicates the direction of the movement. B) When moving to the right, optic flow moves to the left in the observer view which is the opposite direction of self-motion. In both diagrams, the black arrows represent the direction of self-motion, the red lines represent the direction of the objects' movement relative to the observer during self-motion, and the red dotted arrows represent the direction of the optic flow.

Recent advancement of technology using VR has allowed the scientific community to conduct visual perception research using VR as a visual display in which the designer has more control over the visual stimuli presented and can expose people to sensory experiences difficult to create in the real world such as segregating visual and non-visual motion cues. VR head-mounted displays (HMDs) are an effective tool to study visual aspects of self-motion (Kim et al.,

2015; Riecke et al., 2012) and are now widely used in the community to induce vection.

Enhancement in vection indicates an enhancement in self-motion perception and the results from such VR studies reveal various motion cues that contribute to self-motion perception that were not considered before. For example, visual motion cues, other than optic flow, such as simulated viewpoint oscillation (similar to that generated in natural walking) prior to and during exposure to radial visual flow (i.e., forward optic flow) was shown to reduce vection onset time compared to no oscillation (Palmitano & Riecke, 2018). Non-visual cues such as changes in location (Riecke et al., 2009) and intensity (Seno et al., 2012) of sound from an ‘auditory landmark’ in sync with visual motion can also make vection more convincing.

Change in the spatial relationship between the observer and the environment (i.e., optic flow) typically occurs with self-motion, but not always. The perception of self-motion can also be induced using a ‘moving room’ in which all the walls move together (Bertenthal & Bai, 1989) or using VR, as just explained, without actual self-motion. Therefore, visual motion cues alone are vulnerable to trickery and are not enough to reliably judge self-motion. When looking at a moving object during self-motion, e.g., following a ball while running after it, the geometry becomes even more complicated as the observer must extract the object-motion from the visual optic flow which include component due to both object- and self-motion. When vestibular and somatosensory motion cues are available during a self-motion, in addition to visual motion cues, people can detect object-motion more accurately (Fajen & Matthis, 2013; Probst et al., 1986) which demonstrates that non-visual motion cues are helpful for segregating object- and self-motion when both are moving.

From this brief review we can see that various visual motion cues provide motion cues which can be used to perceive self-motion and its direction. Visual motion cues alone can induce

perception of self-motion, but non-visual motion cues are also involved which enhance the experience. The following section will explain the non-visual motion cues for self-motion.

1.2.2. Non-visual (Physical) Motion Cues

When moving, we receive various somatosensory signals from all over our body. From proprioception, we sense the movement and positions of our limbs that tells us we are moving (Cullen, 2011; Harris et al., 2002). On exposed skin, airflow cues have been shown to provide sense of self-motion in humans (Murata et al., 2014) as well as in rats (Polat et al., 2024). During acceleration or deceleration, there are even more self-motion cues available. Change in pressure on our skin (e.g., on the back while sitting in an accelerating car) tell us our body is accelerating. The shear force induced by buttock skin stretch while sitting on a moving seat has been shown to enhancevection (Horie et al., 2018).

Our vestibular system sensory organs sense acceleration: the semicircular canals sense rotational acceleration, and the otolith organs sense linear acceleration (see Figure 1.7). Due to the vestibular system being specialized in detecting acceleration, it is also affected by the constant acceleration that all living creatures on Earth are exposed to, the Earth's gravity. Gravity is consistently pulling everything on earth towards its center, the hair cells in the otolith are bent down producing the sensory signals equivalent to the acceleration of 9.8m/s^2 away from the earth (i.e., upwards) even at rest. For this reason, we must be able to separate vestibular signals caused by bodily acceleration and those caused by gravity to accurately determine our self-motion. When stationary, or moving at a constant speed, the vestibular system is stable which indicates no additional head acceleration. When the head is accelerating, the hair cells in the semicircular canals and the otolith organs are bent (in addition to the bend due to gravity, depending on the type of motion) signalling to the brain that an acceleration is occurring (Day &

Fitzpatrick, 2005). The direction and the type (rotation or linear) of acceleration can be segregated from the otolithic and semicircular canal signals in the vestibular nuclei in the brainstem (Angelaki et al., 2004).

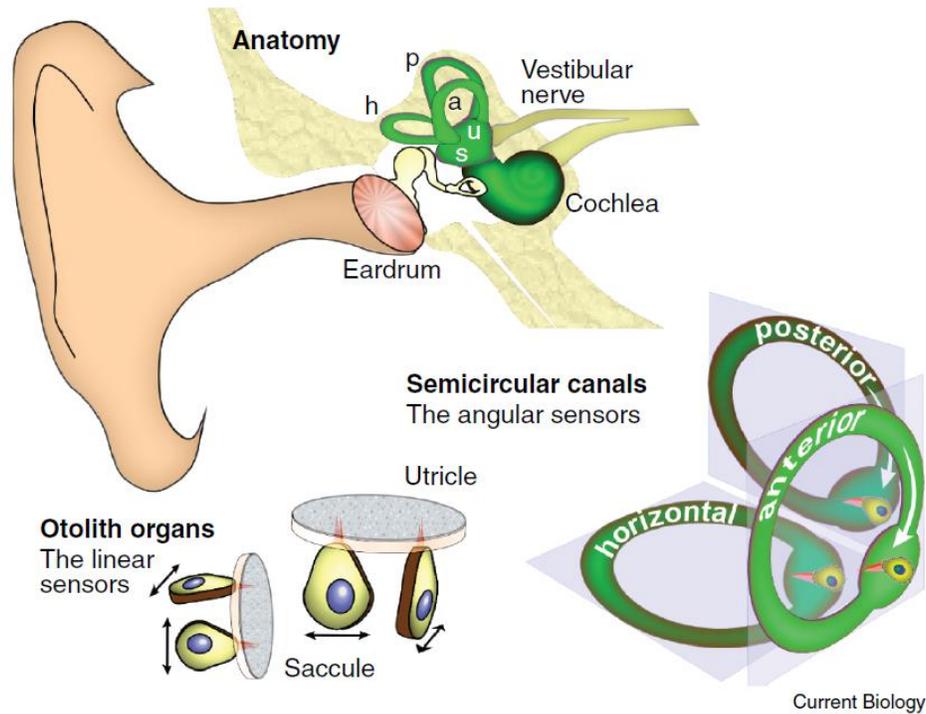


Figure 1.7: Diagram of the vestibular system. The semicircular canals: anterior, posterior, and horizontal canals detect rotational acceleration and the otolith organs: utricle, macula, and sacculle, detect linear acceleration of the head. (from Day & Fitzpatrick, 2005)

Past research had separated non-visual motion cues from visual cues by testing people in dark (Harris et al., 2000; Hlavacka et al., 1996; MacNeilage et al., 2010; Probst et al., 1985). These studies have shown that people can use, based solely on the physical motion cues, to discriminate the direction of linear translations (MacNeilage et al., 2010) and rotations (Probst et al., 1985) although thresholds were larger than when using visual motion cues. People can also estimate their travel distance using physical motion cues when the movement has a constant acceleration (Harris et al., 2000).

While non-visual physical motion cues can be used in self-motion perception, as explained above, any single sensory motion cue can be ambiguous due to their limitations. For example, change in sensory signals from the otolith organs of the vestibular system can be due to head movement or change in body orientation. Also, vestibular system is ineffective when the person is moving in constant speed because it can only detect acceleration. Another example is the sensation of the air flow on the face which can be from the person moving forward or from the wind blowing on the face. For this reason, people rely on more than one sensory cue to judge their self-motion, integrating them for more accurate self-motion perception. In the next section, I will explain how these self-motion cues are believed to be integrated.

1.2.3. Integrating and Reweighing Self-Motion Cues

Self-motion perception is a multisensory process. During a self-motion, people integrate all the motion cues available, including the visual and the non-visual cues, to accurately gauge their self-motion (DeAngelis & Angelaki, 2012; Fetsch et al., 2010; Prsa et al., 2012) and this process of integrating multiple sensory signals (i.e., multisensory integration) is shown in both human and non-human brain (Greenlee et al., 2016; Zhou & Gu, 2023). Multisensory integration of self-motion cues is generally found to improve perception of self-motion (Bayramova et al., 2021; Hlavacka et al., 1996) which may be due to human's ability to integrate sensory information in an optimal way using maximum-likelihood estimation (Ernst & Banks, 2002). Maximum-likelihood estimation (MLE) in multisensory integration estimates the most probable parameter, e.g., travel distance, from multiple sensory cues where each cue is weighted differently based on their reliability (Campos et al., 2014; Saunders, 2014; ter Horst et al., 2015). This method is useful especially when some sensory cues become unreliable, e.g., noisy visual motion signals due to blurry vision, or when there is a conflicting sensory signal, e.g., visually

induced vection indicating self-motion without any physical motion cues. If the conflict is too large some discrepant cues can even be ignored (McManus & Harris, 2023). For example, when a person's whole body (both the body and the head) is tilted it can induce perception of forward and backward translation, but not when only the head is tilted (Crane, 2014).

As explained in this section, integration of the visual and non-visual motion cues leads to more accurate perception of self-motion when cues are consistent. Therefore, more self-motion information available to be integrated should also lead to more accurate spatial updating.

1.3. Cognitive Factor in Human Perception

Based on the review so far, it is evident that the sensory input signals (visual, audio, somatosensory, etc.) from various sensory organs can be used to create a representation of the world around us in our brain (i.e., a bottom-up effect). In this process, some sensory signals contribute to our perception while some are ignored depending on their relevance and reliability (McManus & Harris, 2023; Probst et al., 1985). However, past studies have shown evidence of our cognition affecting visual perception (i.e., top-down effect) as well.

In distance perception, observers judge a person facing them as closer than a person facing away from them despite the fact they are standing at the same distance (Jung et al., 2016). Carrying a heavier physical load on their back (Abe, 2013) or more effort required to move through a distance (Witt et al., 2004) have been shown to increase people's judgement of distance to the target. In size perception, lying down has been shown to result in observers judging a rod to be shorter regardless of the direction of gravity pulling on their body (Kim et al., 2022). In self-motion perception, adding an engine sound enhanced the experience of vection (Väljamäe et al., 2008) suggesting that a learned association (between an engine and motion), i.e., a cognitive process of relating them together, had an effect on the subjective feeling of vection.

People sitting on chair that is known to be able to rotate, feel strongervection compared to when sitting on a fixed chair (D'Amour et al., 2021). When sitting on a oscillating swing, people tend to think that they swung more (larger peak-to-peak displacement) and for longer, even after the swing stopped, when it was suggested to them that the swing will continue oscillating compared to those that were told the swing will stop (Reuten et al., 2022).

These are all cognitive factors that do not contribute directly to the perception of distance, size or self-motion, yet they can still influence observers' responses indirectly through cognitive processes. While the effect of cognition in our perception should be addressed with caution as it is difficult to separate them from other effects such as attention and memory led by the experimenter effect or the instructions given (Firestone & Scholl, 2016; Woods et al., 2009) these observations should not be ignored. Therefore, I will take them into account at various points in this dissertation.

1.4. Aims of this Dissertation

In this dissertation, I evaluate people's ability to visually perceive and remember the position of an object in the environment. Perceiving an object's position accurately requires using various visual information such as the object's visual size, height in the visual field, and its retinal location. Remembering its position accurately after moving requires additional self-motion information such as optic flow. Correct scaling of the space is also important as a distorted perception of space can distort spatial properties such as the relative distance of the surrounding environment which would lead to a misperception of the position of the object relative to the observer. To investigate how these different visual cues influence people's perception of an object's position, I conducted a series of experiments comparing people's judgements of object positions whilst varying the visual information available.

1.4.1. Spatial Updating with a Lateral Translation

Chapter 2 compares people's remembered position of objects at varying eccentricity with and without visual lateral motion in VR. I compared the errors they made after being stationary, after a short translation, and after a long translation to evaluate the effect of translation distance on people's updating ability. I hypothesized that (H1) a longer translation would result in larger updating errors which would suggest that the errors are, at least partially, related to the amount of change in eccentricity of the target created by visual lateral translation and also possibly from a misperceived translation distance.

People's errors were compared to the errors found in a past study (Kim, 2019) which showed no updating error after testing on a MOOG motion platform. I hypothesized that (H2) the errors people make after a short translation, equivalent to the translation distance used in the past study (Kim, 2019), when tested on a regular non-moving chair would differ from when they were stationary, hence an updating error. Such a result would suggest possible cognitive effect of the knowledge of possible physical motion due to sitting on a moving platform on people's ability to accurately update object position after a lateral translation.

1.4.2. Perception of Size and Distance in a 2D Scene Displayed on Computer Screens

Chapter 3 describes three online experiments run during COVID that compared size and position adjusted in a 2D visual scene displayed on computer screens. Participants adjusted a target object's size in accordance with its apparent absolute distance inferred from its position, or its position in accordance with its apparent absolute distance inferred from its visual size in the scene.

In experiment 1, I compared the target sizes participants adjusted to make them appear to be at the simulated distances (final sizes) to the geometrically correct visual sizes for the

simulated distance in the scene (initial sizes). In experiment 2, I compared the target positions (i.e., height in the scene) participants placed them to appear to be at the simulated distances (final positions) to the geometrically correct position for the simulated distances in the scene (initial positions). I hypothesized that (H3) the initial and final sizes and positions would not significantly differ which would indicate that people can reliably use an object's position and size in a 2D scene to infer its apparent distance.

Experiment 3 compared the participant adjusted target sizes and positions with and without familiar objects in the visual scene. I hypothesized that (H4) familiar objects would help in judging object distance in the 2D scene, resulting in the sizes and positions of the target being closer to its true (i.e., geometrically correct) size and position when familiar objects are present.

1.4.3. Perception of Size and Distance in a 2D Scene Displayed on a Large Screen

Chapter 4 describes two in-lab experiments conducted using the same tasks as Chapter 3, but different test setup. The target object was presented on a geometrically accurate 2D scene displayed on a large screen instead of computer screens where participants adjusted a target object's size in accordance with its position, or its position in accordance with its size in the scene. Experiment 1 compared the adjusted sizes and positions of the target object to the results from Chapter 3 and to the geometrically correct target sizes and positions for the simulated distances. I hypothesized that (H5) people would be better at using an object's position and size to judge absolute distance in a geometrically correct 2D scene on a large screen compared to computer screens.

Results from the experiment 3 in Chapter 3, as well as the past studies (Maltz et al., 2021; Yoo et al., 2023), suggest that monocular cues such as familiar size influence people's size perception more than the distance perception. Therefore, I hypothesized that (H6) the additional

distance cues in the new test setup would influence the accuracy in adjusting target to its geometrically correct sizes and positions differently. I also compared the sizes and positions of the target object adjusted while viewing the scene monocularly or binocularly. I hypothesized that (H7) people will be more accurate at estimating the target object's absolute distance from its visual size when viewing the object in a 2D scene monocularly than when viewing binocularly. However, there will be no difference between the viewing conditions when estimating its distance from the object's position in a 2D scene.

In experiment 2, I used a method of continuous tracking method covering a larger range distances to verify the linear relationship between the perceived size and position of the target object on the screen.

2.0. The effect of knowing if you can really move on spatial updating

This chapter is a portion of a manuscript published in PLoS ONE (Kim JJ, Harris LR. (2024)

Updating the remembered position of targets following passive lateral translation PLoS ONE 19(12): e0316469. <https://doi.org/10.1371/journal.pone.0316469>) on which I am the first author and Dr. Laurence Harris is the second author. Exp 1 and 2 of the paper were covered in my Masters thesis. The present chapter describes a new follow up experiment which became Exp 3 of the published paper.

My contribution to the paper includes designing the experiment, programming the tasks, writing up the ethics forms, running the experiment, testing participants, analyzing and interpreting the data and writing up the paper.

Dr. Harris's contributions include helping with the design of the experiment, providing feedback on experimental tasks and interpretations of the results, and editing the paper.

The publication is available at PLoS ONE via <https://doi.org/10.1371/journal.pone.0316469>

2.1. Abstract

Spatial updating, the ability to track the egocentric position of objects during self-motion, is fundamental to navigating around the world. However, people make systematic errors when updating the position of objects after linear self-motion. To determine the source of these errors, I measured errors in remembered target position with or without passive lateral translations. Self-motion was presented visually (simulated in VR) with varying travel distances (0.47 m for short travel and 1 m for long travel distances). People underestimated targets' eccentricity in general even when just asked to remember them for a few seconds (5 seconds), with larger underestimations of more eccentric targets. I hypothesized that updating errors would depend on

travel distance, which was manifested as larger updating errors with long travel distance compared with short travel distance. I also hypothesized that the errors people make after a short lateral translation (0.47 m) would differ from when they were stationary in contrast to my Masters project results, which showed no updating errors when people sat on a MOOG motion platform set up in a room identical to the virtual environment (instead of in an office-like test room) with the equivalent travel distance. The results supported my hypothesis with the errors shifting in the direction of movement after a lateral translation compared to being stationary, which suggest that the efficacy of spatial updating depends on participant's perception of self-motion which is improved by familiarity with an environment and knowing whether a person can move.

2.2. Introduction

Most people can easily navigate around their environment and effectively keep track of their surroundings. To do so they must update the perceived position of surrounding objects relative to themselves as they move. This process is called spatial updating (Riecke et al., 2005; Wang, 2017; Wolbers et al., 2008). Spatial updating is strongly influenced by the observer's perception of the spatial structure of their environment i.e., the egocentric directions and distances of the objects that make up the scene (Burgess, 2006; Wang & Spelke, 2000), and by their self-motion (Burgess, 2006). Knowing about their own movement requires integrating sensory information from visual (optic flow), vestibular, somatosensory and motor systems (Greenlee et al., 2016; Harris et al., 2002; Medendorp, 2011) as well as incorporating internal knowledge of their planned movement, during an active movement. Combining motion cues from these sources allows a person to know about their own movement and hence, theoretically,

to predict the new positions of surrounding, earth-stationary objects relative to them after their movement (Medendorp, 2011).

Any misperception of their travel distance will lead to errors in computing the new egocentric locations of objects after a move. In real life, such errors will usually be corrected by visual information from the environment that allows an observer to derive object locations using allocentric references, i.e., landmarks e.g., one object to the left of another object (Burgess, 2006). To prevent people using such landmarks to derive object locations after a move, previous studies have usually deprived their participants of visual information while moving by conducting the movement in complete darkness (Gutteling et al., 2015; Gutteling & Medendorp, 2016; Klier et al., 2008). However, the visual motion signal (i.e., optic flow) provides important information about both a person's movement and the relative movement of the surrounding environment as a person move through it (Rogers, 2021).

How the position of an object changes relative to an observer is substantially different depending on whether movement is rotational or linear: updating an object's position during linear movement requires more complex computation because only linear movement alters the distance between the observer and object, i.e., the egocentric distance. During rotation, the egocentric distance of objects remains constant. Therefore, the only updating required during rotation is to displace everything by the angle through which the person rotates. During linear movement, however, both the egocentric distance and direction of all the objects in a scene change differently. For example, the egocentric direction of an object will not change at all as a person moves directly towards or away from it. In this special case, the only updating required is to the egocentric distance of the object. When a person updates the position of an object in front of them while moving laterally, both direction and distance to the object change. Therefore,

updating errors made during the movement may be due to mis-updating the direction or distance to the target, or both.

There is no shortage of literature looking into spatial updating as it is a fundamental and necessary ability for people's everyday lives (e.g., Medendorp 2011). Most of these studies look at spatial updating following rotational movements, specifically yaw rotations (Riecke et al., 2005, 2007; Wang & Spelke, 2000) and sometimes forward translation (Klatzky et al., 1998; N. Li & Angelaki, 2005) but there are few studies considering the consequences of linear lateral movement. (Klier et al., 2008) looked at updating after passive linear translation in the fore-aft, up-down and left-right directions but only used targets directly in front of the observer. Previously, in my Masters thesis (Kim, 2019), I conducted experiments using lateral translation and targets at various eccentricities to compare updating performance for objects that start and end at different eccentricities in virtual environment. In general, people underestimated object eccentricity with larger underestimation for more eccentric objects. After a lateral translation the errors shifted in the direction of movement (i.e., updating errors) when participants were visually translated 1m to the left or to the right while sitting on a stationary chair in an office-like test room (Figure 2.1A). However, for the participants translated 0.46 m while sitting on a chair mounted on a MOOG motion platform, set up in a room identical to the virtual environment (see Figure 2.1B and C), their errors did not significantly differ from estimates of target position made when participants remained stationary (Kim, 2019).

There were two differences between these experiments: (1) the travel distance (1m vs. 0.47m) and (2) the test environments (stationary chair vs. moving platform chair). Could these factors have impacted participants' updating, resulting in the differences found between the two experiments? The present experiment was designed to identify which factors could account for

the differences between the results of the past study by using longer distances and a neutral display.

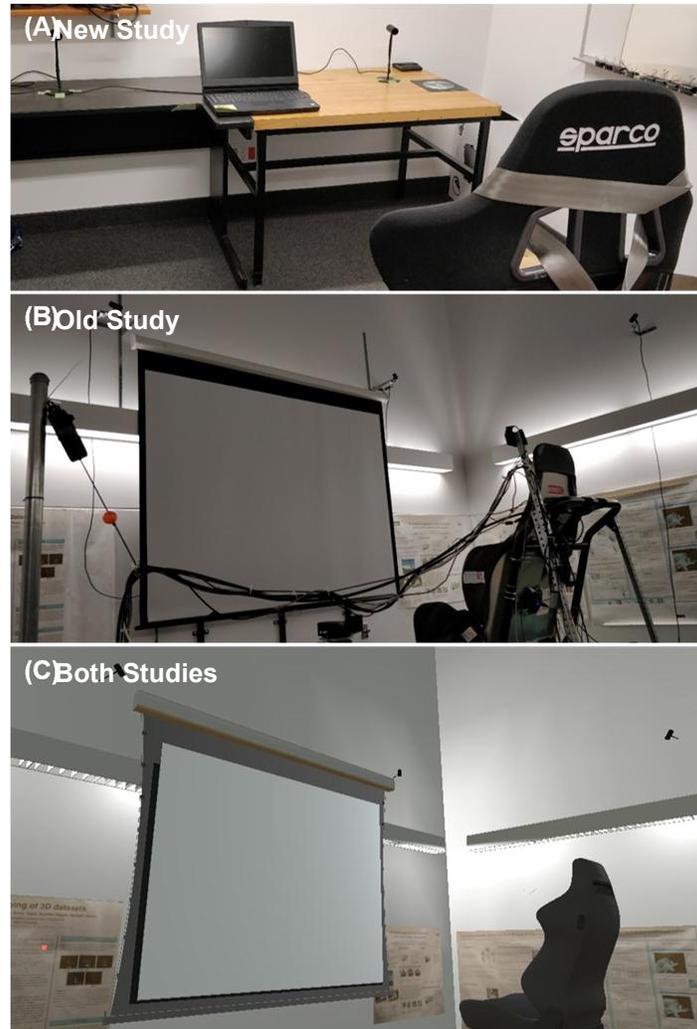


Figure 2.1: The Experimental setup. (A) Photo of the office-like test room with a chair used for the new experiment. (B) Photo of the MOOG test room with a seat on the motion platform used to produce physical motion cues in the old experiment. (C) Example image of the virtual rendition of the MOOG test room used in both experiments with a virtual seat and a screen.

In the new experiment reported here, I compare the two lateral travel distances 0.46 m and 1 m. If the differences in updating errors found in the old study were due to the short travel distance, the errors before and after translation through the short distance (.46 m) would not differ compared to stationery but they might differ for the longer (1 m) distance used in this

experiment. If, on the other hand, the lack of shift in errors were due to the test environment, i.e., exposure to the MOOG room and the moving platform (Figure 2.1B), then the errors before and after translation would differ regardless of travel distance. Here, I compare the errors after short translation with those from the old study to evaluate the effect of test environment.

2.2.1. Hypotheses

To evaluate the effect of travel distance in updating, specifically to compare the different distances used in my old study, 1 m and 0.46 m (Kim, 2019), I varied the translation distance in the present experiment (short vs. long distance). Gutteling and Medendorp (2016) demonstrated that the magnitude of updating errors depends on travel distance, but the longest travel distance they used was 0.18 m (over 3 seconds) which is much shorter than the 0.46 m (over 5 seconds) used in my experiment. I hypothesized that (H1) updating errors would be larger with long travel distance compared with short travel. I also hypothesized that (H2) the errors people make after a short translation, equivalent to the translation distance used in the past study (Kim, 2019), would differ from when they were stationary. This discrepancy from the past results would reveal the cognitive effect of the knowledge of possible physical motion due to sitting on a moving platform on people's spatial updating ability.

2.3. Materials and Methods

2.3.1. Overview

A target ball was presented in virtual reality at various eccentricities. The target then disappeared, and the participant was moved to the left or right. They then had to report the remembered location of the target ball.

2.3.2. Participants

Twenty-six participants (20 females and 6 males, average age of 24.20 years, SD = 11.69), undergraduate students at York University recruited via Undergraduate Research Participant Pool (URPP) participated in the study. They had normal or corrected-to-normal vision and were right-handed. All participants signed a written informed consent form and were given course credit for their participation. The experiments were approved by the York University's Ethics Review Board.

2.3.3. Apparatus

The test was conducted in virtual reality (VR) using an Oculus Rift Head Mounted Display (HMD). The HMD was an Oculus Rift Consumer Version 1 with a resolution of 1080×1200 pixels per eye, maximum 90 Hz refresh rate, 110 degrees FOV, and weighed 470 grams. Participants used a keyboard. The HMD was powered from a Windows 10 computer: Alienware Area-51 R2 equipped with Intel® Core™ i7-5820K Processor, 16.0 gigabyte RAM, and a NVIDIA GeForce GTX 980 graphics card. The VR environment was built in Unity Game Engine (version 2018.3.14f1) and the test was programmed with C# language.

2.3.4. Stimuli

The test environment in VR was a detailed virtual rendition of the actual room in which the MOOG motion platform resides at the Sherman Health Science Research Centre at York University (Figure 2.1C). The virtual room provided a rich environment including a simulated screen with a projector, a seat attached to the MOOG, and an avatar body. Figure 2.1 shows the real MOOG room (B) and the virtual rendition of the room (C). An image of a tennis ball, diameter ≈ 67 mm, was used for the updating test and was presented on the simulated screen

seen in Figure 2.1C. Target locations were spread out horizontally from the center of the screen at eye height. The targets were shown one at a time for each trial.

The experiment took place in an office-like test room (Figure 2.1A) with participants wearing an HMD in which the virtual test room was simulated (Figure 2.1C). The targets were between -1.15 m to +1.15 m from the center of the screen in 0.23 m increments (11 total) and were displayed one at a time, on the virtual screen (1.5 meters x 3.5 meters) 1.75 m away from the viewing camera. The observer was always initially positioned in front of the target at 0m (center of the screen). In the visual motion conditions, they were moved passively either to the left or to the right side (visually only) for 5 seconds. Two translation distances were used: 0.46 m for the Short Travel condition and 1 m for the Long Travel condition. During translations, random dots appeared on the screen which were made to blink between 0.5 to 1 hz. In the response phase, a red dot was displayed on a black wall at the same distance as the screen, which participants moved to indicate their remembered target position on the screen.

2.3.5. Training

Every participant went through a training session at the start of the test session to become familiar with using the keyboard. They were shown the visual target (an actual size image of a tennis ball) on the screen. They had to move a red dot to the position of the ball (always visible) using the left/right arrow keys on the keyboard. Once the red dot was on the target, they pressed a button to “hit” the target. If they pressed the button when the red dot was not on the target, it was considered a “miss”. The training was repeated for 10 trials. Participants had to hit at least 8 out of the 10 targets to continue.

2.3.6. Procedure

Participants sat on a chair in an office-like test room (Figure 2.1A). Then they put on the HMD in which they viewed the virtual environment that simulated the MOOG room and the stimuli (Figure 2.1C) and used a keyboard to respond in each trial. During each experiment, participants were instructed to look straight at the simulated screen in the virtual environment. For each trial, a bell sound was played to let participants know the trial was starting. A) Then a fixation cross was displayed in the middle of their visual field which they were told to fixate. B) After 0.5 seconds, a visual target (the same image of a tennis ball from the training) appeared for 0.5 seconds at one of the pre-selected target eccentricities. C) After the visual target disappeared, randomly generated blinking dots were shown on the screen. D) During the translation phase, participants were either stationary (Stationary condition) or the display with the blinking dots was moved laterally to the left or right (Visual Motion conditions) providing a strong optic flow cue for 5 seconds. The observer always started from the center of the screen, then moved either to the left or right. E) After the movement, or after an equivalent idle period in the stationary condition, the screen went dark, and a second bell sound was played. When the participants heard the second bell, they placed the red dot at the remembered target location with the keyboard (pressing on the up, down, left, or right arrow keys moved the red dot on the screen). When the red dot was at the position where the participant thought the target was, they pressed the return key or a button on the controller to indicate their choice (see Figure 2.2 for the experiment timeline). The pointing location (red dot) was recorded as indicating the participant's remembered target location.

There were three experimental blocks: Stationary, Short Travel, and Long Travel conditions. One block consisted of 3 trials for each target location at each initial observer

position, comprising 66 trials ($3 \text{ trials} \times 11 \text{ target locations} \times 2 \text{ translation direction}$) per block, thus 198 total trials ($66 \text{ trials} \times 3 \text{ blocks}$) per participant (~ 1.5 hours including breaks).

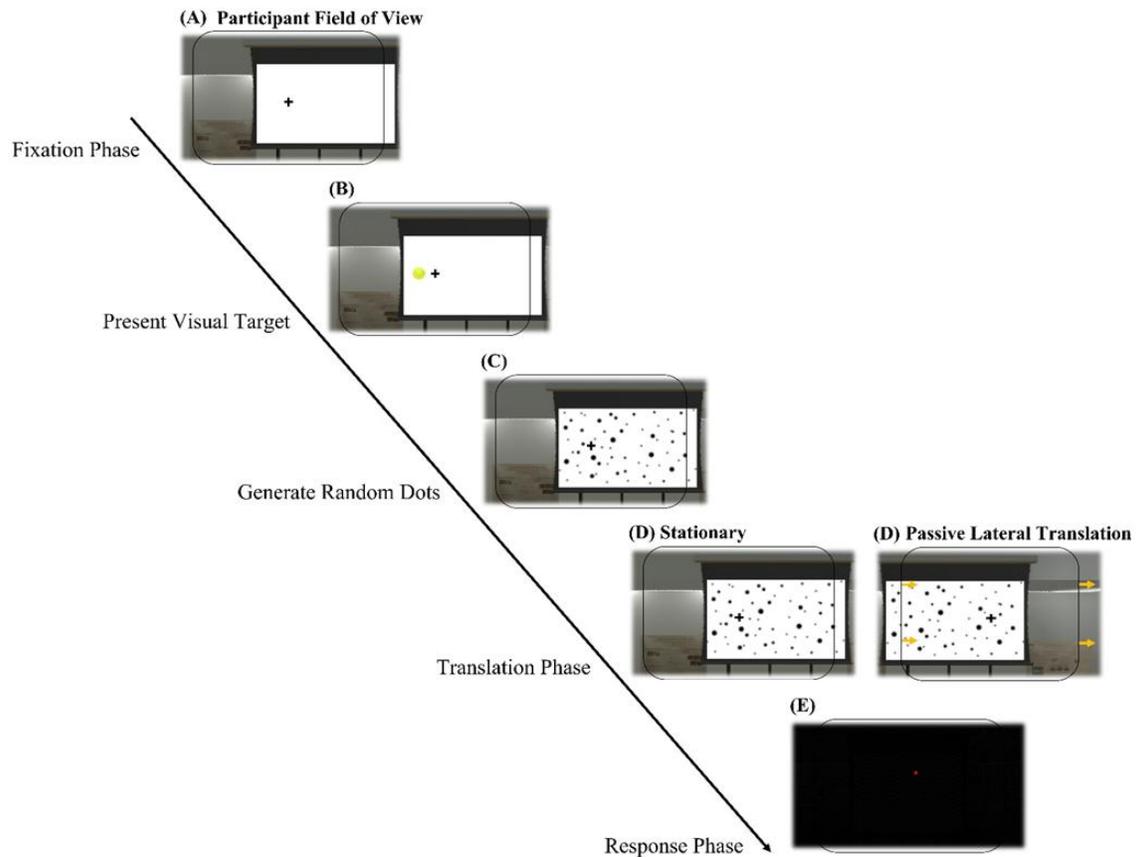


Figure 2.2: Experimental procedures. (A) Participants fixated on the fixation cross for 500 ms. (B) Visual target presented for 500 ms. (C) Randomly generated blinking dots appeared on the screen after the visual target disappeared. (D) Participants were either stationary or passively moved laterally for 5 seconds. (E) The screen turned dark. Participants then moved the indicator to the remembered location of the visual target as fast as they could.

2.3.7. Data analysis

In the data cleaning process, we treated any error value deviating more than 2.5 standard deviations from the mean (for each condition) as a mistake by the participant (e.g., missing the target during presentation or forgetting the position during movement) and removed those points before calculating the average error for each participant. If all the responses for any condition

were removed, that condition for that participant was removed from the analysis. The same process was used in all three experiments.

The data from the two observer start positions (Left and Right) were brought to the same side where the positive (+) values represented errors in the direction towards the opposite side of the observer position for Stationary condition (e.g., errors to left were positive if the observer was on the Right side) and in the direction of the movement for the Visual Motion conditions (e.g., errors to the left were positive if the observer was moved from right-to-left). Only horizontal errors were considered since the movements were in lateral directions only which were calculated by subtracting the actual target position from the remembered target position.

The data cleaning process removed 166 data points from the total data points (3.22 %). As a result, three participants had to be removed from the analysis due to having an incomplete data set leaving 23 participants for the analysis. I performed a two-factors repeated measure ANOVA with a Greenhouse-Geisser correction to compare the effect of motion conditions (Stationary, Short Travel, and Long Travel) and target position (collapsed between the observer positions as described above) on the errors in participants' remembered target positions and used a family-wise alpha of .05.

2.4. Results

Participants placed a red dot on a virtual screen simulated in VR to indicate a target's perceived position after various sideways movements. Errors were calculated by measuring the deviation between the remembered and actual target positions.

Figure 2.3 shows participants' responses for the Stationary, Short Travel, and Long Travel conditions. An ANOVA revealed there was a significant main effect of motion condition, $F(1.451, 31.917) = 15.896, p < .001, \eta^2 = .163$, where errors in both Short ($M = 0.102$ m, $SD =$

0.291 m) and Long travel conditions ($M = 0.266$ m, $SD = .432$ m) significantly differed from Stationary ($M = -0.015$ m, $SD = .210$ m), $p < .001$. A significant main effect of target position was also found, $F(2.085, 45.875) = 26.165$, $p < .001$, $\eta^2 = .184$. There was a significant interaction between travel distance and target position, $F(7.186, 158.097) = 3.950$, $p < .001$, $\eta^2 = .018$.

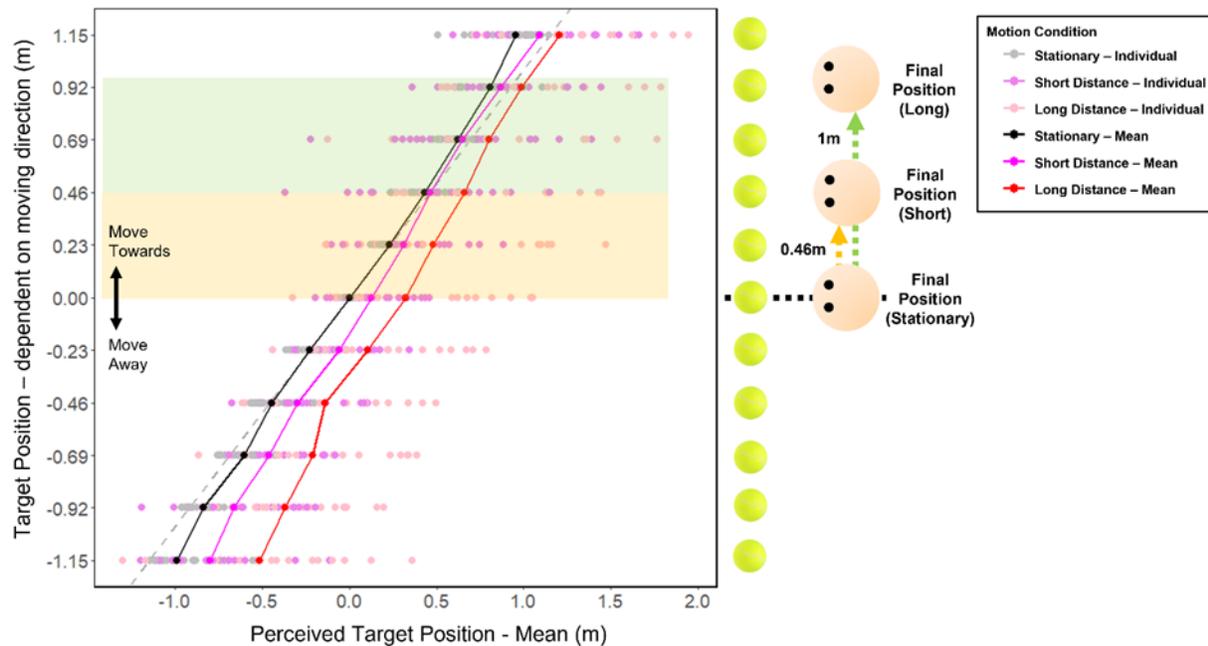


Figure 2.3: The positions at which participants remembered the targets. Black - Stationary, Purple - Short (.46 m), and Red - Long (1 m) travel conditions. The data from the rightward movement directions have been flipped to superimpose on those to the left where positive (+) values represent errors in the direction of observer's movement. The lighter colored dots represent remembered positions from each participant ($N = 23$) and the darker colored dots and lines represent the means. The dashed line represents true target positions, therefore the deviation from the dashed line to the remembered target position is the error made. For both x- and y-axes, positive (+) represents in the direction of the movement and negative (-), the opposite direction. The orange and green shaded area contains the targets within the start and the end positions (orange from 0 m to 0.46 m, and green from 0m to 1m) between which the observer traveled in the moving conditions.

Post hoc evaluation of paired t-test, with Holm-Bonferroni correction showed that the mean errors made for each target position differed significantly between motion conditions, where the errors shifted after moving compared to Stationary in the direction of travel for both

Short ($p = .039$) and Long ($p < .001$) Travels conditions (0.1 m for short and 0.27 m for long travel on average). The errors also differed between the travel distances ($p = .002$), in which longer travel distances resulted in larger errors.

2.5. Discussion

By asking participants to remember the positions of targets following passive lateral translation or after a comparable period with no movement, we were able to assess the contributions of target eccentricity, and translation distance on spatial updating performance. Participants underestimated target eccentricity (slope > 1 in Figure 2.3) even when they did not move at all between when they were shown the targets and when they reported their location (a delay of 5 seconds), with the magnitude of their errors depending on target eccentricity. This was consistent with the results from my old study (Kim, 2019). The errors shifted (in addition to this drift in memory) after lateral translation (by 0.1 ± 0.06 m for Short and 0.27 ± 0.10 m for Long Travel conditions on average) in the direction of their translation.

These shifts in errors portray a change in the observer's mental representation of all the targets seen before they were moved, i.e., updating errors. As hypothesized (H1), updating errors, i.e., the shift in errors after being moved laterally, were larger with longer translation than with shorter translation suggesting that updating errors depends on the amount of change in eccentricity between the target and the observer. The source of this error can be the underestimated object eccentricity (as shown in the stationary condition results), underestimated travel distance, or combination of both.

The errors participants made after moving, when pointing at remembered target positions, significantly differed compared to those made after they remained stationary for both the short and long travel distances which also supports our hypothesis (H2). These results demonstrate that

people indeed make updating errors after a lateral translation of 0.47 m in the present experiment which shows that the lack of shift in errors found in my Masters project was likely due to the exposure to the MOOG room and moving platform, not the shorter travel distance.

2.5.1. Comparing the results to the previous experiment

Figure 2.4 compares the errors found in the old experiment (where the participants sat on a MOOG motion platform where they knew chair could move) to this new experiment where participants were sitting in a regular lab on a chair which was known not to be able to move for the corresponding target eccentricities. This comparison further reveals that the lack of updating errors in the old study were likely due to differences in the simulated rooms and known mobility of the chairs in which they were tested. In the old study participants were tested in a physical room which the virtual environment mimicked convincingly. Although the experiment was done wearing an HMD, participants in the old study had prior experience of the room as they climbed into the chair (Figure 2.1B) with a clear view of the room and the motion platform on which they sat.

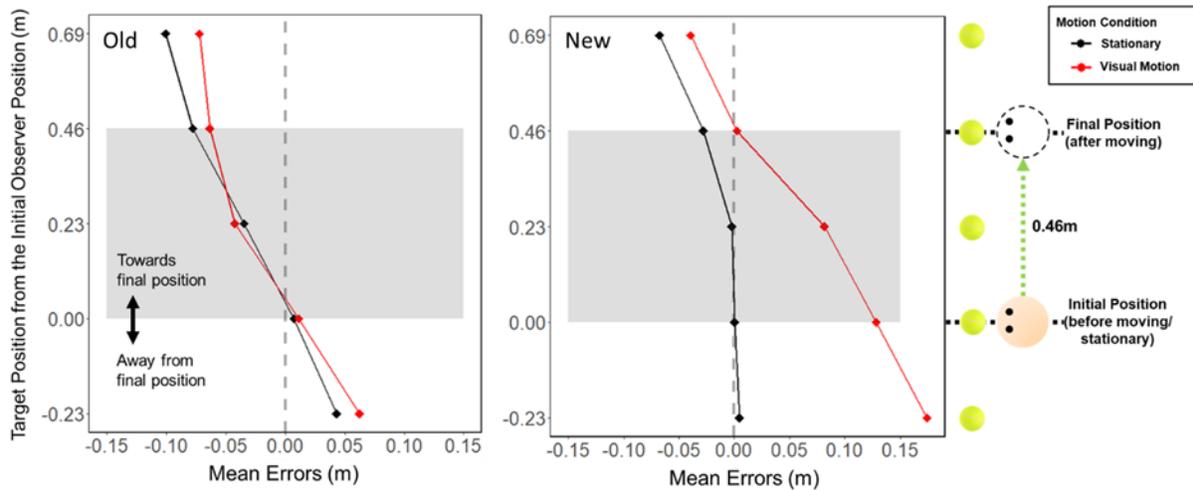


Figure 2.4: Comparison between old and new. Mean errors participants made for each target for Stationary (Black) and Visual Motion (Red) conditions for old and the equivalent targets from new, where 0 on the vertical axis is the start position and collapsed across observer positions: Left and Right. For both x- and y-axis, positive (+) represents in the direction of the movement, and negative (-) the opposite direction. The faded area contains the errors for the targets from the start and the end position (from 0 to .46m) where the observer traveled in the Visual Motion conditions.

It is well known that objects are perceived as closer, i.e., distance compression, in virtual reality (Peer, 2017), but a simple experience of viewing a real-world environment can improve distance perception in a matching virtual environment (Kelly et al., 2018). Distance compression can even disappear when the viewer had previous experience interacting with a matching real-world environment (Interrante et al., 2006). My participants' exposure to the MOOG may have helped them perceive the virtual environment accurately thus reducing the underestimation of target distance and resulting in a more accurate representation and scaling of the virtual environment. If updating errors found in the present experiment were due to distance compression in VR, however, the errors should have shifted in the opposite direction of translation (see Figure 2.5). Instead, the updating errors were in the direction of translation which suggest that the disappearance of updating errors in the old study was due to a better distance perception.

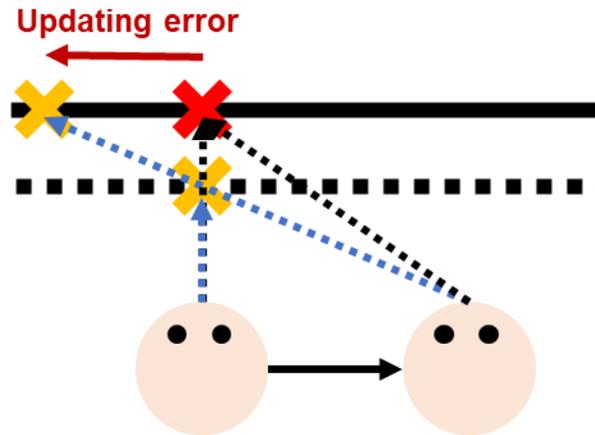


Figure 2.5: Impact of distance compression on updating position of an object after a lateral translation. Underestimated absolute distance of a target results in the viewer overestimating the shift in target position after being translated laterally, remembering it to be more eccentric than it is after moving. Therefore, updating error will be in the opposite direction of the viewer's translation.

Sitting on a chair that can move, even when it doesn't actually move, has been shown to enhance vection (D'Amour et al., 2021; Riecke et al., 2005). Such an enhancement may then lead to a more accurate perception of travel distance compared to sitting on a chair that participants knew could not move. Based on these considerations, it seems that even potential or imagined physical motion cues can play an important role in our spatial updating. Even when there is no actual physical motion involved, knowing that one can move can still improve one's perception of visually induced self-motion and make spatial updating more accurate.

2.6. Conclusion

When a person observes an object, their perception of its position is biased towards their straight ahead, leading to underestimating its eccentricity. Correctly gauging self-motion is important when updating an object's perceived position during and after movement. The results in this experiment, compared to the findings from the old study in my Masters thesis, show evidence that familiarity with an environment and knowing whether a person is free to move

(e.g., if sitting on a moveable platform) both improve updating. This improvement appears to be from enhancement in their perception of self-motion.

3.0. Can People Infer Distance in a 2D Scene Using the Visual Size and Position of an Object?

This chapter was published on Vision (Kim J, Harris LR (2022) Can People Infer Distance in a 2D Scene Using the Visual Size and Position of an Object? Vision 2022, 6, 25). I am the first author of the paper and Dr. Laurence Harris is the senior author.

My contribution to the paper includes designing the experiment, programming the tasks, writing up the ethics forms, running the experiment, testing participants, analyzing and interpreting the data and writing up the paper.

Dr. Harris's contributions include helping with the design of the experiment, providing feedback on experimental tasks and interpretations of the results, and editing the paper.

The publication is available at Vision via <https://doi.org/10.3390/vision6020025>

3.1. Abstract

Depth information is limited in a 2D scene and for people to perceive the distance of an object, they need to rely on pictorial cues such as perspective, size constancy and elevation in the scene. In this study, I tested whether people could use an object's size and its position in a 2D image to determine its distance. In a series of online experiments, participants viewed a target representing their smartphone rendered within a 2D scene. They either positioned it in the scene at the distance they thought was correct based on its size or adjusted the target to the correct size based on its position in the scene. In all experiments, the adjusted target size and positions were not consistent with their initially presented positions and sizes and were made larger and moved further away on average. Familiar objects influenced adjusted position from size but not adjusted size from position. These results suggest that in a 2D scene, (1) people cannot use an object's

visual size and position relative to the horizon to infer distance reliably and (2) familiar objects in the scene affect perceived size and distance differently. The differences found demonstrate that size and distance perception processes may be independent.

3.2. Introduction

The relationship between visual size and distance has been obvious to humans for a long time: as an object moves further away, its visual size shrinks in a systematic way. Size constancy, the ability to recognize that an object is the same size despite its visual size changing as its distance from the observer changes, is present from a very young age (Cruikshank, 1941; McKenzie et al., 1980), even at birth (Slater et al., 1990). It seems likely that the concept is even hardwired in our brains (Chen et al., 2019; Sperandio et al., 2015) as knowing the size and distance of objects is very important to our survival. The primary visual cortex (V1) serves as an important node in the integration of signals from various sources to produce stable perceived size (Sperandio et al., 2015). Chen et al. (2019) observed stronger neural responses in V1 to stimuli at far distances compared to stimuli at near distances for objects with the same retinal image size, indicating that human V1 can distinguish the distance and size of an object. Such activity is possibly regulated by feedback of distance information from other higher-order areas (Granrud, 2009).

Although size constancy may be innate, accurate size estimation requires the development of cognitive ability and the use of strategies such as the distance compensation strategy: deliberately inflating one's size estimate to compensate for a reduced perceived size at a far distance that supplements size perception (Viguier et al., 2001). In recent centuries, since the discovery of perspective in art in the 1400s, people have been exposed to images that represent depth on a 2D platform (e.g., paintings, photographs, or monitor screens). Such images use

various techniques to simulate natural cues to depth such as texture gradient, perspective, haze, and size constancy. The purpose of this study was to test whether people can determine the egocentric distance of an object in a 2D scene.

3.2.1. Representing depth in a 2D scene

In a 3D world, stereopsis is one of the most reliable indications of depth for a person for relatively close objects (Viguier et al., 2001). The disparities between the views in the two eyes tell the observer how far away an object is from them (Wheatstone, 1838). The reliability of stereopsis quickly declines with distance as the disparities become too fine to detect although the range of distance over which stereopsis is useful is debatable (Palmisano et al., 2010). Stereopsis is, of course, not at all useful when determining the relative depicted distances of items in a 2D platform as those images of items are presented on a single surface; hence, the simulated distance of different items in such a display must be represented using other depth cues. Here, I investigated the usability of the two depth cues of ‘visual size’ and ‘distance below the horizon’ within a 2D scene.

As mentioned above, visual size is a useful feature to represent the distance of an object in a 2D scene taking advantage of size constancy. Objects far away are presented smaller in angular size compared to closer objects (see Figure 3.1). To determine the absolute distance of an object from its visual size, however, the viewer must know the real size of the object. For example, the sizes of the running person, the dog, and the bird in Figure 3.1 are identical. However, typically, a person is larger than a dog and a dog is larger than a bird. So, it appears that the running person is the furthest from the viewer and the bird is the closest in the scene. If the size of the object were ambiguous - for example, the plane in the scene could be a real

airplane and large or a toy airplane and small - then size cannot be used to determine how far away it is.



Figure 3.1: Example of a 2D scene with objects at various distances. Distances are represented with the objects' (1) size —objects of smaller visual size (e.g., the running person and the tree on the right side) are perceived as further away compared to the same objects of larger visual size (e.g., the standing person and the tree on the left of the figure); and their (2) distance below the horizon —objects placed closer to the horizon (the running person and the tree on the right side) are perceived as further away compared to objects further below the horizon (e.g., the standing person and the tree on the left side). The ground plane and the horizon play big roles in depth perception in this example. The running person, the dog, the airplane, and the bird in this scene are all the same visual size. When comparing the person and the dog, it appears obvious the person is further away. However, when comparing them to the bird or the plane, it is ambiguous whether the bird/airplane is: (i) flying at a distance close to the viewer and is smaller in physical size or (ii) flying at a far distance but is larger in size.

Another way distances of objects can be represented in a 2D scene is to use the position, i.e., the elevation in the visual field. People can use angular declination below the horizon to determine the absolute distance of an object (Ooi et al., 2001). In a 2D scene, this can easily be implemented with the objects placed at different elevations. An object placed near the horizon

provides a smaller angular declination to the viewer, appearing to be far from them compared to an object placed further below the horizon, appearing to be near (see Figure 3.1). However, this is only a useful cue for the objects on the ground and not those that are floating or above the horizon. For example, the plane in Figure 3.1 is the same visual size as the running person and the dog, but it is ambiguous as to whether it is a small plane flying at a close distance or a large plane flying at far distance. Additionally, although the bird in the scene is at the same distance below the horizon and the same visual size as the dog, it is perceived as closer to the viewer than the dog because it is flying (i.e., not on the ground).

3.2.2. Quantifying depth in a 2D scene.

These depth cues, the elevation and the visual size, representing the distance of an object on a 2D platform are quantifiable using basic geometry. Figure 3.2 presents an example of a target (with a height of H_{object}) at a distance (D_{object}) on the ground represented on a screen placed at a certain distance (D_{screen}) away from the observer's eyes (at the height H_{eye}). The distance of the object below the horizon on screen (d_{bottom}) and its height on screen (h_{object}) can be computed using these parameters (see Figure 3.2). Hence, theoretically, the distance between the observer and the target (D_{target}) should also be determinable from the object's position on the screen, i.e., its distance below the horizon (P ; where $P = d_{bottom}$):

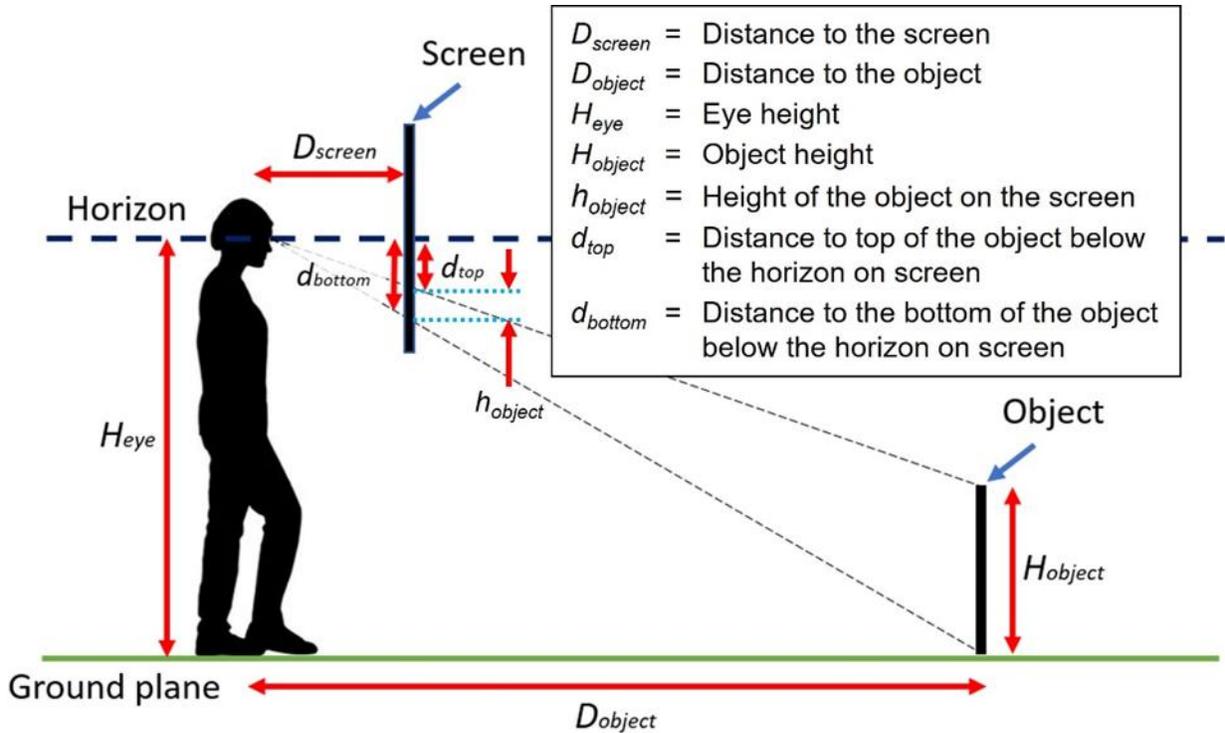


Figure 3.2: Geometry of an object represented on a 2D platform (the screen). The elevation of the object on the screen can be calculated with: $d_{bottom} = \frac{D_{screen} \times H_{eye}}{D_{object}}$, and the height of the object on the screen can be expressed as h_{object} or: $(d_{bottom} - d_{top})$, where $d_{top} = \frac{D_{screen} \times (H_{eye} - H_{object})}{D_{target}}$. Hence, $h_{object} = d_{bottom} - d_{top} = H_{object} \times \frac{D_{screen}}{D_{object}}$.

$$D_{object} = \frac{D_{screen} \times H_{eye}}{p} \dots\dots\dots (3.1)$$

Based on the dependency of the perceived size of an object on size constancy and Emmert’s law (Irwin, 1969), the relationship between the object displayed on a 2D screen and the real object can be represented as (as shown in Figure 3.2):

$$S = \frac{h_{object}}{H_{object}} = \frac{D_{screen}}{D_{object}} \dots\dots\dots (3.2)$$

where the size of the object on the screen (S) is expressed in proportion to the real target size. Therefore, I can determine the object distance (D_{object}) from its size on screen in proportion to the true object size (S):

$$D_{object} = \frac{D_{screen}}{S} \dots\dots\dots (3.3)$$

Throughout this paper, the distance below the horizon of the object on screen (i.e., d_{bottom} from Figure 3.2 in mm) will be referred to as the ‘position’ of the object (or target) on the screen (P) and the ‘size’ of the object (or target) on the screen will be expressed as its size as a proportion of its true physical size (S) - see Equation (3.2).

3.2.2. Hypotheses

In the present study, I evaluated whether people are able to use an object’s position (P) and/or its size rendered in a 2D scene (S) to estimate an absolute distance. The perceived distance of the target (of a familiar size) was measured indirectly by asking participants to either change a target’s position (P) or its size (S) to what they perceived to match their smartphone – an object with which they were intimately acquainted – in the scene standing on the ground at a distance (D_{object}). Perceiving correct distances and sizes requires that the perceived scale of space which shrinks with distance (Boring, 1964) and past studies have typically found systematic errors in egocentric distance judgements (Z. Li et al., 2011; Viguier et al., 2001). Past models such as Stevens’ (1957) power law (with the exponent 0.67; Table 1) as well as Gilinsky’s (1951) formulas for perceived distance (Equation I) both show perceived distance is reduced compared to the real distance. Therefore, it does not seem likely that people can determine the correct absolute distance of an object from a 2D scene which only provides limited cues to distance. Rather, I expect people to underestimate distance compared to the object distance I simulate in the scene in this experiment. However, if people could infer a distance (apparent distance) from position and size using their relationship as shown in Figure 3.2, then the adjusted positions and sizes of the targets should be consistent.

Here, the participants performed two tasks consecutively. They first adjusted target position while its size was fixed (Size-to-Position task) or adjusted target size while its position was fixed (Position-to-Size task). They then adjusted the target size while its position was fixed to the size that they had previously set in the first task or adjusted its position while its size was fixed to the value set in the first task. I hypothesized that (H3) people would be able to reliably use a target's position and size in a 2D scene to infer its apparent distance and hence be able to adjust target's size and its position to the values at which it was first presented to them in the first task. I also hypothesized that (H4) familiar objects in a 2D scene would improve people's perception of distance of the objects in the scene, resulting in the adjusted target sizes and positions to be closer to the geometrically correct sizes and positions for the simulated distances. To test these hypotheses, I conducted the following experiments.

3.3. Materials and Methods

Three online experiments were conducted following guidelines approved by the York University's Ethics Review Board and were carried out using the principles of the Declaration of Helsinki. The participants were recruited from the York Undergraduate Research Participant Pool (URPP) and received course credit for their participation. They had to be between 18 and 45 years old to be able to sign up for this study online, but the average age and number of male/females are unknown due to many participants keeping that information hidden from the researcher. All participants gave informed consent before taking part in the experiment.

3.3.1. Apparatus

All experiments were designed and programmed using PsychoPy3 that was run on Pavlovia – an online platform to run experiments. Participants followed a web link provided when they signed up for this study, on their own computer (laptop or desktop). I wanted to

simulate an object of a familiar size as the target. Therefore, participants used their personal smartphone as a reference size to compare to visual renderings of their phone shown on the screen. Because the sizes of participants' screens and their smartphones varied, I needed to know the screen and smartphone sizes for each participant. The participants performed the following tasks at the beginning of the experiment to provide the information to allow us to estimate each participant's screen and phone size.

3.3.1.1. Enter phone size.

Participants entered their phone size by physically placing their smartphone on their screen over a black rectangle and then adjusting the rectangle size using sliders—width and height (see Figure 3.3A). When the rectangle matched their phone size, the participants pressed the spacebar to continue to the next step. The size of the rectangle was saved as the full size of their smartphone. Because the experiment was performed online and participants used their own screens, the screen sizes varied. For this reason, the experiments recorded the phone height and width as a proportion of the height of the screen (using the height units in PsychoPy3 where the full height of the screen = 1).

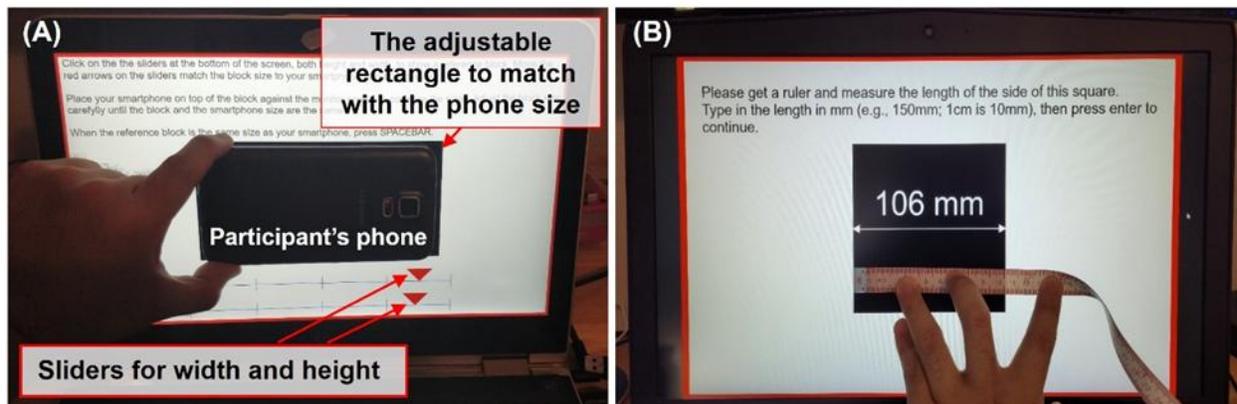


Figure 3.3: Entering reference size. (A) Example photo of the phone size setting stage at the start of the experiment. The width (W_{phone}) and the height (H_{phone}) of the phones were recorded as a fraction of the height of the screen. (B) Example photo of the reference length ($H_{reference}$) setting stage.

3.3.1.2. Enter reference length.

Participants measured the length of a black square (each side was $\frac{1}{2}$ of the height of their personal screen) shown on the screen and entered it in mm (see Figure 3.3B). The participants then pressed the enter key to continue to the next step. This provided us with the actual size of their screen.

3.3.1.3. Estimating participant screen size.

Combining the phone size and the reference length entered, I computed each participant's screen height. I estimated their smartphone sizes by averaging the most popular smartphones in Canada for 2019 – height = 143.3 mm and width 69.4 mm. First the phone size, width (W_{phone}) and height (H_{phone}) entered (in proportion to the screen height) were divided by the estimated real phone size and averaged to compute the first estimate of the screen height ($H_{screen1}$):

$$H_{screen1} = \left(\frac{W_{phone}}{69.4} + \frac{H_{phone}}{143.3} \right) \times \frac{1}{2} \dots\dots\dots (3.4)$$

Because I only asked participants to enter the phone sizes in Exp 1, I used $H_{screen1}$ as the estimated screen sizes for each participant. Then, in Exp 2 and 3, reference length was collected in mm. Since the reference length ($H_{reference}$) was $\frac{1}{2}$ of the height of the screen, I multiplied the entered length by 2 to compute the second estimate of the screen height ($H_{screen2}$):

$$H_{screen2} = 2 \times H_{reference} \dots\dots\dots (3.5)$$

To minimize the disparity between experiments, two estimated screen heights were compared to verify any possible errors and then averaged to compute the final estimate of the screen height for Exp 2 and 3 (H_{screen}):

$$H_{screen} = \frac{H_{screen1} + H_{screen2}}{2} \dots\dots\dots (3.6)$$

3.3.1.4. Visual stimuli.

The scene displayed was a 2D rendering of a long, grassy path, open to the sky and with stony walls on each side. The scene was created on Unity using a camera view 1.7 m above the ground. The scene contained a grass-textured floor and stone walls on the left and right sides (Figure 3.4A). The scene was designed to provide limited distance information (i.e., stones on the walls are of unfamiliar shape and size, and there are no other objects in the scene other than the target) so that the observer could focus on the target itself when determining its size/distance. Because the screen sizes and ratios varied between participants, the width and the height of the scene was kept at a 9:5 ratio, where the height of scene rendered on the screen was always the full height of the screen. For most screens, this produced a full screen image.

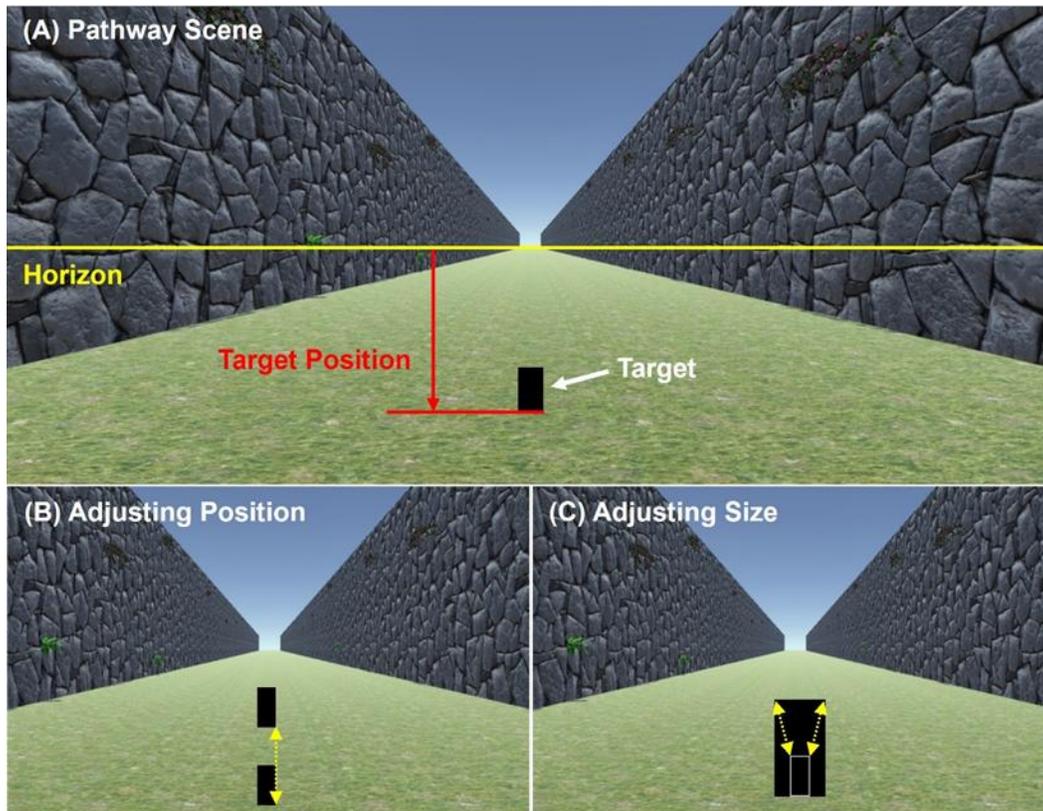


Figure 3.4: Visual stimuli and tasks. (A) The pathway scene with an example of a target (black rectangle) displayed in the scene. The yellow horizontal line is the horizon in the scene – the midline of the image. The target position is the distance from the horizon to the target base (red line). Participants were instructed to imagine themselves standing in the pathway looking at their smartphone standing on the ground. (B) During the Size-to-Position task, the position of the target could be moved upward by moving the mouse up, and downwards by moving the mouse down. (C) During the Position-to-Size task, the target size could be made larger by moving the mouse up, and smaller by moving the mouse down.

The target was a black rectangle positioned along the center line of the scene. Participants were asked to imagine that the target was their personal smartphone standing on the ground at some distance from them. Depending on the task (see below), the target could be made smaller/larger or moved up/down in the scene using a mouse.

3.3.2. Tasks

Having calibrated the participant's equipment with the Enter Reference Length task and the Enter Phone Size task, they then performed two tasks in order: a Size-to-Position task in which they set the position (P) of a rendering of their smartphone of fixed size, and a Position-to-

Size task in which they set the size (S) of their smartphone rendered at a given position. The order of the tasks was varied for each experiment. At the start of the experiment, and throughout the experiment, participants were instructed to keep their eyes 0.5 m from the screen.

3.3.2.1. The size-to-position task.

During the Size-to-Position task, a pathway scene and a target was displayed on the participants' screen (see Figure 3.4A). Target size (S) was fixed, but participants could adjust its vertical position (P). Five different target sizes were used for each test. For each target size, the participants moved the target up or down using their mouse (see Figure 3.4B) to position it where they perceived it to be at the correct position for that size rendering of their smartphone in the scene. When they thought the target's size and position were correct, they pressed the spacebar to continue to the next trial.

Target positions were recorded as a fraction of the height of the screen. The full height (bottom to the top of the screen) was defined as 1 unit, where $[0, 0]$ represented the center of the screen (see Figure 3.4A). $[0, 0.5]$ represented the top, and $[0, -0.5]$ represented the bottom of the center line. The horizon extended from $[0.9, 0]$ to $[-0.9, 0]$ in this notation, hence keeping the 9:5 ratio between the width and the height of the scene.

3.3.2.2. The position-to-size task.

The pathway scene and the target displayed on the participant's screen were identical to as they were in the Size-to-Position task (Figure 3.4A). The target position (P) was fixed but its size (S) could be adjusted. Five different target positions were used for each test. For each target position, participants had to adjust the target size by moving their mouse up and down (Figure 3.4C). When they perceived the target to be the same size as their smartphone placed at that distance, they pressed the spacebar to end the trial.

The adjusted target sizes were recorded as a fraction of the participant's actual phone size entered at the Enter Phone Size task, where a value of 1 represented the full size of the phone. For example, a target size of 0.5 corresponds to the target being set to the half of the actual size of the phone (in both height and width) rendered according to the geometry of Figure 3.2.

3.3.3. *Methods for experiment 1*

3.3.3.1. **Participants.**

Out of the 150 people that signed up for the experiment, 122 completed the experiment. After post-test data cleaning (details below), 40 participants had to be removed, leaving 82 participants for the analysis.

3.3.3.2. **Procedures.**

Experiment 1 started with Enter Phone Size task to obtain the phone size to be used to determine target size throughout. It was followed by the Size-to-Position task, where I measured each participant's judgement of where each target should be positioned in the scene based on the target size presented. The target size was equivalent to the visual size of their phone (based on the phone size from the Enter Phone Size task) when seen at each of the five distances. The visual sizes were determined using the formula $\frac{D_{screen}}{D_{object}}$, where D_{screen} is the viewing distance (0.5 m) and D_{object} is the simulated target distance, 4, 8, 12, 16, and 20 m (see Figure 3.2).

After completing the Size-to-Position task, the participant then continued to the Position-to-Size task (see Figure 3.5 for the task order for Exp 1). I considered the positions where the participant placed targets in the Size-to-Position task to be the position at which they perceived the target as the same size as their smartphone. Therefore, the targets in the Position-to-Size task were presented at the average positions recorded from the Size-to-Position task for each intended target distance.

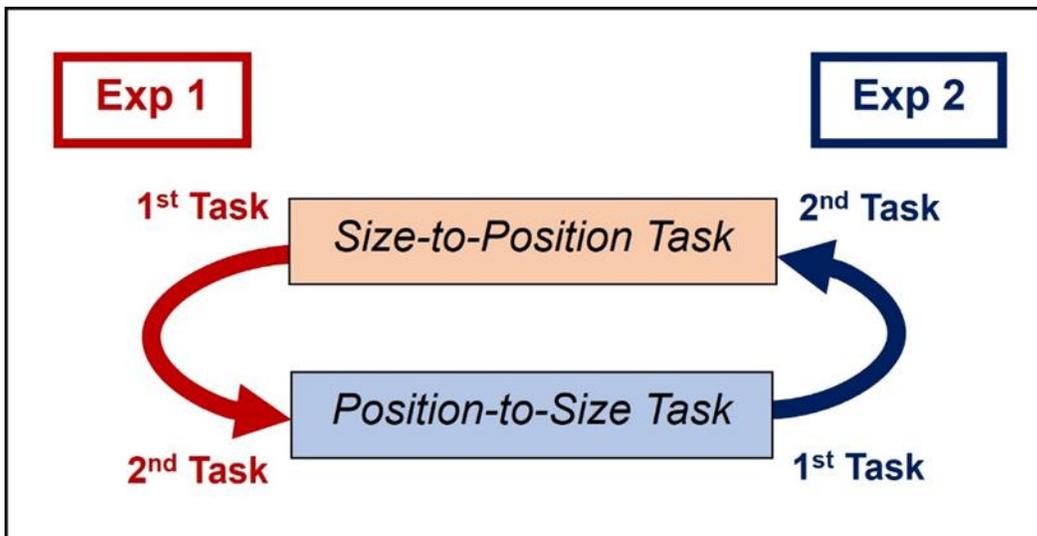


Figure 3.5: Task orders for Exp 1 and 2. The participants in Exp 1 performed the Position-to-Size task first (1st task) and then the Size-to-Position task (2nd task); and in Exp 2, they performed the Size-to-Position task first (1st task) and then the Position-to-Size task (2nd task). The results from the 1st task (target positions from the Size-to-Position task, and target sizes from the Position-to-Size task) were used in the 2nd task.

There was a total of 25 trials (5 trials per target sizes \times 5 target sizes) in the 1st task, and 25 trials (5 trials per target position \times 5 target positions) in the 2nd task. It took approximately 20 min to complete the experiment.

3.3.4. Methods for experiment 2

3.3.4.1. Participants.

Out of the 150 people who signed up for the experiment, 100 completed the experiment. After post-test data cleaning (see below), 48 participants had to be removed, leaving 52 participants for the analysis.

3.3.4.2. Procedures.

The test procedures were similar to Exp 1. The Enter Reference Length task (see above) was added following the Enter Phone Size task to obtain additional information about the participant's screen size. After the Enter Reference Length task, participants performed the

Position-to-Size task, where the targets were placed at positions corresponding to the test distances: 4, 8, 12, 16, and 20 m (0.417, 0.191, 0.123, 0.089, and 0.069 below the horizon, where 1 is the full screen height). The positions were determined from the position of the targets when they are placed in the 3D model in Unity which was used to render the 2D scene (see Figure 3.2 and Visual Stimuli section). Then, the participants continued to the Size-to-Position task, using the average sizes recorded from the Position-to-Size task for each intended target distance (see Figure 5 for the task order for Exp 2).

There was a total of 25 trials (5 trials per target sizes \times 5 target sizes) in the 1st task, and 25 trials (5 trials per target position \times 5 target positions) in the 2nd task. It took approximately 20 min to complete the experiment.

3.3.5. *Methods for experiment 3*

3.3.5.1. Participants.

A total of 210 people completed the experiment (93 in group 1 and 117 in group 2). After post-test data cleaning (see below), 126 participants had to be removed, leaving 84 participants for the analysis (42 in each of two groups).

3.3.5.2. Procedures.

The test procedures were similar to Exp 1 and 2 except the 2D pathway scene had familiar objects added to the scene. The familiar objects used were a bicycle and a door. They were either near (at 8 m) or far (at 16 m) from the observer in the rendered scene (see Figure 3.6). There were 2 conditions: task order and familiar object distance. Groups 1 and 2 were separated by task order: group 1 followed the task order for Exp 2 (see Exp 2 order in Figure 3.5), and the group 2 followed the task order for Exp 1 (see Exp 1 order in Figure 3.5).

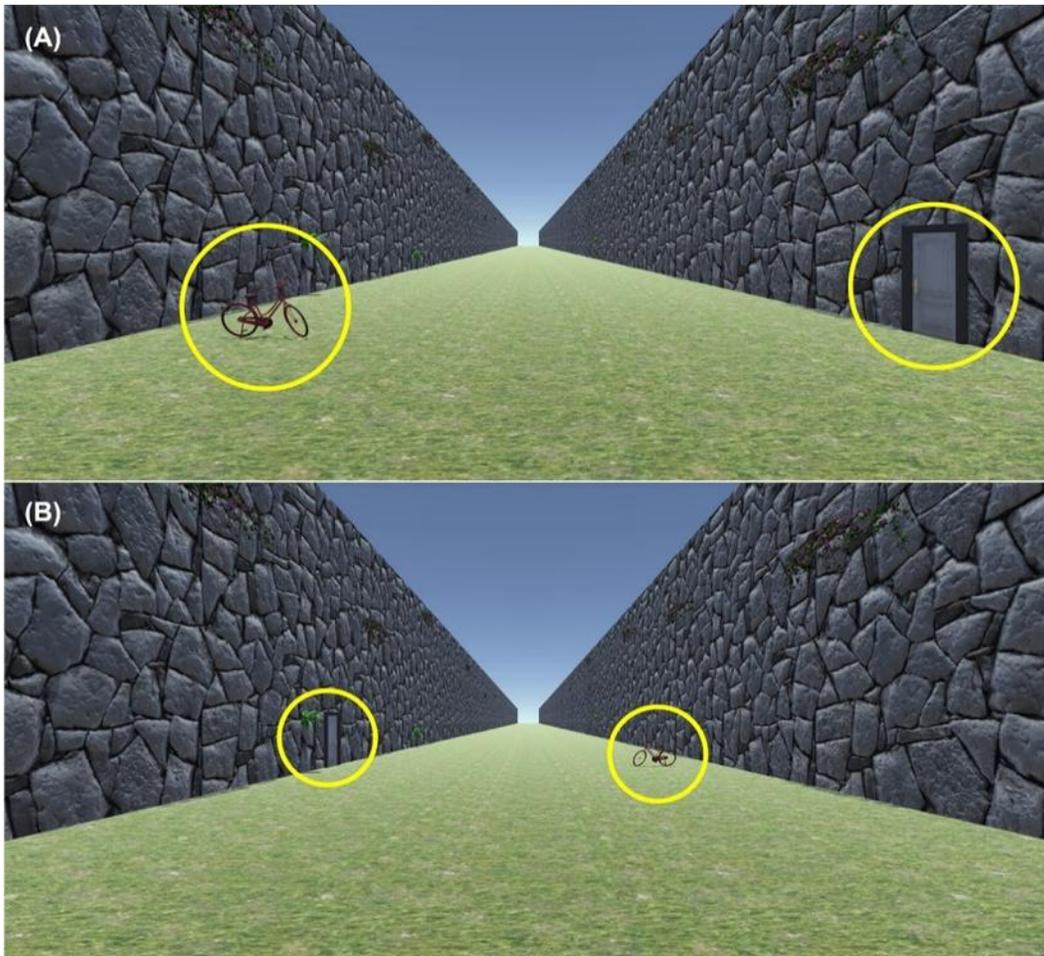


Figure 3.6: The pathway scene with familiar objects. The scene is identical to Figure 3.4A but contained a bicycle and a door (highlighted by yellow circles) at near (A) and far (B) distances. The yellow circles were not present on the display.

The order of the familiar object scenes (near or far) was counterbalanced to monitor any order effect of the familiar object position in the scene. In the 1st task, those with odd participant numbers (in both groups 1 and 2) viewed the scene with the familiar objects at the near distance for the first 20 trials then at the far distance for the next 20 trials and the order was reversed for those with even participants numbers. In the 2nd task, 40 trials were presented using the responses obtained from the near trials in the 1st task (20 with near familiar objects – near-to-near; and 20 with far familiar objects in the scene – near-to-far) and 40 trials were presented using the responses from the far trials in the 1st task (20 with near familiar objects – far-to-near;

and 20 with far familiar objects in the scene – far-to-far). All 80 trials in the 2nd task were randomly presented to the participants. All participants experienced the four familiar object orders (near-to-near, near-to-far, far-to-near, and far-to-far). Therefore, there were a total of 120 trials for each participant: 40 trials (4 trials per target distances \times 5 target distances \times 2 familiar object distances) in the 1st task, and 80 trials (4 trials per target distances \times 5 target distances \times 4 familiar object distance orders) in the 2nd task. It took approximately 30 min to complete the experiment.

3.3.6. *Data analysis*

In each experiment, the two tasks were performed in an order which depended on the experiment as described above. The target position or size results from the 1st task were used to set the initial values used in the 2nd task. Therefore, if participants could reliably estimate a target's visual size from its position and its position from its visual size, then the initial target position or size (presented in the 1st task) should match the final target size or position (results from the 2nd task). To determine whether participants could use target size to determine the correct target position in the scene or use target position to determine the correct target size, I compared the initial target size/position combination to the final target size/position combination after completing the tasks for each experiment. I also compared the target size and position between experiments.

3.3.6.1. **Converting positions to mm below the horizon.**

To convert target positions in proportion to the screen height (P_{screen}) to millimeters (mm), I multiplied the estimated screen height (H_{screen}) for each participant (see Section 3.3.1.3 for details on how that was computed) to the target positions results:

$$P = H_{screen} \times P_{screen} \dots \dots \dots (3.7)$$

3.3.7. *Post-test data cleaning*

Each participant's responses were evaluated for any possible misunderstanding of the tasks and irrational responses due to lack of enthusiasm (e.g., button mashing) that is inevitably found amongst undergraduates running unsupervised experiments online. Participants were removed from analysis if any of the followings were evident:

1. Incomplete dataset;
2. Impossible target position (target position > 0 – above horizon);
3. Unfeasible screen/phone sizes (i.e., the phone height-to-width ratio > 0.7 or <0.3, and the discrepancy in estimated screen heights from the phone size and the reference size is larger than 50 mm);
4. Uncorrelated responses to the target size/position conditions (suspected of not paying attention to the task, i.e., standard deviation of their responses was 0 in one or more conditions).

3.4. Results

3.4.1. *Experiment 1: the position-to-size task and then the size-to-position task*

In Exp 1, participants first adjusted the position of targets that were set to the sizes they should be when viewed at 4, 8, 12, 16, or 20 m (the Size-to-Position task) so that they appeared correct. Then, in a second task, they adjusted the size of targets that were positioned at the locations they had set in the 1st task (the Position-to-Size task). I conducted a repeated-measures, i.e., within-subject, analysis of variance (ANOVA) to analyze the differences among the mean sizes (task order: initial vs. final sizes) between the simulated target distances (see Appendix A: Table A1 for the results). The main effect of task order was significant, $F(1, 81) = 33.034$, $p < 0.001$, $\eta^2 = 0.207$. The main effect of simulated target distance was also significant, $F(4, 324) =$

130.388, $p < 0.001$, $\eta^2 = 0.124$. These main effects were qualified by a significant interaction between the task order and simulated target distance, $F(4, 324) = 6.958$, $p < 0.001$, $\eta^2 = 0.007$. Further evaluations with pairwise t-test using Bonferroni correction, adjusting the alpha level to accommodate multiple comparisons and maintain the familywise error rate of 0.05, showed that the final sizes were significantly larger than the initial sizes for each target distance ($p < 0.001$), which suggests that the participants could not successfully adjust targets to their original sizes from their positions in the 2nd task, even though they had themselves placed the targets there in the 1st task. It was also revealed that the final sizes for the targets simulated at closer distances were significantly larger than those simulated at further distance ($p < 0.001$) except for those between 12 and 16 m ($p = 1.000$) and between 16 and 20 m ($p = 1.000$). These results suggest that participants could match the target sizes in the correct order based on their simulated distances up to 12 m (See Table 1A for the average values from Exp 1).

Table 1: Average initial and final sizes/positions for all tasks.

A. The Average Initial and Final Target Sizes for Size-to-Position-to-Size Task Order					
	4 m	8 m	12 m	16 m	20 m
Initial size	0.125	0.063	0.042	0.031	0.025
Position (Exp 1) * (mm)	-48.0 ± 2.8	-26.5 ± 1.8	-18.7 ± 1.4	-13.9 ± 1.3	-12.8 ± 1.3
Position (Exp3: Group 2) * (mm)	-62.6 ± 4.9	-41.2 ± 3.0	-31.1 ± 2.1	-24.1 ± 1.8	-21.6 ± 1.8
Final size (Exp 1) **	0.29 ± 0.02	0.19 ± 0.02	0.15 ± 0.02	0.13 ± 0.02	0.14 ± 0.03
Final size (Exp 3: Group 2) **	0.28 ± 0.04	0.20 ± 0.04	0.17 ± 0.04	0.16 ± 0.04	0.16 ± 0.04
B. The Average Initial and Final Target Positions for Position-to-Size-to-Position Task Order					
	4 m	8 m	12 m	16 m	20 m
Initial position (Exp 2) (mm)	-73.7 ± 1.5	-33.8 ± 0.7	-21.7 ± 0.4	-15.6 ± 0.3	-12.1 ± 0.2
Size (Exp 2) **	0.47 ± 0.04	0.35 ± 0.03	0.28 ± 0.03	0.25 ± 0.03	0.19 ± 0.03
Final position (Exp 2)* (mm)	-55.0 ± 2.6	-41.7 ± 2.3	-33.0 ± 2.4	-27.4 ± 2.0	-20.9 ± 2.2
	4 m	8 m	12 m	16 m	20 m
Initial position (Exp 3: Group 1) (mm)	-76.9 ± 2.7	-36.1 ± 1.2	-23.5 ± 0.8	-17.3 ± 0.6	-14.1 ± 0.5
Size (Exp 3: Group 1) **	0.48 ± 0.05	0.35 ± 0.05	0.32 ± 0.05	0.32 ± 0.05	0.28 ± 0.05
Final position (Exp 3: Group 1) * (mm)	-62.3 ± 2.8	-46.5 ± 2.8	-40.3 ± 3.0	-35.8 ± 3.2	-31.3 ± 2.9

Note 1: The initial sizes (grey background) were the geometrically correct size for each simulated target. Note 2: The initial positions (grey background) are the average of the positions where targets were rendered in the 2D scene for each simulated distance. These were different for each participant due to the different screen sizes. Note 3: The values that are stated after each mean size/positions (\pm) are the standard errors. * The positions are how far below the horizon the target was placed in the Size-to-Position task. ** The sizes are in proportion to the phone size entered.

I also conducted a single-factor repeated-measures ANOVA to compare the positions at which targets were placed in the Size-to-Position task (1st task) with the simulated target distances (see Appendix A: Table A2 for the results). A significant main effect of position was found, $F(4, 324) = 143.443$, $p < 0.001$, $\eta^2 = 0.639$, which was followed by pairwise t-test using Bonferroni correction. The positions were significantly different for each target size ($p \leq 0.05$) except for those between 16 and 20 m ($p = 1.000$). That is, participants could match target size and position in the correct order in the scene up to 16 m away for their positions and up to 12 m away for their sizes. Participants may not have been able to infer target distances from their sizes or positions in a 2D scene because they could not match the targets' final sizes to their initial sizes.

Individual participant responses are shown in Figure 3.7 (Exp 1), which plots the position set as a function of size for the two tasks. The data patterns for the two tasks are very different from each other and from the black dashed line which indicates the correct geometry. To discover whether the order of the tasks mattered, I conducted Experiment 2 in which the task order was reversed.

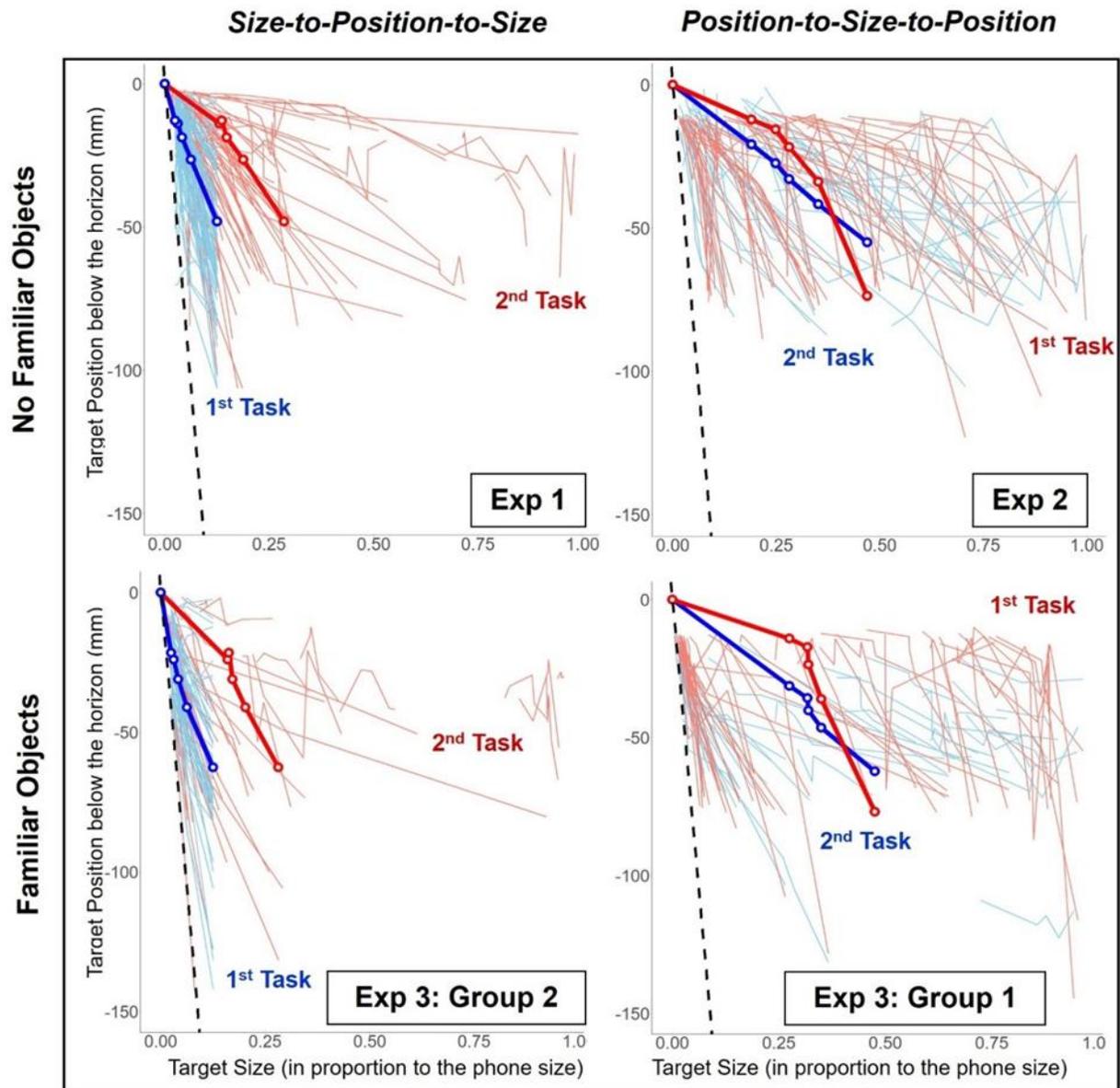


Figure 3.7: Participants' settings for all tasks. Target position plotted as a function of target size. Top row, with no familiar objects (Exp 1 and Exp 2); bottom row, with familiar objects (Exp 3). Left column, the Size-to-Position task first followed by the Position-to-Size task; right column,

the Position-to-Size task first followed by the Size-to-Position task. The dashed black lines represent target sizes and positions with the correct geometry. The solid light-colored lines are the average responses from each individual participant and the darker-colored lines with white dots represent the mean of responses from all participants for each task, where the dots are the means for each simulated target: the Size-to-Position task (Blue) and the Position-to-Size task (Red). (The means and the standard errors can be found in Table 1).

3.4.2. *Experiment 2: size-to-position then position-to-size*

In Exp 2, the task order was the reverse of that used in Exp 1: participants first adjusted target sizes from their given positions simulated to be at 4, 8, 12, 16, and 20 m (the Position-to-Size task). Then in the 2nd task, they adjusted the target positions for targets set to the sizes that they had just set for each simulated distance in the 1st task (the Size-to-Position task). The data are plotted in Figure 3.7 (Exp 2).

A repeated-measures ANOVA was conducted to compare the positions (task order: initial vs. final positions) between the simulated targets distances (see Appendix A: Table A3 for the results). The analysis revealed a significant main effect of task order, $F(1, 51) = 6.582$, $p = 0.013$, $\eta^2 = 0.010$, and a significant main effect of simulated target distance, $F(4, 204) = 385.818$, $p < 0.001$, $\eta^2 = 0.664$. These main effects were qualified by a significant interaction between the task order and simulated target distance, $F(4, 204) = 46.905$, $p < 0.001$, $\eta^2 = 0.077$. Further evaluations with pairwise t-test using Bonferroni correction showed that the final positions were significantly different compared to the initial positions for each target distance ($p < 0.050$), which suggests that the participants could not successfully adjust targets to their original positions from their sizes in the 2nd task which they had set in the 1st task. It was also revealed that the final positions were significantly further below the horizon for targets that were simulated at closer distances than those simulated at further distances except for the targets between 12 and 16 m ($p = 0.185$). These results suggest that participants could place the targets

in the correct order based on their simulated distances up to 12 m (See Table 1B for the average values from Exp 2).

In addition, a single-factor repeated-measures ANOVA was conducted to compare the sizes set in the Position-to-Size task (1st task) between intended target distances (see Appendix A: Table A4 for the results). A significant main effect of task was found, $F(4, 204) = 68.352$, $p < 0.001$ $\eta^2 = 0.573$. Pairwise t-test using Bonferroni correction showed that the positions were significantly different for each target size except for those between 12 and 16 m ($p = 0.770$). Therefore, participants could adjust the target sizes to the correct order in the scene up to 12 m.

These results are similar to the findings from Exp 1. The results further suggest that participants could adjust the target sizes and positions to the geometrically correct order to a degree, but that their responses were again very different from the correct geometry (see Figure 3.7 – Exp 2; the blue and the red lines compared to the black dashed line). From the results of Exp 1 and 2, it appears that the participants were not able to infer distances from their sizes or positions in a 2D scene.

3.4.3. Experiment 3: familiar objects in the 2D scene

In Exp 3, I repeated Exp 1 and 2 with the simulated target distances at 4, 8, 12, 16, and 20 m but provided an additional distance cue—familiar objects (a bicycle and a door, see Figure 3.6) in the scene. Group 1 followed the task order of Exp 2 and Group 2 followed the task order of Exp 1. The familiar objects in the scene were either at near or far distances, the order of which was counterbalanced during the experiment (see Section 3.3.5.2 for details).

I conducted a repeated-measures ANOVA comparing target positions for Group 1 and sizes for Group 2 to determine whether there was any effect of the order of the familiar object positions: near-to-near, near-to-far, far-to-near, and far-to-far. The analysis revealed no

significant differences between familiar object orders for Group 1, $F(3, 123) = 0.117$, $p = 0.950$, $\eta^2 < 0.001$, or Group 2, $F(3, 123) = 0.173$, $p = 0.915$, $\eta^2 < 0.001$. Therefore, I pooled the participant responses by averaging the sizes and positions across the familiar object order conditions for the following analyses.

3.4.3.1. Group 1: position-to-size-to-position.

I performed a repeated-measures ANOVA to compare the positions (task order: initial vs. final positions) between the simulated target distances (see Appendix A: Table A5 for the results). The analysis revealed a significant main effect of task order, $F(1, 41) = 29.855$, $p < 0.001$, $\eta^2 = 0.054$, and a significant main effect of simulated target distance, $F(4, 164) = 422.301$, $p < 0.001$, $\eta^2 = 0.658$. I also found a significant interaction between task and target position, $F(4, 164) = 61.639$, $p < 0.001$, $\eta^2 = 0.090$. Pairwise comparisons using Bonferroni correction showed that the final positions were significantly different from the initial positions for each simulated target distance ($p \leq 0.001$). Further evaluation also showed that the final positions were significantly further below the horizon for the targets simulated at closer distances than those simulated at further distances except for those between 12 and 16 m ($p = 0.574$), and between 16 and 20 m ($p = 0.639$). These results are consistent with the previous results from Exp 2 (see Table 1B for the average values from Exp 3: Group 1).

A single-factor repeated-measures ANOVA was also conducted to compare the sizes matched in the Position-to-Size task (1st task) between simulated target distances - see Appendix A: Table A6 for the results - and a significant main effect of task was found, $F(4, 164) = 40.487$, $p < 0.001$, $\eta^2 = 0.497$. Pairwise t-test using Bonferroni correction showed that the sizes were significantly different for each intended target distance except for those between 12 and 16 m ($p = 0.099$) and 16 and 20 m ($p = 1.000$).

3.4.3.1. Group 2: size-to-position-to-size.

I performed a repeated-measures ANOVA to compare the sizes (task order: initial vs. final positions) between the simulated target distances (see Appendix A: Table A7 for the results). The analysis revealed a significant main effect of task order, $F(1, 41) = 12.123$, $p = 0.001$, $\eta^2 = 0.193$, where the final target sizes were significantly larger compared to the initial sizes on average (see initial and final sizes for Exp 3: Group 2 in Table 1A). In addition, there was a significant main effect of simulated target distance, $F(4, 164) = 64.329$, $p < 0.001$, $\eta^2 = 0.067$. However, there was no significant inter-action between order and intended target distance found, $F(4, 164) = 0.916$, $p = 0.456$, $\eta^2 < 0.001$.

In a single-factor repeated-measures ANOVA conducted to compare the positions set in the Size-to-Position task (1st task) between simulated target distances (see Appendix A: Table A8 for the results), a significant main effect of simulated target distance was found, $F(4, 164) = 77.179$, $p < 0.001$, $\eta^2 = 0.653$. Pairwise t-test using Bonferroni correction showed that the positions were significantly different for each target size except for those between 12 and 16 m ($p = 0.099$) and 16 and 20 m ($p = 1.000$). These results are similar to those found in Exp 1.

3.4.4. *Between experiments analysis*

3.4.4.1. Effect of familiar objects on the 1st task.

When comparing the adjusted target sizes from the Position-to-Size task in Exp 2 and Exp 3 - Group 1, responding to the initial target positions, I used a repeated-measures ANOVA. No significant effect of familiar object was found, $F(1, 92) = 0.498$, $p = 0.482$, $\eta^2 = 0.005$. However, when comparing the positions set during the Size-to-Position task in Exp 1 and Exp 3 - Group 2, responding to the initial target sizes, there was a significant effect of familiar object, $F(1, 122) = 20.267$, $p < 0.001$, $\eta^2 = 0.065$. These results suggest that familiar objects in a 2D

scene did not influence the participants' judgement of the target size based on target position (Position-to-Size), but they affected their judgement of the target position based on target size (Size-to-Position).

3.4.4.2. Effect of familiar objects on the 2nd task.

When comparing the target positions set during the Size-to-Position task in Exp 2 and Exp 3 - Group 1, responding to the target sizes they adjusted to in the 1st task, I used a repeated-measures ANOVA. A significant effect of familiar object was found, $F(1, 92) = 6.628$, $p = 0.012$, $\eta^2 = 0.032$. However, when comparing the adjusted target sizes from the Position-to-Size task in Exp 1 and Exp 3 - Group 2, responding to the target positions they set in the 1st task, no significant effect of familiar object was found, $F(1, 122) = 0.188$, $p = 0.665$, $\eta^2 = 0.001$. These results are consistent with the findings from the comparisons of the 1st task, where familiar objects in a 2D scene did not influence the participants' judgement of the target size based on target position (Position-to-Size), but they affected their judgement of the target position based on target size (Size-to-Position).

3.5. Discussion

In my study, participants positioned a rendition of their cell phone of a given size to where they thought it should be in a 2D scene, the Size-to-Position task, or they adjusted its rendered size until it appeared correct while it was at some fixed position in the scene, the Position-to-Size task. The distance of a given object determines its image size. Hence, participants would have had to estimate the distance to the rendition of their cell phone to be able to adjust its size correctly and vice versa. Given the known familiar physical size of their phone, I hypothesized that (H3) they should have been able to adjust the target's position below the

horizon and the size at which it should appear. Instead, in all the experiments, the final target size/position combinations did not match the initially simulated values.

I interpret my results as indicating that my participants could not infer target position from its size or its size from its position, which implies that my participants were not able to estimate the distance to the target reliably. In general, they made the rendition too large or placed it too close to the horizon (too far away) compared to the geometrically correct values (see Figure 3.7). Pictorial cues such as wall patterns (although the stone sizes for my walls were ambiguous compared to standard sized bricks) and linear perspective from the intersections between the walls and the floor can help in people's distance perception (Hornsey & Hibbard, 2021) but participants could not use these cues to determine absolute distance. This may be due to the 2D scene being seen as smaller (minified) as was found in past studies when a real-world scene that was presented on a synchronized image display had to be magnified substantially to be seen as correct (Roscoe, 1984). If the 2D space were perceived as smaller than simulated, then an object would have had to be made larger for it to appear to be the correct size and placed further away for it to appear to be at the correct distance, which is what my participants did on average. My results could not be represented using Stevens's power law (Stevens, 1957) or Gilinsky's formulas for perceived size and distance (Gilinsky, 1951). Both models propose underestimation of an object distance, whereas my results show the opposite effect for some tasks (i.e., the Size-to-Position task). Additionally, the responses from each individual participant varied greatly, therefore making it difficult to be represented using a single model.

Overall, it would seem that my participants could not use a familiar object's position (angular declination below the horizon) or its visual size to derive its absolute distance in a 2D scene. Participants not being able to infer absolute distance from size is consistent with past

studies of perceived size and distance in the real world (Gogel et al., 1963). However, not being able to infer distance from target position is not consistent with previous research - where angular declination below horizon helped in distance perception (Ooi et al., 2001) - which may be due to using a 2D scene. Ooi et al. (2001) demonstrated that the perceived eye level is important in computing the angular declination below horizon. In the 2D scene presented in my experiment, the ground information from their feet to the bottom of the screen was missing and participants may therefore have mis-perceived the eye level which was set to a fixed value of 1.7 m. Dixon et al. suggested that eye level in general is not scaled correctly when people view non-immersive displays because the altitude of the horizon is indeterminate (Dixon et al., 2000). Misperceived eye level would then lead to misperceived angular declination of the target, resulting in errors when determining a target's distance from its position. Although participants could not determine the absolute distance of the targets from their size and position in the scene, they could use what they saw to, at least mostly, determine the correct distance order. In the real world, a given object at a further distance from the observer has a smaller visual angle compared to that same object viewed at a closer distance. When viewed more distantly, the object will also be closer to the horizon in the visual field. In all experiments, my participants set targets that were closer to the horizon to smaller sizes compared to the targets further below the horizon in the Position-to-Size tasks. They also placed the smaller targets closer to the horizon compared to the larger targets in the Size-to-Position tasks. This shows that they had some idea of the three-dimensional nature of the world that was depicted. Gogel et al. (1963) found that the perceived absolute size and distance of an object were positively correlated, but that the ratio between them varied rather than following strict size-distance constancy rules. My results are consistent with this idea. When the horizon is clearly defined in a picture, observers can use distance to the

horizon to determine relative distance to objects and ignore other possibly erroneous information such as the height in the picture plane, i.e., “the distance from the picture’s lower border to the bottom of the object” (Gardner et al., 2010) (p. 445). It would seem that my participants could not derive absolute distances from target size or position, but they could correctly infer relative distances using the angular declination below the horizon and size constancy. Most past studies on perceived size and distance have been conducted in the real or virtual environment in a 3D space (Chen et al., 2019; Gogel et al., 1963; Maltz et al., 2021). When using a 2D image, they were designed to provide sufficient distance information, such as using a live video or a photo of a real scene (Boring, 1964; Roscoe, 1984), rather than to limit the distance cues as I did in this study. My study fills this methodological gap in the literature and shows that people can judge relative distance to the objects in a 2D scene from using only their visual sizes or positions.

Out of the five targets rendered at distances between 4 and 20 m, participants could do this for targets at 4–12 m (corresponding to the base of the object 76.9 to 21.7 mm below the horizon on average; see Table 1). The horizon ratio, the ratio between the visual height and the distance below the horizon of an object in pictures, can help judge relative sizes which are typically most accurate at eye level, i.e., near horizon (Bertamini et al., 1998; S. Rogers, 1996). Instead, my participants were less accurate in determining the relative sizes of targets closer to the horizon (simulated distance beyond 12 m). This may be due to the difference in the size of the target used. Bertamini et al. (1998) used poles at 60% of the observer eye height (the shortest pole was 96 cm), whereas my participants used their own smartphone as reference size (approximately 14.3 cm). As objects are simulated at further distances, their angular size and displacement from the horizon become smaller, hence more difficult to distinguish as the differences in size and distance also become smaller. The largest difference between the targets

simulated beyond 12 m were 7 mm for positions and 0.03 for sizes (approximately 4.3 mm for height) which correspond to visual angles of 0.8° and 0.5° , respectively. These differences may have been too small for the participants to distinguish them.

3.5.1. Why is there a task order effect?

How well my participants performed on the tasks I set them depended, unexpectedly, on the order in which they performed them. The targets positioned based on their size, the Size-to-Position task, were more geometrically accurate compared to the target sizes chosen for a given position, the Position-to-Size task, when they performed the Size-to-Position task first (i.e., blue lines were closer to the dashed line compared to the red lines in Figure 3.7 Exp 1 and Exp 3: Group 2). However, when they performed the Position-to-Size task first, the size was consistently set too large and subsequently performed the Size-to-Position task did not improve their accuracy. In the 1st task, participants were presented with targets that were set to the geometrically correct size (for the Size-to-Position task) or position (for the Position-to-Size task). Based on these results, it appears that participants were more accurate at placing targets at the geometrically correct positions when the correct target sizes are presented to them during the Size-to-Position task. However, being presented with targets at the correct position during the Position-to-Size task did not help them match the targets to the correct sizes. My data confirms that the absolute distance of an object may not be determined from its size and position presented in a 2D scene, but it is unclear why participants' responses differed between the tasks. Such an order effect suggests there might be fundamental differences between these tasks. Do people perceive size and distance differently depending on the task they are performing?

Despite the popularity of the size–distance hypothesis, studies have shown that size perception and distance perception may, to some extent, be independent from each other (Haber

& Levin, 2001; Kim, 2018). Kim (2018) suggested “size and distances are two independent perceptual processes with each determined directly by the corresponding information sources” (p. 16). Haber and Levin (2001) also claimed that size perception is based on properties of the object such as prior knowledge or experience that the observer had, and that distance perception is based on the environmental information that describe its distance. The two tasks used in the present study asked the participants to determine different aspects of the targets. In the Size-to-Position task, they had to determine target position (i.e., their distances), while in the Position-to-Size task, they had to determine target size. If size and distance perceptions are fundamentally different, then this might explain the differences between the two tasks shown here. That is, being presented with geometrically correct target sizes results in more accurate estimates of target positions in the Size-to-Position task but being presented with correct positions did not result in more accurate target sizes in the Position-to-Size task.

During the Size-to-Position task, they may have used the environment sources (the 2D scene) to determine the target position and focused on placing each target based on where they saw the targets to be in the environment (the pathway scene), ignoring the visual sizes of the targets. If this were the case, then it is possible that participants simply distributed targets along the pathway in the scene based exclusively on their relative sizes. The targets were still placed in the appropriate order, which shows they understood the overall relationship between size and position in a 2D scene. The plots for individual participants shown in Figure 3.7 show large individual differences. Some participants set target sizes that were so large that their correct geometrical position would have been too close to even be rendered on the screen during the Size-to-Position task (Figure 3.7, refer to the individuals on the far-right side of the plots). However, instead of placing targets clustered towards the bottom of the screen, they still placed targets

distributed along the pathway, further suggesting that they could not use the visual size of the targets to determine their correct positions in a 2D scene.

Similarly, during the Position-to-Size task, many targets were made much larger than their geometrically correct size, despite their position in the scene. The participants may have been unable to use the environment (2D pathway scene) and focused instead on matching the targets to the physical size of their smartphone. When viewing a picture, there are two distinct distances a person can perceive simultaneously: (1) a distance from the eye to the picture, and (2) the distance from the point of view of the picture, i.e., within the three-dimensional scene in the picture (Gibson, 1954). These participants may have failed to perceive the scene as a space within the picture during this task. Some participants, however, were able to use the size and position of targets to determine their geometrically correct distances (Figure 3.7; refer to the individuals on the far-left side of the plots, close to the geometrically correct lines). Individual differences have been shown in past experiments evaluating absolute sizes and distances for objects beyond 3 feet (Gogel & Mertens, 1967; Higashiyama, 1983); Higashiyama (1983) suggested that there may be different populations of observers using different types of strategies. My participants seem to have used different individual strategies when determining the size of an object as opposed to when determining its distance. These results demonstrate the independence of the size and distance perception further, at least in a 2D scene.

3.5.2. Can familiar objects improve object distance judgements in a 2D scene?

There are many studies looking at the effect of familiar size on determining the perceived distance of a familiar object (Fitzpatrick et al., 1982; Haber & Levin, 2001; Hochberg & Hochberg, 1952). However, results are mixed as to how familiarity affects perception. Hochberg and Hochberg (1952) suggested that familiar size may not affect perception of depth at all.

Changing the visual size of a familiar object can sometimes affect its perceived distance (Fitzpatrick et al., 1982), but this may depend on the person's viewing attitude, i.e., using different strategies. Being instructed to use the knowledge the person has about an object can result in them relying less on the perceptual information (Higashiyama, 1984; Predebon, 1992), cognitively fixing the absolute size of an object in their mind and only adjusting its apparent distance according to its visual size as Fitzpatrick et al. (1982) showed in their study. Haber and Levin (2001) found that an observer's familiarity with an object helped determine their perception of the size and distance of a far object (50–100 m) where distance information was limited, but not for close objects (0–50 m) where distance information was clear. They also found that people were more accurate at determining an object's size and distance when a given familiar object's size varied less in general, e.g., bikes vary less in size compared to house plants (Haber & Levin, 2001).

In this study, the targets represented participants' smartphones which they are very familiar with, but these familiar visual stimuli were displayed in a 2D scene which provided limited distance information. Recognizable familiar objects in the environment can help in determining the size and distance of other objects (Park et al., 2021). Therefore, adding a bike and a door to the scene - two of the items used by Haber and Levin (2001) as objects with low variance in size (see Haber & Levin, 2001, Table 1, p. 1142) - was expected to improve participants' size and distance perception (H4). My results show familiar objects in the scene did indeed influence participants' responses in the Size-to-Position task. Having familiar objects in the scene resulted in participants' responses being more geometrically correct in the Size-to-Position task. It appears that familiar objects may play a role as anchors which observers can use when determining an object's position from its size.

The familiar objects, however, did not affect responses in the Position-to-Size task. Interestingly, therefore, the objects seem to have helped determine a target's position but not its size in a 2D scene. This does not align with the past findings, where familiar size was found to affect both perceived size and distance (Maltz et al., 2021). Maltz et al. (2021) found people perceived Rubik's cubes as larger and further away than dice when they were matched to the same physical sizes and distances. However, familiar size affected size perception more when viewing an object monocularly compared to when viewing it binocularly. Additionally, judgement of size was found to vary with depth information in the real-world scene displayed on a screen, but judgement of distance did not (Meehan & Triggs, 1992). My data suggests that when viewing an object in a 2D scene, the presence of familiar objects does not affect perceived size and distance equally, which is a novel finding that, as far as I know, has not been observed before. The differences in the effect of familiar size on my tasks further demonstrate that size and distance perception processes are independent.

3.5.3. Limitations and future studies.

The present study was conducted online, with each participant using their own computer screen which provided a limited field of view, especially in the vertical dimension. The 2D scene was presented on a regular screen and much of the ground information was missing. Unlike in the lab, I did not have the ability to blank out the rest of the world so that only the screen was visible or to fix my participant's eye height above the simulated ground plane. The lack of sufficient ground plane information in my displays may have contributed to my participants not being able to infer correct distances from target sizes and positions. Future studies should utilize a larger screen, fully extended to the floor, which may allow more accurate distance perception even when viewing an object on a 2D screen.

Although I instructed participants to imagine themselves standing in the pathway scene, some may not have been able to do so effectively. If they failed to consider the environment in the 2D scene as a life-size pathway but instead viewed it like a painting, then they would not have perceived it at the correct scale. The perceived size of an object is “driven from the underlying scale of the environment...within which individual objects are located” (Sedgwick, 2021, p. 15). Misperceiving the scale of the environment would result in misperceiving the size of an object, leading to inaccurate perceived distance. More immersive experience of the scene, such as in a cave automatic virtual environment (CAVE) or other virtual environment, would help observers perceive an environment at the correct scale, hence perceiving object size correctly and perhaps leading to a more accurate determination of its distance.

3.6. Conclusion

When determining an object’s distance from a 2D scene displayed on a computer screen, people cannot use its visual size and position (elevation in the scene) are not reliably to compute the distance simulated in the presented scene. Familiar objects in the scene seem to help in determining the target’s size more accurately, improving their performance when adjusting its position according to its visual size, but they do not help in determining its correct position in the scene. These results suggest that size and distance perception processes may be independent and rely on different information to determine them.

4.0. Can people determine object distance from its visual size and position in a correctly scaled 2D scene displayed on a large screen with aligned ground plane?

This chapter is of a manuscript which is currently under preparation. I am the first author of the paper and Dr. Laurence Harris is the second author. My contribution to the paper includes designing the experiment, programming the tasks, writing up the ethics forms, running the experiment, testing participants, analyzing and interpreting the data and writing up the paper.

Dr. Harris's contributions include helping with the design of the experiment, providing feedback on experimental tasks and interpretations of the results, and editing the paper.

4.1. Abstract

The results in Chapter 3 demonstrated that people can infer relative distance of an object reliably from its visual size and position in a 2D scene, but not its absolute distance when viewed, for example, on a computer screen. Here, I repeated the experiment described in Chapter 3, that I had originally done remotely, with visual stimuli projected onto a wall. The simulated scene was displayed at natural size with the ground plane aligned with the floor of the display room. Might realistically scaling the environment improve people's distance perception in a 2D scene? Participants sat on a chair and held a familiar reference object (a physical cereal box) in their hands while viewing a target (an image of a cereal box) projected onto a wall 2 m away. They compared their sizes and adjusted either the target's size based on its position in the scene (Position-to-Size task) or its position based on its visual size (Size-to-Position task) to match the true reference size. In experiment 1, the adjusted target position was consistent with the object's visual size. The adjusted size, however, was consistently larger than the correct size. These results suggest that people can use the visual size of an object to infer distance from a 2D scene if it is presented in a correctly scaled large display unlike in the study described in chapter 3

where the target and the scene were presented on a small monitor screen. Experiment 2 used a tracking task using a wider range of simulated distances with smaller intervals between them. Estimating participants' perceived horizon and eye height, by fitting their responses to a linear model, revealed that participants may have perceived the horizon to be lower than it was in the scene. The eye height was perceived to be shorter in the Position-to-Size task than in the Size-to-Position task. The fact that we observed different errors in the two tasks supports the idea that they are fundamentally different tasks, where one relies on size perception and the other on distance perception using different mechanisms, confirming the results of my previous study.

4.2. Introduction

Humans can easily recognize and perceive objects (e.g., from drawings on a cave wall) or even a 3D scene (e.g., from landscape paintings with perspective) represented in images presented on a flat surface. Most animals on the other hand cannot do this, or at least not without significant training (Jitsumori, 2010). Images on cave walls, canvas, paper, and more recently 2D displays have been used to expose people to images simulating distance for many purposes from simple entertainment to operating equipment remotely. Such images use various cues to simulate distance such as size, perspective, and elevation. In Chapter 3, I tested whether people can use visual size or position (i.e., distance below the horizon) of an object in a 2D scene to infer its distance and demonstrated that people can infer the relative distances (distance between objects) using their visual sizes or positions, but not their absolute distances (distance from the observer). The results of that experiment showed that generally targets were made larger and higher in the scene (closer to the horizon) than their correct size and position, which I interpreted as participants perceiving the 2D scene as smaller than they were simulated to be (i.e., a miniature world). Roscoe (1984) found that for a real-world scene presented on a display (a synchronized

image) to be seen as the correct size it had to be magnified substantially and that such errors in perceived scale influenced the observer's perception of distance in such a space. Knowing the correct eye height is important in computing angular declination relative to eye level (Ooi et al., 2001) and any errors can significantly influence people's estimate of both absolute distance and the dimension of a room (Leyrer et al., 2011), or judging height of surfaces relative to their body (Mark, 1987). When people view non-immersive displays, such as a scene on a computer screen or a landscape painting or photograph, they generally cannot scale their eye level correctly because the altitude of the horizon is indeterminate (Dixon et al., 2000). Will presenting an object in a realistic, immersive scene enable people to correctly scale their eye level hence correctly judge absolute distance of an object from its visual size and its position in the scene, even in a 2D presentation?

4.2.1. Judging Distance in an Immersive Environment

The experiment in Chapter 3 was conducted online using each participant's personal computer monitors. Therefore, the screen size, eye height, and distance to the screen could not be well controlled which would have made it difficult for participants to estimate their correct eye height in the scene. A carefully designed lab setting using a correctly scaled 2D scene displayed on a large screen with the horizon carefully aligned with their eye height while sitting should give the observer better cues to scale their eye height correctly. When viewing a picture of a scene, two distinct perception of space could occur: 1) the three-dimensional space presented in the picture, and 2) the room that the observer is in which contains the picture of a scene (Gibson, 1954). Some participants' responses in Chapter 3 showed evidence of failing to perceive the scene as its own space, where the target was made to be the full size of the reference size as if it were on the screen (i.e., at the same distance as the screen) rather than at a distance simulated in

the scene. Using a full-scale projection of a scene on a wall in comparison to a scene on a computer monitor, used in Chapter 3, should eliminate this conflict in their perceived space giving observers a more accurate perception of space in the presented scene.

Additionally, human perception can be influenced by a person's cognitive state of thinking something is real or not real. Misperception of distance in a virtual environment is reduced or can even disappear when participants are previously exposed to an identical real room (Interrante et al., 2006). Similarly, in my study looking into people's spatial updating in virtual reality (Chapter 2), sitting in a real room identical to the virtual room used in the study, resulted in observers producing smaller errors compared to when sitting in a room that was completely different. A more realistic test environment, giving participants the feeling of looking at a real object in a real hallway, may affect people's interpretation of the visual stimuli at a higher cognitive level (e.g., from their previous experience or memory) resulting in better distance perception.

4.2.2. Task Dependent Effect of Distance Cues: Using Size or Position

There are many distance cues other than size and position that are important when determining the distance of an object. Near-ground-surface (i.e., the ground from the feet to an object) information is important in judging distance accurately (Wu et al., 2004). Familiar size has also been found to affect both perceived size (Maltz et al., 2021; Park et al., 2021) and distance (Fitzpatrick et al., 1982; Gogel & Mertens, 1967; Maltz et al., 2021; Park et al., 2021) although it can also depend on the person's strategies (Fitzpatrick et al., 1982) or the instructions given by the experimenters (Higashiyama, 1984; Predebon, 1992). In theory, more distance information provided in a scene should improve people's accuracy in perceiving an object's distance. However, in Chapter 3 the presence of familiar objects in the scene (such as a bicycle),

used to give a clearer idea of scale, affected the two tasks differently (Chapter 3): it improved accuracy when moving an object to the correct position based on its visual size (Size-to-Position task), but not when making the object the correct visual size based on its position (Position-to-Size task). There is some evidence in the literature that distance information in the real-world scene when displayed on a screen affects judgement of size but not the judgement of distance (Meehan & Triggs, 1992). The distance cues added in a 2D scene may impact participants' size and distance perception differently which may be additional evidence that size and distance perceptions use different mechanisms.

4.2.3. Judging Size and Distance in a 2D Scene Monocularly or Binocularly

When determining the size and distance of a familiar object, people tend to be affected by the object's known size in such a way that a typically larger object (e.g., a Rubik's cube) is perceived as larger and farther than a typically smaller object (e.g., a die) even when their physical sizes and distances are matched, i.e., made to be the same physical size and placed at the same distance (Maltz et al., 2021). This 'familiar size effect' demonstrates that humans are influenced by prior knowledge of an object's typical size when judging its size and distance even when their idea of an object's size may not accurately represent its actual size. When people view familiar objects monocularly the 'familiar size effect' is even stronger compared to viewing it binocularly for perceived size, but not perceived distance (Maltz et al., 2021). This finding suggests people are not affected as much by familiar size when judging distance as they are when judging size when viewing an object monocularly, likely due to relying more on their memory of the familiar object size than its visual size. When viewing a 2D scene, binocular cues provide conflicting distance information (i.e., everything is at the same distance; no depth) from the monocular cues and it would distract rather than help judging distance. Past study shows that

when monocular and binocular cues are in conflict, monocular cues contribute more to size perception than binocular disparity (Yoo et al., 2023). People may be better at judging an object's size and distance in a 2D scene when viewing it monocularly than binocularly with conflicting distance information.

4.2.4. Hypotheses

In this study, a representation of a 3D scene was presented in a 2D display where binocular cues were not available to help determine target size or distance. I used a correctly scaled 2D scene displayed on a large screen (projection on a wall, see Figure 1A) to give the observer a better perception of scale and tested whether it improved people's distance perception in a 2D scene compared to the past study (Chapter 3). The scene provided additional distance cues, except for the binocular cues, such as familiar objects (both as the target and as objects in the environment), and ground information (real carpet from the observer's feet to the wall, lined up with the scene floor with identical pattern). With the rich distance information in the present study, people should be more accurate at inferring an object's absolute distance.

With a convincing semi-immersive set up of a 2D scene displayed on a large screen, I hypothesized that (H5) people would more likely to experience looking at a real hallway scene leading to more accurate inference of an object's absolute distance (i.e., will adjust the target to be closer to the correct size/position) compared to the results of my past study (Chapter 3). Although I expect people to improve with the added ground surface information, the magnitude of improvement may depend on the task. In the past study (Chapter 3), familiar-sized objects in the scene helped in the Size-to-Position task, but not in the Position-to-Size task. I interpreted that this difference between the tasks were due to the size and the distance judgements relying on different visual cues, likely size judgement relies more on monocular cues in the 2D scene.

Therefore, I hypothesized that (H6) people will again be less accurate in the Position-to-Size task than in Size-to-Position task even with the additional distance cues in the present study such as ground information (e.g., near-ground-surface and carpet pattern).

Past studies (reviewed above) suggest that monocular cues such as familiar size, impact perceived size but not distance when viewing an object monocularly. The present study was conducted on a 2D display of a scene and all the distance cues provided were monocular cues, but the scene could still be viewed monocularly or binocularly. I hypothesize that (H7) people will be more accurate when viewing monocularly at estimating an object's absolute distance from its visual size (Size-to-Position task) in a 2D scene than when viewing binocularly with conflicting stereoscopic cues indicating everything as being at the same distance (that of the screen), but not when estimating size from its position (Position-to-Size task).

4.3. General Methods

4.3.1. Materials and Methods

4.3.1.1. Apparatus.

The experiment was programmed using PsychoPy with Python language and ran on a Windows 10 computer, model Alienware 17 R4 with Intel® Core™ i7-7700HQ CPU @ 2.80 GHz, 16.0 gigabyte RAM, and NVIDIA GeForce GTX 1070 graphics card. The visual stimuli (images of a cereal box in a hallway scene Figure 4.1A) were projected on a wall using a short-throw projector, BenQ Digital Projector, model HT1085ST, Full HD 1080P. The scene was made using Unity3D then a screenshot was taken looking into the hallway. Participants held a physical family-size cereal box, identical to the target in the image (see Figure 4.2B and C), to compare to

the visual target and they used a hand-held mouse (Figure 4.1B) to adjust target size or position in the scene. An eye-patch (Figure 4.1C) was used for monocular viewing condition.

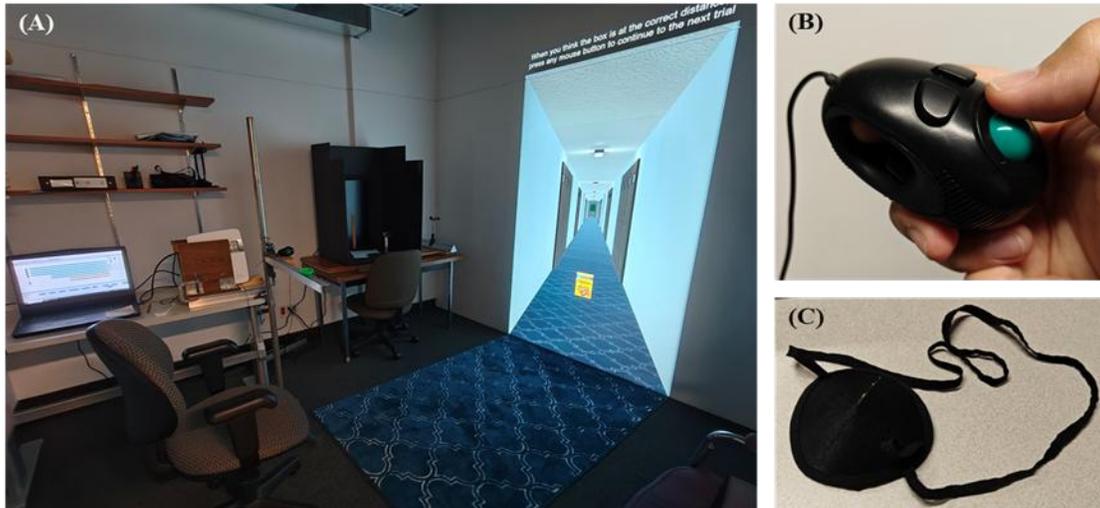


Figure 4.1: Test apparatus and setup. (A) Test room setup with a chair, chinrest, and a projector presenting visual stimuli on the wall. Note the carpet in the room was continuous with the same-patterned carpet in the image. (B) Hand-held mouse participants held on their right hand and used to adjust target size/position. (C) Eye-patch used for the monocular viewing condition.

4.3.1.2. Stimuli.

A target, an image of a cereal box (Figure 4.2C), was displayed in a correctly scaled 2D scene of a hallway (Figure 4.2A). The hallway, 2.205 m high and 1.245 m wide with a simulated length of 25 m, was projected on a wall, 2 m away from the participant's eyes, to give the observer an impression they were looking into a realistically sized hallway. Participants sat on a chair and looked at the scene with their chin on a chinrest (1.1 m above the floor) to maintain their eye level (H_{eye}) at approximately 1.2 m above both the real and the simulated floor. They were instructed to imagine they were sitting on a chair in a real hallway.

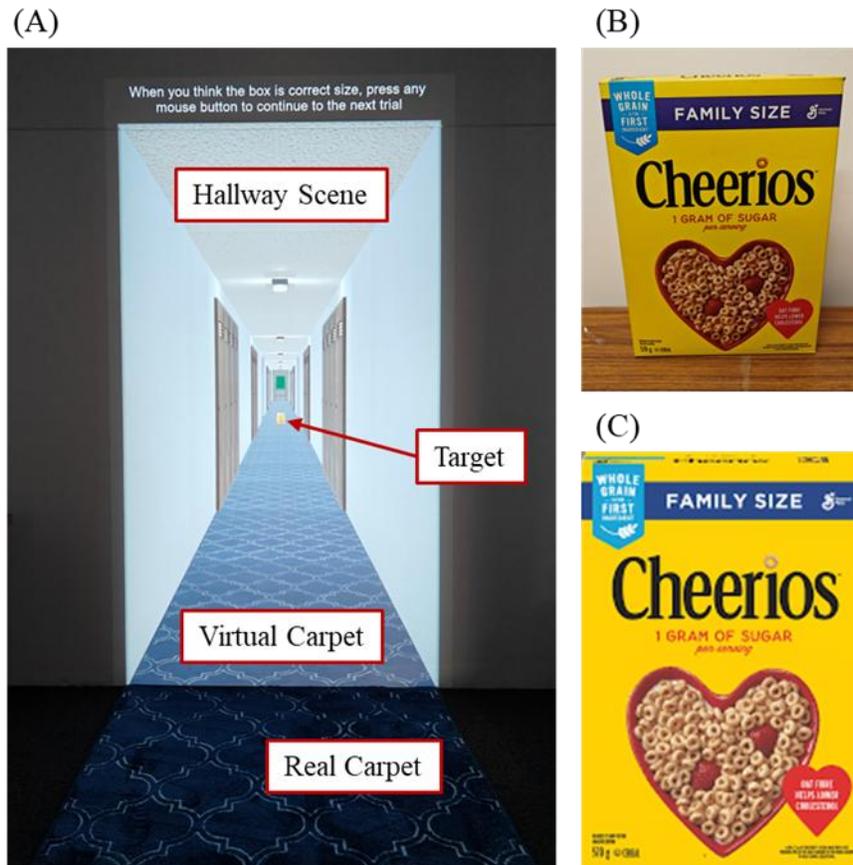


Figure 4.2: Test stimuli. (A) The hallway scene, projected on a wall, contained lockers on the side walls, ceiling lights, a green board, and virtual carpet. There was a real carpet with identical color and pattern on the ground in front of the projected scene made to be seen as continuing into the hallway floor. (C) The target was an image of a family-size cereal box identical to the (B) physical reference box held by participants during the experiment.

Participants held a physical cereal box (Figure 4.2B) as a reference during the experiment. The box was 0.344 m in height, 0.241 m in width, and 0.076 m thick. They compared this reference box to the target on the screen which was simulated to be the same size as the reference box (Simulated size; $S_{simulated}$) standing on the floor at a set distance (Simulated distance; $D_{simulated}$) away from the observer. The target's size and its position projected on the screen varied for each simulated distance as shown in Figure 4.3.

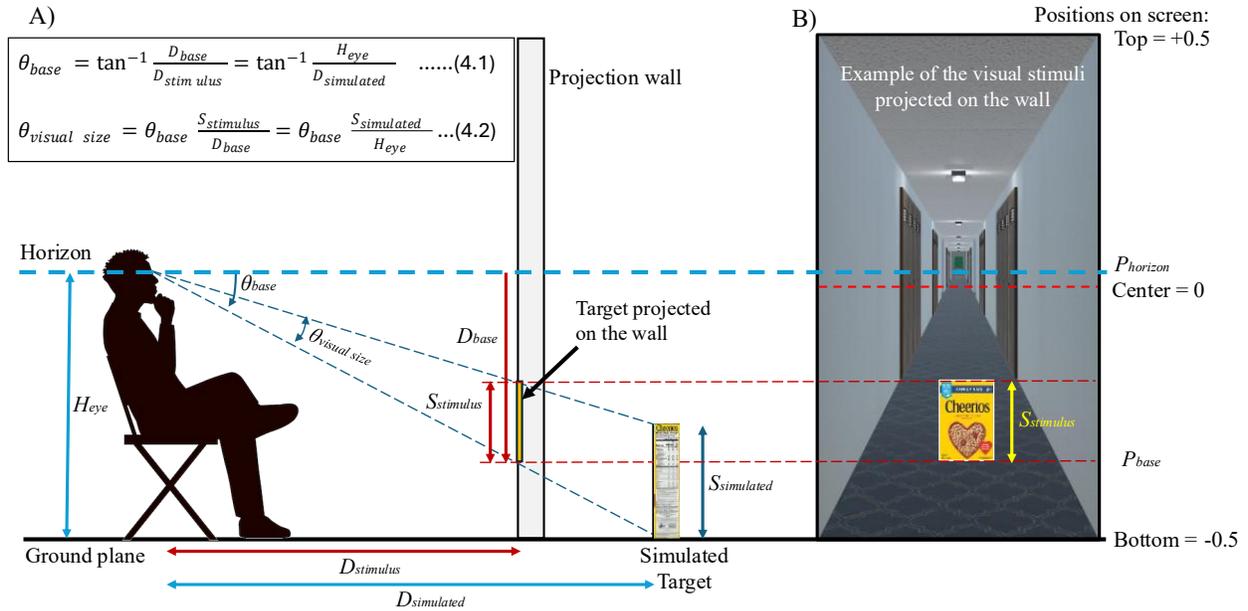


Figure 4.3: Target distance simulated in a projection. A) A diagram of the geometry of a target simulated to be at a set distance away from the observer presented in the projected scene. $S_{simulated}$ was the simulated target size, i.e., the size of the reference box, and $S_{stimulus}$ is its size portrayed on screen. $D_{simulated}$ is the simulated target distance and $D_{stimulus}$ is its distance portrayed on screen. P_{base} was the position of the target base on the screen simulating target distance, which was portrayed as its base at a distance (D_{base}) below the horizon. The horizon was at the observer's eye height (H_{eye}), i.e., the eye level. As shown in the equations, the target's angular position, i.e., the declination of the target base from the eye level (θ_{base}), and its angular visual size ($\theta_{visual\ size}$) were identical for both the simulated and stimulus target. B) An example of the visual stimuli projected on the wall simulating the target from A. The position on the screen was based on its horizontal length where the full length was 1 with the middle position set to 0, i.e., +0.5 was the top of the screen and -0.5 was the bottom of the screen. The horizon, i.e., the eye level, was at 0.015. The correct eye height was the distance between the horizon and the bottom of the screen (0.515).

With a fixed screen distance ($D_{stimulus}$), the target size on the screen ($S_{stimulus}$) varies in proportion to its simulated size ($S_{simulated}$) in accordance with its simulated distance ($D_{simulated}$) where $S_{stimulus}$ becomes larger with shorter $S_{simulated}$ and smaller with longer $S_{simulated}$. The stimulus size ($S_{simulated}$) can be represented as a scale factor (S_{scale}), a ratio between the stimulus size and the simulated size, which is equivalent to the ratio between the stimulus (i.e., distance to the screen) and the simulated distance. See Figure 4.4 for details.

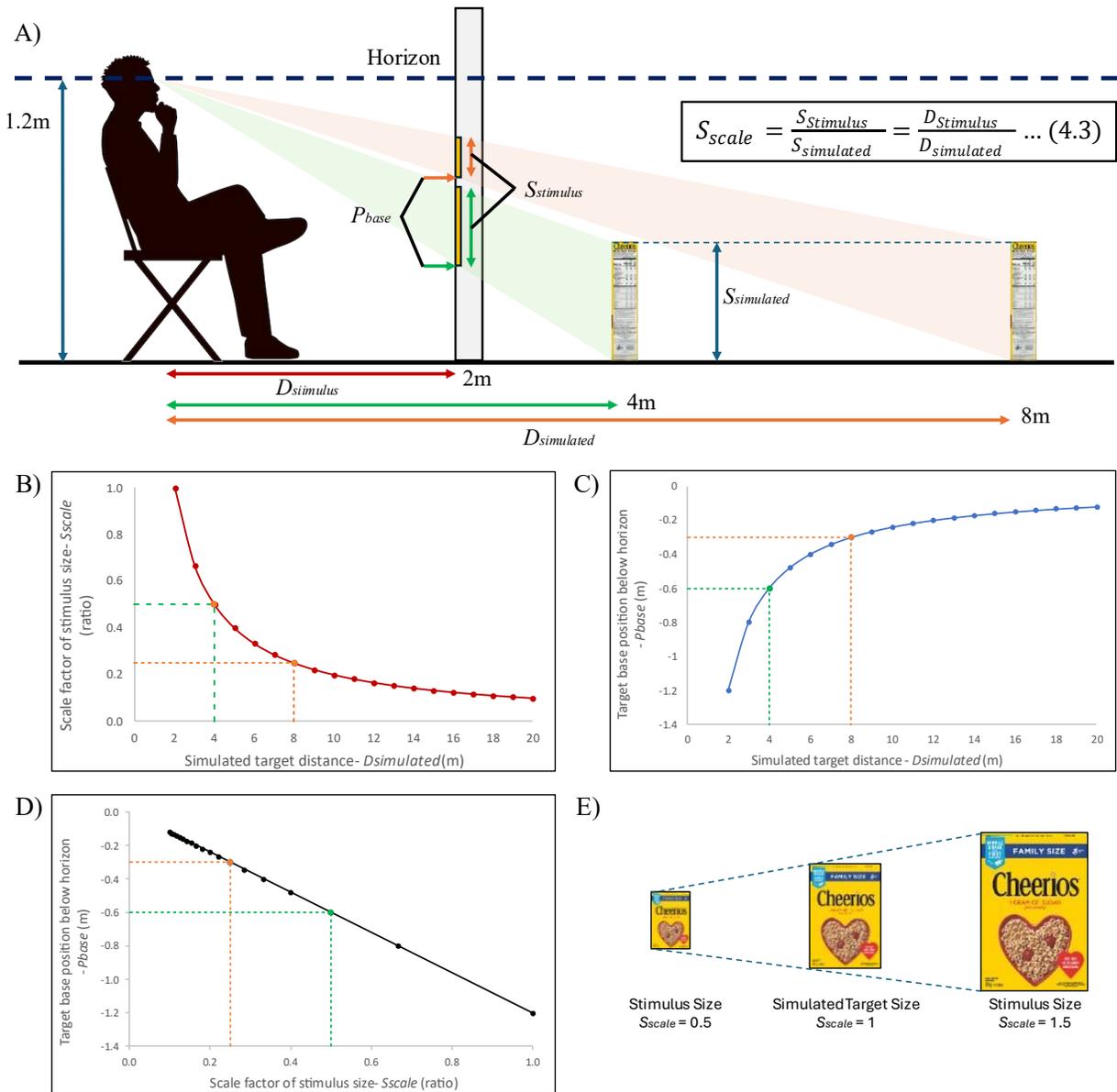


Figure 4.4: Relationship between stimulus size ($S_{stimulus}$), target base position (P_{base}), and simulated distance ($D_{simulated}$) presented graphically. The target size on the screen can be represented as a scale factor (S_{scale}) between the stimulus size ($S_{stimulus}$) and the simulated size ($S_{simulated}$) which is equivalent to the ratio between the stimulus distance ($D_{stimulus}$) and the simulated distance ($D_{simulated}$) as shown in Equation 4.3. A) Diagram of target stimuli displayed on a screen 2 m away. Targets simulated to be at 4 m (green) and 8 m (orange) are shown as examples. The plots represent the relationship between: B) stimulus size in scale factor (S_{scale}) and simulated distance ($D_{simulated}$), C) target base position (P_{base}) and simulated distance ($D_{simulated}$), and D) target base position (P_{base}) and stimulus size in scale factor (S_{scale}) as shown in Equation 4.5. In all plots, the green point represents target simulated at 4 m, and the orange point represents target simulated at 8 m. E) some examples of stimulus size ($S_{stimulus}$) in relation to the simulated size ($S_{simulated}$) using the scale factor (S_{scale}) following the Equation 4.3. The simulated target size is the same as the physical reference box size participants held during the experiment.

Similarly, the position of the target also varies in accordance with its simulated distance ($D_{simulated}$) where the target on screen moves farther below the horizon with shorter $D_{simulated}$ and closer to the horizon with longer $D_{simulated}$. Since the target was simulated to be on solid ground, its position (P_{base}) was represented by how far the bottom of the target was below the horizon.

The distance to the target's base below horizon in the scene (D_{base}) was computed with:

$$D_{base} = H_{eye} \times \frac{D_{stimulus}}{D_{simulated}} = H_{eye} \times S_{scale} \dots\dots\dots (4.4)$$

Therefore, the target position (P_{base}) was:

$$P_{base} = P_{horizon} - D_{base} = P_{horizon} - (H_{eye} \times S_{scale}) \dots\dots\dots (4.5)$$

which shows that stimulus size in scale factor (S_{ratio}) and target position (P_{base}) had an invariant linear relationship when representing the distance of an object on the screen.

When looking at a real world scene, the horizon and eye height is aligned in the visual field. When looking at a 2D scene (e.g., projected on a screen), however, the horizon (represented with perspective lines and a vanishing point; see Figure 4.5A) may not align with the eye height as shown in Figure 4.5B. For this reason, all participants had their chin on the chin rest to keep their eye height aligned with the horizon in the 2D scene presented in the present experimental set up. Figure 4.3B shows how positions on screen were measured and with the horizon ($P_{horizon}$) in the scene set up to be at the eye height (H_{eye} on screen was the distance between the horizon and the bottom of the screen, i.e., the ground). Figure 4.4 shows the model relationship between the target size in ratio (S_{ratio}), the position of the target base on screen (P_{base}), and its simulated distance ($D_{simulated}$) graphically where the scene was presented on a wall 2 m away. In the present test setup, the relationship between S_{scale} and P_{base} was:

$$P_{base} = 0.015 - (0.515 \times S_{scale}) \dots\dots\dots (4.6)$$

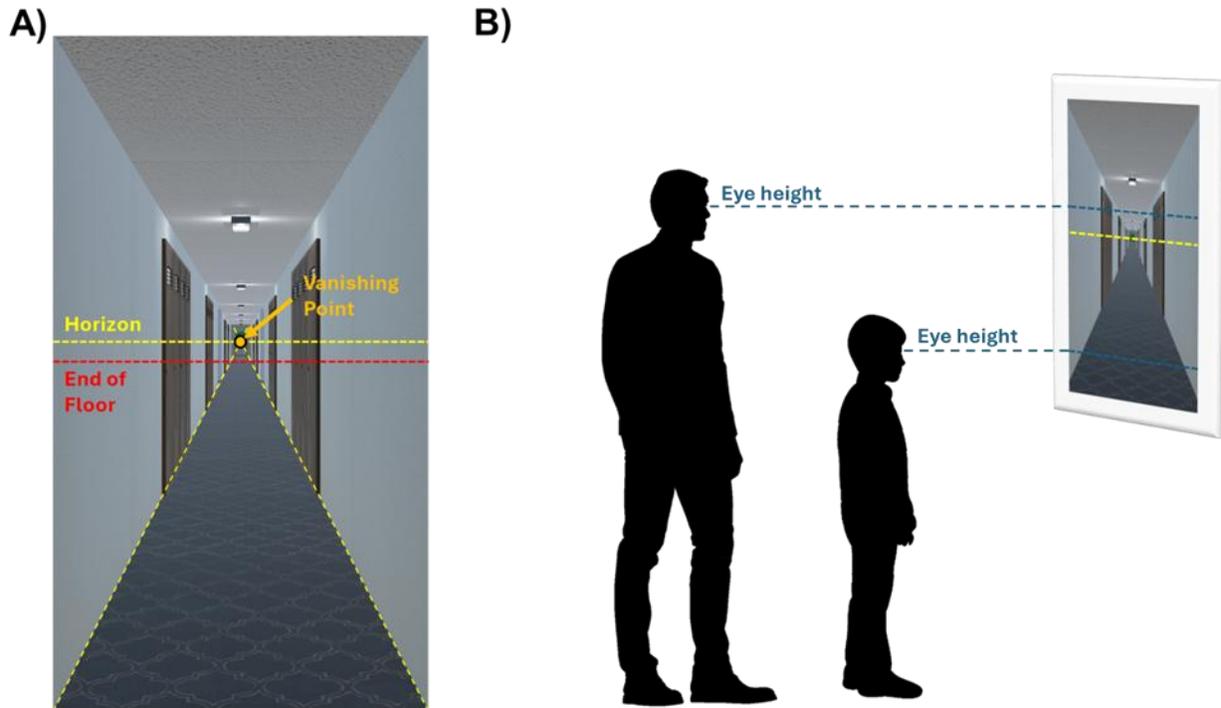


Figure 4.5: Diagram of the horizon in a visual scene and eye height of the observers. A) The horizon can be determined by using perspective as the perspective lines intersect at the vanishing point which aligns with the horizon. In the real world, the horizon also aligns with the eye height of the observer. B) The eye height of each observer can differ, and the horizon may not align with their observer's eye height when they look at a 2D scene if it is not set up to be aligned (e.g., in such a way my experiments were set up; see Figure 4.3).

4.3.1.3. Tasks.

There were two experimental tasks: Participants had to either adjust the target's position based on its size (Size-to-Position task) or adjust the target's size based on its position (Position-to-Size task).

4.3.1.3.1. The Size-to-Position task.

Participants viewed a target presented in the corridor scene sized ($S_{stimulus}$) to simulate experimental distances ($D_{simulated}$) while holding the reference box. Then they adjusted its position up or down using the hand-held mouse until the target appeared to be at the simulated distance for its visual size. The adjusted positions of the target on screen (P_{base}) were recorded. Refer to Figure 4.3B for details on how the target position on the screen was measured.

4.3.1.3.2. *The Position-to-Size task.*

Participants viewed a target presented in the corridor scene positioned (P_{base}) to simulate experimental distances ($D_{simulated}$) while holding the reference box. Then they adjusted its size using the hand-held mouse until the target appeared to be the same size as the reference box standing at the simulated distance. The adjusted sizes ($S_{stimulus}$) were recorded in scale factor (S_{scale}) following the Equation 4.3. These recorded sizes were between 0 (target not visible) and 1 (full size of the simulated box, i.e., the reference box) depending on how big the participant made the target on the projected screen, e.g., the recorded size was 0.5 if the participant made the target half the size of the reference box. Refer to the Section 4.3.1.2 for the full size of the physical reference box (i.e., size of 1).

4.3.1.4. *Analysis.*

A family-wise error rate of .05 was used for all analysis. For any ANOVA violating the assumption of sphericity, I adjusted the degrees of freedom using the Huynh-Feldt correction as the estimated epsilon (ϵ) was generally less than .75.

4.4. Experiment 1

4.4.1. *Materials and Methods*

4.4.1.1. *Overview.*

Here I report the methods for experiment 1 which share some similarities to the previous study that was done remotely conducted on people's personal home computer screens (Chapter 3) but with important differences. In this experiment, a target (a familiar cereal box) was presented in a virtual 2D hallway scene at various positions and sizes. Using the method of adjustment, participants adjusted target's position based on its size or its size based on its position to make it

appear to be the same size at a simulated distance as a physical reference box they were holding in their hands.

4.4.1.2. Participants.

Forty undergraduate students (20 females and 20 males, average age 19.9 years, SD = 3.04) at York University were recruited via Undergraduate Research Participant Pool (URPP). They had normal or corrected-to-normal vision. All participants signed a written informed consent form and were given course credit for their participation. The experiments were approved by the York University's Ethics Review Board.

4.4.1.3. Experimental Conditions.

There were four experimental variables: task, task order, viewing conditions, and target distances. The two tasks were Size-to-Position and Position-to-Size tasks and two possible task orders which separated participants into two experimental groups, Size-to-Position-to-Size group and Position-to-Size-to-Position group. Participants either started with the Size-to-Position task followed by the Position-to-Size task, or with the Position-to-Size task followed by the Size-to-Position task depending on which group they were in. Regardless of task order, participants' responses from the 1st task were used to determine the target size/position for the 2nd task. For example, when the Position-to-Size task was done first, the response sizes to which the participant set the targets were saved, averaged and used in the subsequent Size-to-Position task. Therefore, participants first determined the size at the geometrically correct target positions and then determined the position from those response sizes (Position-to-Size-to-Position). If people could reliably compute target size from its position and vice versa, the final position in the 2nd task should be the same as the initial position regardless of the size they had set the target to in the 1st task. The order was reversed for the other task order (Size-to-Position-to-Size).

There were two viewing conditions, binocular and monocular. For the binocular viewing condition, participants viewed the stimuli normally with both eyes. For the monocular viewing condition, they wore an eye-patch (Figure 4.1C) on their left eye. Half the participants did the experiment binocularly first then monocularly, the other half did the experiment monocularly first then binocularly. The order of the viewing conditions was counterbalanced within each task order group.

The experiment contained five simulated target distances: 4, 8, 12, 16, and 20m presented on a screen (i.e., a wall) which were simulated using either the target's visual size (for Size-to-Position task) or its position (for Position-to-Size task) in the scene. See Section 4.3.1.2. for details. The order of the target distances presented within each task block was randomized.

4.4.1.4. Procedure.

After participants signed the consent form, they sat on a chair in the test room (Figure 4.1A) with their chin on the chinrest and looked at the wall on which the visual stimuli were projected. After reading the instructions, they held a hand-held mouse (Figure 4.1B) in their right hand and the reference box (Figure 4.2B) with both hands. For each trial, participants fixated on a fixation cross (no scene) for 1 second until the target was shown with the test scene. Then, participants adjusted either the target position or its size depending on the task. They were instructed to make the target appear to be the same size as the reference box “as accurately as possible” at the distance it appears to be. After participants adjusted target to the position (Size-to-Position task) or the size (Position-to-Size task) they thought was correct for its visual size or position, they pressed a mouse button to record their response (P_{base} or S_{scale} respectively). The experiment then proceeded to the next trial. The experiment comprised 30 trials (5 target

distances \times 6 trials) in each block with total of four blocks (2 viewing conditions \times 2 tasks). The sequences are summarized in Figure 4.6.

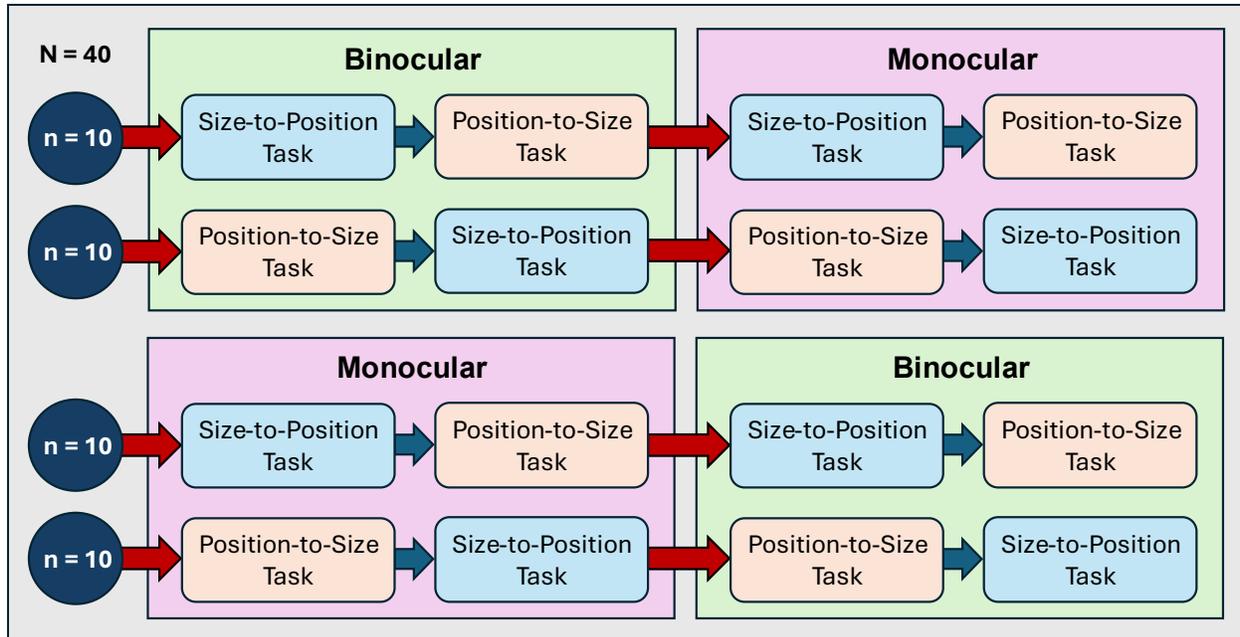


Figure 4.6: Experimental conditions and orders. Out of the 40 participants, 20 did the binocular condition first followed by the monocular condition and the other 20 did the opposite. Within each viewing condition order, 10 did the Size-to-Position task first then Position-to-Size task (Size-to-Position-to-Size group) in each viewing condition and the other 10 did the opposite order (Position-to-Size-to-Position group).

At the end of each viewing condition block, participants filled in a questionnaire (see Appendix B for the questions) which asked about the mental workload they required to do the task – NASA Task Load Index (Hart & Staveland, 1988) to evaluate whether participants it to be more difficult to the tasks either monocularly or binocularly. They were also asked to write down any strategies they used in an open answer format. The entire experiment lasted approximately 1 hour.

4.4.1.5. Data analysis.

There were two types of responses: response position (P_{base}) from the Size-to-Position task and response size (S_{scale}) from Position-to-Size task. Before conducting statistical testing,

each participant's responses were evaluated for any possible misunderstanding of the tasks and irrational responses, i.e., target positioned above the horizon ($P_{base} > 0.015$). I considered any such responses as errors and removed them from the data set. This data cleaning process removed 98 data points from the total data set (2.04%). As a result, one participant had to be removed from the analysis due to having an incomplete data set leaving 39 participants for the analysis.

Comparing the average responses (P_{base} and S_{scale}) between conditions allowed me to evaluate the effect of each experimental factor. To evaluate accuracy in participant responses, I also compared the S_{scale} and P_{base} against the geometrically correct size and position to evaluate how accurately participants adjusted target size or position relative to the correct values for the simulated distances ($D_{simulated}$) in each condition. If participants could accurately estimate a target's visual size from its position and its position from its visual size, then the size and the position from participants' responses should not differ from the correct values.

Because participant responses for the two tasks were from different paradigms, positions (P_{base}) in the Size-to-Position task and sizes (S_{scale}) in Position-to-Size task, participant's perceived relationship between target size and position (see Figure 4.4D for details) were evaluated from the intercept and slope between the stimuli and their responses in each condition. The P_{base} and S_{scale} from each task were fit with a linear model (Equation 4.5) to compute the intercept and slope and compared them between conditions to evaluate the effect of each experimental factor, and against the correct values (.015 for the horizon and 0.515 for the eye height; see Figure 4.3). Using these intercepts and slopes to estimate participant's perceived horizon ($P_{horizon}$) on the screen and eye height (H_{eye}), I evaluated whether the participants accurately judged the horizon in the scene and their eye height.

4.4.2. Results

4.4.2.1. Effect of Viewing Condition.

I compared the average response positions (P_{base}) and response size (S_{scale}) between task orders (Size-to-Position-to-Size and Position-to-Size-to-Position) – to explore any order effect, between viewing conditions (Monocular and Binocular conditions) – to evaluate whether viewing the stimuli monocularly or binocularly impact participant performances, and across simulated target distance (4, 8, 12, 16, and 20 m) – to verify they could recognize the relationship between distance and visual size (S_{scale}) or the target's position in the scene (P_{base}).

4.4.2.1.1. Comparing Response Position (P_{base}) Between Conditions.

A mixed-design ANOVA comparing the average response positions (P_{base}) from each participant between the task order, viewing conditions, and simulated target distance revealed no main effect of task order, $F(1, 37) = 2.074, p = .158, \eta^2 = .005$, or viewing condition, $F(1, 37) = .249, p = .621, \eta^2 < .001$. There was a main effect of simulated target distance, $F(1.244, 46.016) = 592.831, p < .001, \eta^2 = .815$, in which targets simulated to be further away (smaller S_{scale}) were positioned closer to horizon compared to those simulated to be at closer distances (larger S_{scale}). No significant interaction between any of the conditions was found. See Figure 4.7A.

4.4.2.1.2. Comparing Response Sizes (S_{scale}) Between Conditions.

A mixed-design ANOVA comparing the average response sizes (S_{scale}) from each participant between the task order, viewing conditions, and simulated target distance revealed no main effect of task order, $F(1, 37) = .175, p = .678, \eta^2 = .001$, or viewing condition, $F(1, 37) = .932, p = .341, \eta^2 < .001$. There was a main effect of simulated target distance, $F(1.263, 46.713) = 534.185, p < .001, \eta^2 = .698$, where targets simulated to be further away (positioned closer to the horizon) were made smaller in S_{scale} compared to those simulated to be at closer distances

(positioned further below the horizon). No significant interaction between any of the conditions was found. See Figure 4.7A.

4.4.2.1.3. Self-reported Mental Load between Viewing Conditions.

According to the NASA Task Load Index score interpretation (Prabaswari et al., 2019), with equal weight for each question, doing the task required ‘Somewhat high’ mental workload (scores between 30 and 49) in both viewing conditions on average. Comparing the scores using an independent t-test between the viewing conditions showed that participant found doing the task monocularly ($M = 49$, $SD = 14$) required more mental load compared to doing it binocularly ($M = 40$, $SD = 10$), $t(39) = 5.00$, $p < .001$, with a Cohen’s d of .79.

4.4.2.2. Comparing participant responses to the geometrically correct positions and sizes.

Because there was no statistical difference found in participants’ responses between the viewing conditions and task orders, I averaged them between the viewing conditions and combined the responses together between the two task order groups. I compared participant responses to the geometrically correct positions and sizes for each simulated target distance to evaluate how accurate they were at doing each task.

4.4.2.2.1. Size-to-Position Task.

The response positions (P_{base}) for each simulated target distance (4, 8, 12, 16, and 20 m) condition were compared to the geometrically correct positions using one sample t-tests. The analysis revealed no significant difference from the correct position for any of the simulated target distance: 4 m ($p = .734$), 8 m ($p = .457$), 12 m ($p = .210$), 16m ($p = .182$), 20 m ($p = .082$). The response data are plotted in Figure 4.7B.

4.4.2.2. Position-to-Size Task.

The response sizes (S_{scale}) for each simulated target distance (4, 8, 12, 16, and 20 m) condition were compared to the geometrically correct sizes using one sample t-tests. The analysis revealed that the response sizes were significantly larger than the correct size for all the simulated target distance: 4 m ($p = .003$), 8 m ($p < .001$), 12 m ($p < .001$), 16m ($p = .004$), 20 m ($p = .004$). The response data are plotted on Figure 4.7A.

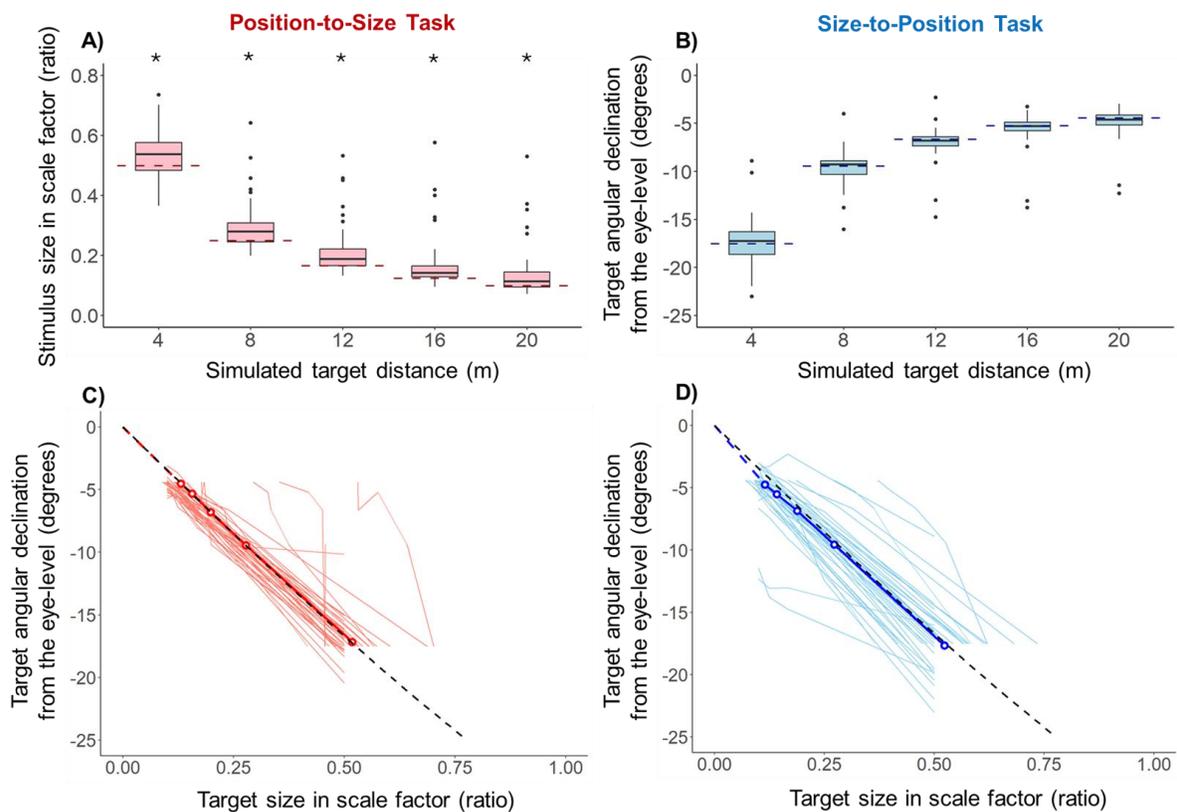


Figure 4.7: Exp 1 – Target size and position from each task. A) shows the response sizes and B) shows the response positions compared to the geometrically correct values for each simulated target distance. For each task, the response positions (Blue) and sizes (Red) are presented in box plots (with medians). C) shows the relationship between the presented target positions and response sizes and D) shows the relationship between the presented target sizes and response positions. The dashed lines represent target sizes and positions with the correct geometry. The light-colored lines are the average responses from each individual participant and the darker-colored lines with white dots represent the mean of responses from all participants for each task, where the dots are the means for each simulated target distance. For easier understanding of the data, response positions (P_{base}) were converted to the angular declination from the horizon (θ_{base} ; See Figure 4.3A).

4.4.2.3. Estimated Perceived Horizon and Eye Height.

The correlation coefficients (r) between the stimuli and responses were between -0.500 and -0.992 in the Position-to-Size task (see Figure 4.7C), and between -0.442 and -0.993 in the Size-to-Position task (see Figure 4.7D) which showed moderate to strong correlation between position and size ($|r| > 0.4$), except for 2 participants. Therefore, I estimated participants' perceived horizon ($P_{horizon}$) and eye height (H_{eye}) from the intercepts and slopes computed by fitting the target sizes (S_{scale}) and positions (P_{base}) in each condition with a linear model (see Section 4.4.1.5).

I compared the intercepts and slopes between viewing conditions (Monocular and Binocular conditions) – to evaluate whether viewing the stimuli monocularly or binocularly impact participant perception of the relationship between size and position, and tasks (Size-to-Position task and Position-to-Size task) – to test whether participants' perceived distances were task dependent. The results from analyzing intercepts and slopes are shown in the following sections and plotted in Figure 4.8.

4.4.2.3.1. Comparing Intercepts.

A mixed-design ANOVA comparing the average intercepts between viewing conditions, and tasks revealed no main effect of viewing condition, $F(1, 38) = 2.499, p = .122, \eta^2 = .025$, but there was a significant main effect of task, $F(1, 38) = 4.442, p = .042, \eta^2 = .046$. No significant interaction between the conditions was found.

When compared to the correct horizon position of .015 (see Figure 4.3), using one sample t-tests, the intercepts for both Size-to-Position task ($M = -.003, SD = .044$), $t(38) = -2.558, p$

= .015, *Cohen's d* = -.410, and for Position-to-Size task ($M = .003$, $SD = .033$), $t(38) = -2.340$, $p = .025$, *Cohen's d* = -.375, were significantly different from .015.

4.4.2.3.2. Comparing Slopes.

A mixed-design ANOVA comparing the average slopes between viewing conditions, and tasks revealed no main effect of viewing condition, $F(1, 38) = 1.929$, $p = .173$, $\eta^2 = .021$, or task, $F(1, 38) = 0.096$, $p = .758$, $\eta^2 < .001$. No significant interaction between the conditions was found.

Compared to the correct eye height of 0.515 (see Figure 4.3), using one sample t-tests, slopes for Size-to-Position task ($M = .474$, $SD = .129$) were not significantly different, $t(38) = 1.993$, $p = .053$, *Cohen's d* = .319, but the slopes for Position-to-Size task ($M = .461$, $SD = .104$) were significant difference compared to 0.515, $t(38) = 3.271$, $p = .002$, *Cohen's d* = .524.

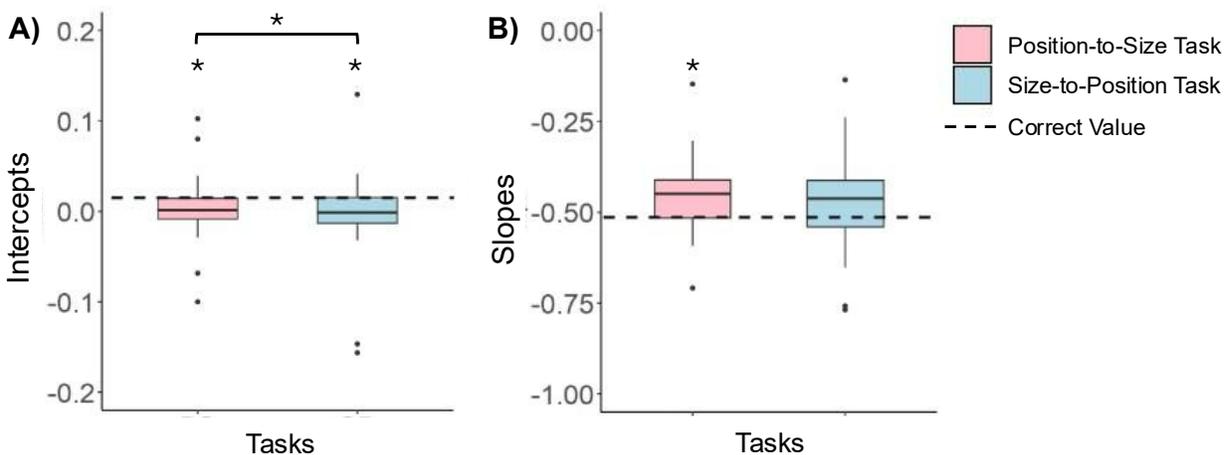


Figure 4.8: Exp 1 – Intercepts and Slopes. A) Mean intercepts from participant responses compared between tasks (Position-to-Size task and Size-to-Position task). These intercepts were also compared to the correct horizon on screen (.015). B) Mean slopes from participant responses compared between tasks. These slopes were also compared to the correct eye height on screen (.515). The dashed lines are the correct intercept for A and the correct slope for B.

4.4.3. Discussion of Experiment 1

These results show that participants accurately placed targets in the scene using their visual sizes (Size-to-Position task) but made the targets too big when adjusting their sizes using their positions (Position-to-Size task). There was no difference found between the task orders, therefore these results appear to be strictly task-dependent. Considering that participants in my past study (Chapter 3) could not accurately derive target size/position from their simulated distance in any conditions, the visual stimuli and setup used in the present study seem to have improved participant performance to some degree.

Figure 4.9 highlights the differences in participant performance between the past (Chapter 3) and the present study. In the past study (Chapter 3), the target sizes were set bigger on average in the Position-to-Size task and the positions were set higher on average (closer to the eye level) in the Size-to-Position task compared to the correct values. They were also much noisier and deviated further from the correct answers (dashed black line) compared to the present study. I hypothesized that (H5) people would be more accurate at inferring an object's absolute distance with my semi-immersive 2D scene displayed on a large screen than on computer screens, and the results from the present study in comparison to the past study (Chapter 3) seem to support this hypothesis.

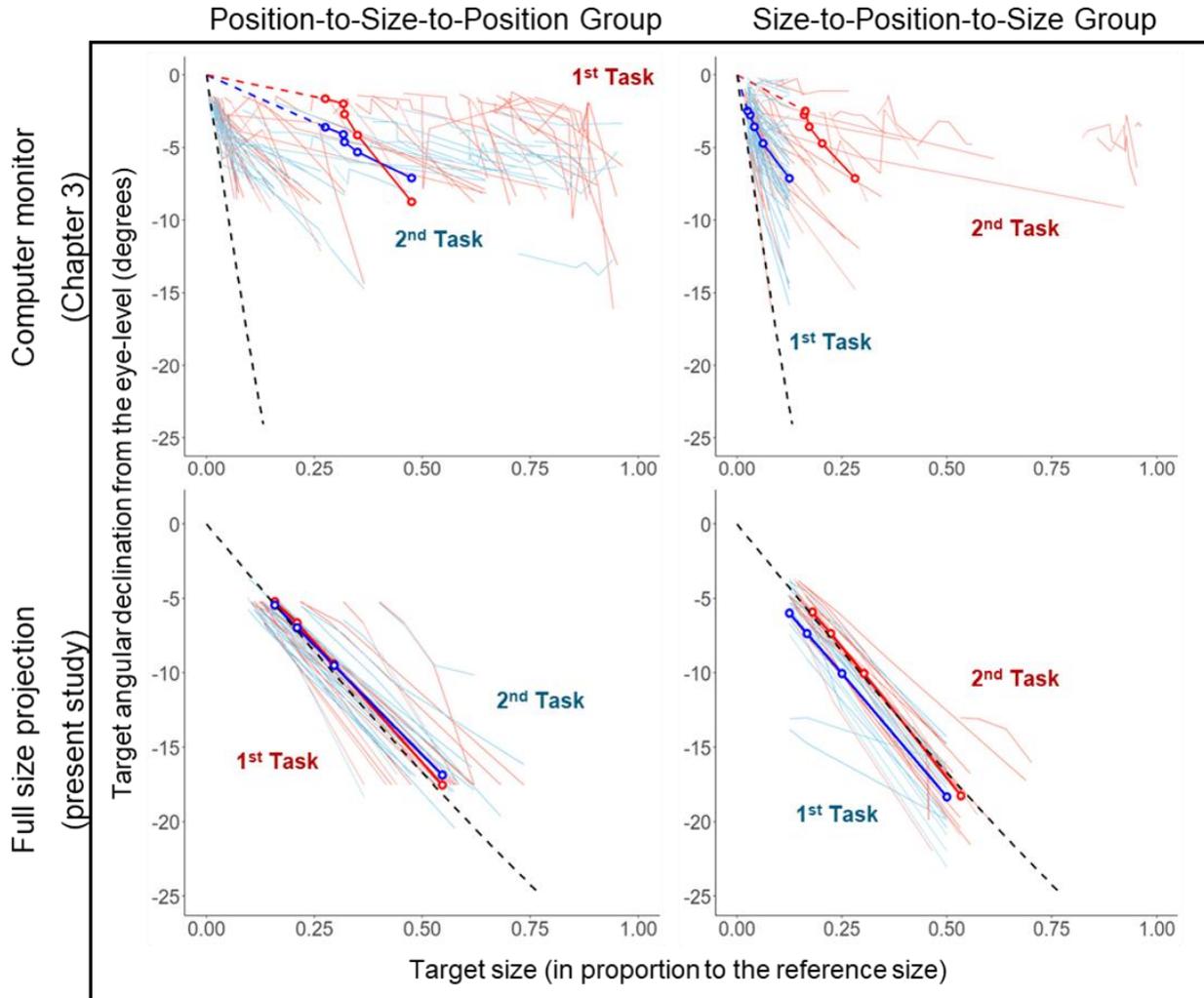


Figure 4.9: Participants' responses in the past and the present experiments. Target positions are plotted as a function of target size. Top row is from the past experiment where the experiment was conducted online on computer screens (Chapter 3); bottom row is from the present experiment where the experiment was conducted in lab on a full-scale projection. Left column, the Position-to-Size task first followed by the Size-to-Position; right column, the Size-to-Position task first followed by the Position-to-Size task. The dashed black lines represent target sizes and positions with the correct geometry. The solid light-colored lines are the average responses from each individual participant and the darker-colored lines with white dots represent the mean of responses from all participants for each task, where the dots are the means for each simulated target: the Size-to-Position task (Blue) and the Position-to-Size task (Red).

Comparing the participants' responses to the correct position and size showed that targets were made too big in Position-to-Size task (Figure 4.7A), but they were placed at the correct positions in Size-to-Position task (Figure 4.7B) which supports my hypothesis that (H6) people would be less accurate in the Position-to-Size task than in Size-to-Position task. These results

suggest that the additional distance cues, e.g., ground information, in the present study improved participant performances in the Size-to-Position task more than in the Position-to-Size task.

My hypothesis that (H7) people would be more accurate in the Size-to-Position task when doing the tasks monocularly compared to doing them binocularly was not supported by the results as participant responses did not differ between the two viewing conditions in either task. Interestingly, despite the fact that binocular vision could not help in any way in determining depicted distance in a 2D scene and even provides conflicting information from the screen itself, participants felt it required less mental load when doing the tasks binocularly compared to doing them monocularly which may be because participants are not used to viewing an object with just one eye.

4.4.3.1. Difference between the tasks.

There is some evidence in the literature that distance information impacts size and distance perception differently (Meehan & Triggs, 1992; Yoo et al., 2023). The differences in participant accuracy I report here between the tasks add further support to this idea. The difference in the participants' performance between the two tasks in both studies, the past study with and without familiar size in the scene (Chapter 3) and the present study with various visual distance cues in the 2D scene, may indicate that size and distance judgements are fundamentally different tasks requiring different strategies and being vulnerable to different sources of error. In the Size-to-Position task, participants first must judge the target's size to estimate its distance then move the target to the position in the scene accordingly. Therefore, it may be more dependent on correctly judged size aided by additional distance information, such as familiar object, displayed on screen as suggested by Meehan and Triggs (1992). In the Position-to-Size task, however, some of these distance cues may not have been useful enough to judge target

position (angular declination below eye level) and its distance correctly, resulting in participants estimating the visual size for its distance erroneously. Where might this error come from?

4.4.3.1.1. Misperceived Size?

Errors in the Position-to-Size task may not be from misperceived distance but instead may arise from misperceived size while re-sizing the target. If targets were perceived as smaller than they actually were but perceived to be the correct size at the distances at which they were simulated, participants would then have had to make them bigger for the target sizes to appear accurate. However, if the error was from misperceived size instead of distance, then the targets in the Size-to-Position task would have been perceived smaller as well and participants should have placed targets further away (smaller angular declination from eye level). Instead, my participants were accurate at placing targets in the scene based on their sizes and therefore it is unlikely the errors in the Position-to-Size task were from misperceived size rather from misperceived distance.

4.4.3.1.2. Misperceived Horizon and Eye Height?

The visual angle of an object increases as it gets closer to the observer. Setting the targets as larger in the Position-to-Size task therefore suggests that they may have been perceiving the target as closer than simulated. This is consistent with other studies where perceived distance has generally been found to be compressed compared to the real distance (Gilinsky, 1951; Z. Li et al., 2011; Stevens, 1957; Viguier et al., 2001). Estimating the horizon as too low results in misjudging the angular declination below eye level toward the target base. Dixon and colleagues (2000) suggested that the elevation of the horizon in a scene viewed on a non-immersive display is indeterminate and can lead to incorrectly scaled eye level. The estimated perceived horizons (intercepts; see Section 4.4.2.3) in the present experiment were indeed lower than the actual

simulated horizon height in the scene which may have led to the participants incorrectly scaling their eye level. The perceived horizons differed significantly between the tasks where they were judged as higher in the Position-to-Size task than in the Size-to-Position task. The estimated perceived eye height (slopes; see Section 4.4.2.3) was also shorter than the correct eye height in the Position-to-Size task, but not in the Size-to-Position task. These results may suggest participants' perceived horizon and eye height depended on the task, but this is unlikely as the scene presented in this experiment was identical in all conditions. What is more likely is that their perceived horizon and eye height in the scene influenced their responses differently in the two tasks.

The perceived horizons and eye heights were estimated using Equation 4.5 which assumes that participants recognized and used the linear relationship between target size and position demonstrated in Figure 4.4. Although participant performances plotted in Figure 4.9 also appear to suggest a linear relationship, I only used five simulated distances between 4 m and 20 m with gaps of 4 m in between target distances. To further verify whether participants estimated the elevation of horizon in the scene correctly and hence their eye level and their eye height, I conducted experiment 2 using the full range (between 2 m and 25 m) with more simulated distance within the hallway scene.

4.5. Experiment 2

Purpose of experiment 2 was to verify the intercepts and slopes computed in experiment 1 using more spread-out distances for my simulated targets. Whereas experiment 1 had used the method of constant stimuli, here I used the continuous tracking method following the continuous psychophysics paradigm Bonnen et al. (2015) introduced which potentially shows the relationship between target size and the position on the screen better as this method utilizes the

full length of the simulated hallway with more simulated distances than just the five distances used in experiment 1. This method also allowed collecting more participant responses in a shorter time.

I hypothesized that intercepts and slopes from continuous tracking would not differ from those obtained using the method of adjustment which would verify that five simulated target distances were sufficient for computing the linear relationship between target size and position from fitting them to a linear model. I also hypothesized that there would be a difference in intercepts between my two tasks as was revealed in experiment 1 (and in the experiment of Chapter 3). When estimating the perceived horizon from the intercepts, the results of Experiment 2 were expected to further support the hypothesis that people use the horizon information differently in the two tasks resulting in different participant performance.

4.5.1. Materials and Methods

4.5.1.1. Overview.

Here I report the methods for the experiment 2 using continuous tracking. Participants also did the tasks using the previous method of adjustment (used in Experiment 1 and Chapter 3) to verify whether the new method is compatible to the traditional method. Similar to in experiment 1, participants again adjusted a target's position based on its size, or its size based on its position to make it appear to be the same size as a physical reference box they were holding in their hands. However, in the method of continuous tracking the target size or position was continuously changing whereas in the method of adjustment the target size or position were constant while participants adjusted its position in the Size-to-Position task, or its size in the Position-to-Size task.

4.5.1.2. Participants.

Twenty-seven undergraduate students (17 females and 10 males, average age 21.9 years, $SD = 3.78$) at York University were recruited via Undergraduate Research Participant Pool (URPP). They had normal or corrected-to-normal vision. All participants signed a written informed consent form and were given course credit for their participation. The experiments were approved by the York University's Ethics Review Board.

4.5.1.3. Experimental Conditions.

There were two experimental methods used, the Method of Adjustment and Continuous Tracking. Because the continuous tracking method requires participants to be familiar with using the hand-held mouse for their motor control to adjust target size or position, the method of adjustment condition was conducted first followed by the continuous tracking condition. Identical to the arrangement in experiment 1, there were two tasks, Position-to-Size task and Size-to-Position task. See Section 4.3.1.3 for details.

4.5.1.3.1. Method of Adjustment.

Identical to experiment 1, the method of adjustment condition contained five simulated target distances: 4 m, 8 m, 12 m, 16 m, and 20 m presented one at a time on a screen which were simulated using either the target's visual size (for Size-to-Position task) or its position (for Position-to-Size task) in the scene. See Section 4.3.1.2 for details. Because I did not find any order effect in experiment 1, the order of the tasks was always the Position-to-Size task followed by the Size-to-Position task. The order of target distances presented within each task block was randomized.

4.5.1.3.2. *Method of Continuous Tracking.*

For the method of continuous tracking, target size or position was continuously changed during each block (total of 10 blocks; 5 blocks of Size-to-Position task and 5 blocks of Position-to-Size task). A block started with simulated target distances at 4 m, 8 m, 12 m, 16 m, or 20 m and the target's size or position changed in a random walk every 0.5 seconds such that the new target size (S_{scale}) or position (P_{base}) was the previous value plus a value determined from a normal distribution with a mean of 0 and a standard deviation of .0125 within a range of ± 0.05 for size (maximum change of $\sim 10\%$ of the full size) and ± 0.025 for position (maximum change of $\sim 10\%$ of the full range of the position). To prevent the target from not appearing to be on the ground, or from going outside the screen, its position (P_{base}) was limited to be between -0.024 (end of the hallway) and -0.5 (bottom of the screen). To prevent the target from becoming too large, its size was also limited to be smaller than 1.5 (150 % of the reference box size). Each block lasted for 1.5 minutes, and the order of the blocks was randomized.

4.5.1.4. *Procedure.*

After participants signed the consent form, they sat on a chair in the test room (Figure 4.1A) with their chin on the chinrest and looked at the wall on which the visual stimuli were projected. After reading the instructions, they held a hand-held mouse (Figure 4.1B) in their right hand and the reference box (Figure 4.2B) with both hands. They first did the experiment using the method of adjustment, which started with the Position-to-Size task followed by the Size-to-Position task. For each trial, participants fixated on a fixation cross for 1 second until the test scene was shown with a target. In the Position-to-Size task participants adjusted the target size depending on its position. When participant thought the target was the correct size, they pressed a mouse button to record their response (S_{scale}). The experiment proceeded to the next trial until

all 20 trials were complete, then was followed by the Position-to-Size task. In the Position-to-Size task, participants adjusted the target position depending on its size. When participant thought the target was at the correct position, they pressed a mouse button to record their response (P_{base}). The experiment proceeded to the next trial until all 20 trials were complete. The method of adjustment comprised of 40 trials (2 tasks \times 5 target distance \times 4 trials).

Participants then did the experiment using the method of continuous tracking. For each block, they were presented with a target in the scene simulating one of five starting distances. After 10 seconds, the target's position or its size (depending on the task) started changing every 0.5 seconds. Participants had to continuously adjust the target size depending on its current position (Size-to-Position task) or adjust its position depending on its current size (Position-to-Size task) using the rollerball on the hand-held mouse to keep it the same size as the reference block. Participant responses were recorded automatically at each frame (60 Hz). The continuous tracking comprised 10 blocks (2 tasks \times 5 starting distances) and each block was 1.5 minutes.

At the end of the experiment, participants filled in a questionnaire (see Appendix B for the questions) which asked about the mental workload they required to do the task – NASA Task Load Index (Hart & Staveland, 1988) to assess how difficult participants thought the task was.

4.5.1.5. Data analysis.

Before the analysis, each participant's responses were evaluated for any possible misunderstanding of the tasks and irrational responses using the following criteria.

4.5.1.5.1. *Post-test data cleaning.*

For the method of adjustment, I considered any such responses as errors and removed them from the data set. This data cleaning process removed 0 data points from the total data set in experiment 2. However, one participant was removed from the data set for responding in the

opposite way from the geometrically correct answers in one of the tasks, i.e., placing larger target closer to the horizon and smaller target further away from the horizon.

For the continuous tracking method, I considered that the participant did not perform the task properly if the time lag of highest correlation fell outside the interval between 0 and 3 seconds. Those blocks were removed from the data set resulting in 10 blocks from the total blocks (3.85%). No participants were removed from this process.

4.5.1.5.2. Participant Response Analysis.

There were two experimental variables: experimental methods (Method of Adjustment vs. Continuous Tracking) and task (Size-to-Position vs. Position-to-Size). I computed interceptions and slopes for each condition using the same method as experiment 1, i.e., by fitting target size (S_{scale}) and position (P_{base}) into a linear model using Equation 4.5. I compared the average intercepts and slopes between the experimental conditions to evaluate contribution of each of these factors on participants' perceived horizon in the scene and their eye height. I also compared them against the correct values (0.15 for the horizon and 0.515 for the eye height) to evaluate how accurately they could estimate where the horizon was and their eye height in the scene.

4.5.1.5.3. Cross-correlogram Analysis.

For continuous tracking, I employed a cross-correlogram analysis to estimate how fast participants reacted to the changing stimulus size or position following a procedure from the paper by Bonnen et al. (2015). Cross-correlogram analysis compares the correlations between the stimulus and participants' responses for different delays between the stimulus and responses. From the analysis, I picked the time lag with the highest correlation as indicating how long participants needed to adjust the target according to the stimulus (see Figure 4.10A). Taking these lags (i.e., reaction times) from the continuous tracking task into account, participant

responses were recorded as being delayed by the time lag with the highest correlation for each block (see Figure 4.10B) to compute their responses to the stimulus size or position at the time.

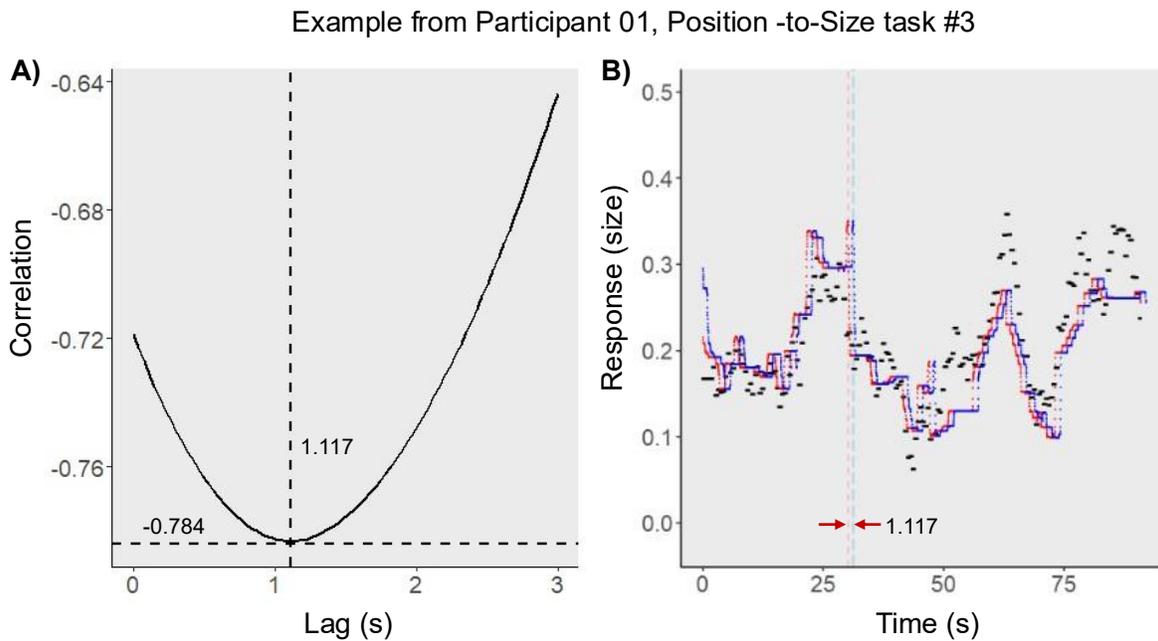


Figure 4.10: Example of cross-correlation analysis and participant responses. A) Cross-correlogram from participant 01, Position-to-Size task block #3. The highest correlation between correct target sizes, based on the target position presented, and participant response sizes was -0.784 (dashed horizontal line), which associated with the time lag of 1.117 (dashed vertical line). B) The correct target sizes (in black) and participant responses (in blue) plotted over the course of the block. The response sizes were delayed by 1.117 seconds (time lag with the highest correlation computed from the cross-correlation analysis shown in A) were plotted (in red) to show the participant's performance with the reaction time considered. The delay in participant responses is represented with dashed vertical lines (in light blue and light red).

4.5.2. Results

4.5.2.1. Evaluating Accuracy in Method of Adjustment Responses.

To verify participants responded in the same manner in experiment 2 as those in experiment 1, before comparing them to the new method of continuous tracking, I evaluated their accuracy with the same analysis conducted in experiment 1 comparing response sizes (S_{scale}) and positions (P_{base}) to the geometrically correct positions and sizes.

4.5.2.1.1. *Position-to-Size Task.*

The average response sizes (S_{scale}) for simulated target distance condition 8 m was compared to the geometrically correct positions using Wilcoxon signed-rank test due to violating normality assumption ($p = 0.041$ in Shapiro-Wilk Test of Normality). The analysis revealed significant different from the correct size for the target simulated to be at 8 m ($p = .003$). One sample t-tests conducted for the rest of the simulated target distance conditions revealed significant difference from the correct size for: 12 m ($p = .004$), 16 m ($p = .003$), and 20 m ($p = .002$). The response size data are plotted in Figure 4.11A.

A repeated-measures ANOVA comparing the average response sizes (S_{scale}) between simulated target distances (4, 8, 12, 16, and 20 m) revealed a significant main effect of simulated target distance, $F(1.204, 32.295) = 319.564, p < .001, \eta^2 = .925$, where targets simulated to be further away (positioned closer to the eye level) were made smaller in visual size compared to those simulated to be closer (positioned further below the eye level).

4.5.2.1.2. *Size-to-Position Task.*

The average response positions (P_{base}) for simulated target distance conditions 4, 8, 12, and 16 m were compared to the geometrically correct positions using Wilcoxon signed-rank test due to violating normality assumption ($p < 0.05$ in Shapiro-Wilk Test of Normality). The analysis revealed significant difference from the correct position for the simulated target distances: 4 m ($p = .049$), 8 m ($p = .038$), and 16 m ($p = .019$). A one sample t-test conducted for the simulated target distance condition 20 m revealed no significant difference from the correct position ($p = .358$). The response position data are plotted in Figure 4.11B.

A repeated-measures ANOVA comparing the average response positions (P_{base}) from between simulated target distances (4, 8, 12, 16, and 20 m) revealed a significant main effect of

simulated distance, $F(1.254, 32.606) = 126.486$, $p < .001$, $\eta^2 = .829$, in which targets simulated to be further away (smaller visual size) were positioned closer to eye level compared to those simulated to be closer (larger visual size).

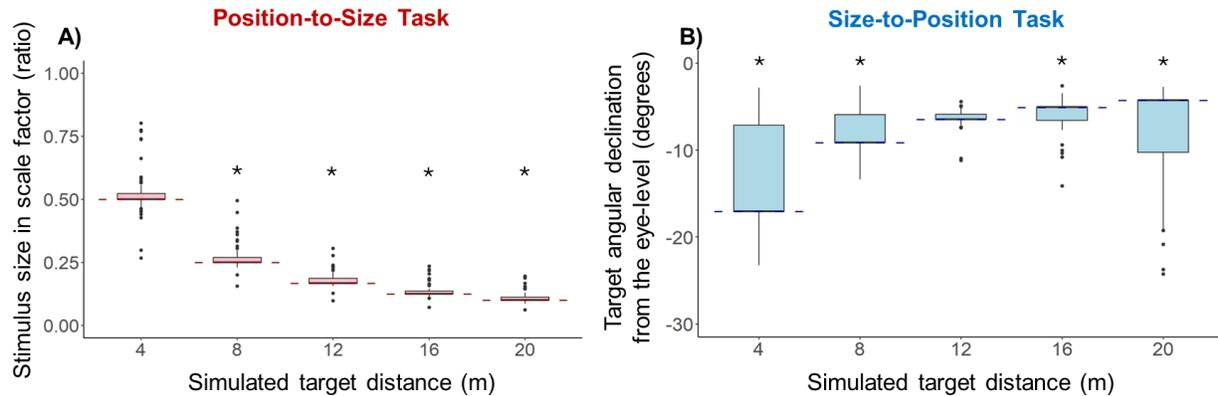


Figure 4.11: Exp 2 – Target size and position from each task. A) shows the response sizes and B) shows the response positions compared to the geometrically correct values for each simulated target distance. For each task, the response positions (Blue) and sizes (Red) are presented in box plots (with medians). The dashed lines represent target sizes and positions with the correct geometry. For easier understanding of the data, response positions (P_{base}) were converted to the angular declination from the horizon (θ_{base} ; See Figure 4.3A).

4.5.2.2. Comparing Between Experimental Methods.

For each participant, the linear correlation (Pearson Correlation Coefficient; r) between target size (S_{scale}) and position (P_{base}) for each task was computed. Correlations for Method of Adjustment were between -0.619 and -0.997 in the Size-to-Position task, -0.781 and -0.998 in the Position-to-Size task. For Continuous Tracking, the correlations were between -0.771 and -0.982 in Size-to-Position task, -0.743 and -0.991 in Position-to-Size task. Since all the correlation coefficients (r) showed strong correlation between position and size ($|r| > 0.5$), I estimated participants' perceived horizon ($P_{horizon}$) and eye height (H_{eye}) from intercepts and slopes from fitting a linear model using Equation 4.5 with response sizes (S_{scale}) and positions (P_{base}) for each condition (see Section 4.4.1.5).

4.5.2.2.1. *Evaluating Intercepts.*

A repeated-measures ANOVA comparing the average intercepts between test methods and tasks revealed a significant main effect of experimental method, $F(1, 25) = 11.372, p = .002, \eta^2 = .204$, where further evaluation revealed a significant simple effect of experimental method in both the Position-to-Size task, $t(25) = 3.167, p = .004, \text{Cohen's } d = .621$ and the Size-to-Position task, $t(25) = 2.900, p = .008, \text{Cohen's } d = .569$. There was no main effect of task found, $F(1, 25) = 1.953, p = .174, \eta^2 = .014$. See Figure 4.12A. No significant interaction between test method and task was found.

Compared to the correct value of .015, the intercepts for the Size-to-Position task ($M = .008, SD = .019$) did not significantly differ, $t(25) = -1.921, p = .066, \text{Cohen's } d = -.377$. However, they were significantly different for the Position-to-Size task ($M = .001, SD = .015$), $t(25) = -4.488, p < .001, \text{Cohen's } d = -.880$, in method of adjustment and for both Position-to-Size task ($M = -.010, SD = .019$), $t(25) = -6.869, p < .001, \text{Cohen's } d = -1.347$, and for Size-to-Position task ($M = -.009, SD = .024$), $t(25) = -5.229, p < .001, \text{Cohen's } d = -1.347$, in method of continuous tracking.

4.5.2.2.2. *Evaluating Slopes.*

A repeated-measures ANOVA comparing the average slopes between test methods and tasks revealed a significant main effect of test method, $F(1, 25) = 7.959, p = .009, \eta^2 = .146$, where further evaluation revealed significant simple effect of experimental method in Size-to-Position task, $t(25) = -2.927, p = .007, \text{Cohen's } d = .160$, but not in Position-to-Size task, $t(25) = -1.889, p = .071, \text{Cohen's } d = .227$. There was also a significant main effect of task, $F(1, 25) = 4.370, p = .047, \eta^2 = .034$, but further evaluation revealed no significant simple effect of task in

either method of adjustment or continuous tracking conditions. See Figure 4.12B. No significant interaction between test method and task was found.

Compared to the correct value of 0.515, the slopes for Size-to-Position task ($M = .500$, $SD = .140$) did not significantly differ, $t(25) = .550$, $p = .587$, *Cohen's d* = .108. However, they were significantly different for Position-to-Size task ($M = .451$, $SD = .116$), $t(25) = 2.802$, $p = .010$, *Cohen's d* = .550, in method of adjustment and for both Position-to-Size task ($M = .409$, $SD = .082$), $t(25) = 6.609$, $p < .001$, *Cohen's d* = 1.296, and for Size-to-Position task ($M = .420$, $SD = .068$), $t(25) = 7.193$, $p < .001$, *Cohen's d* = 1.411.

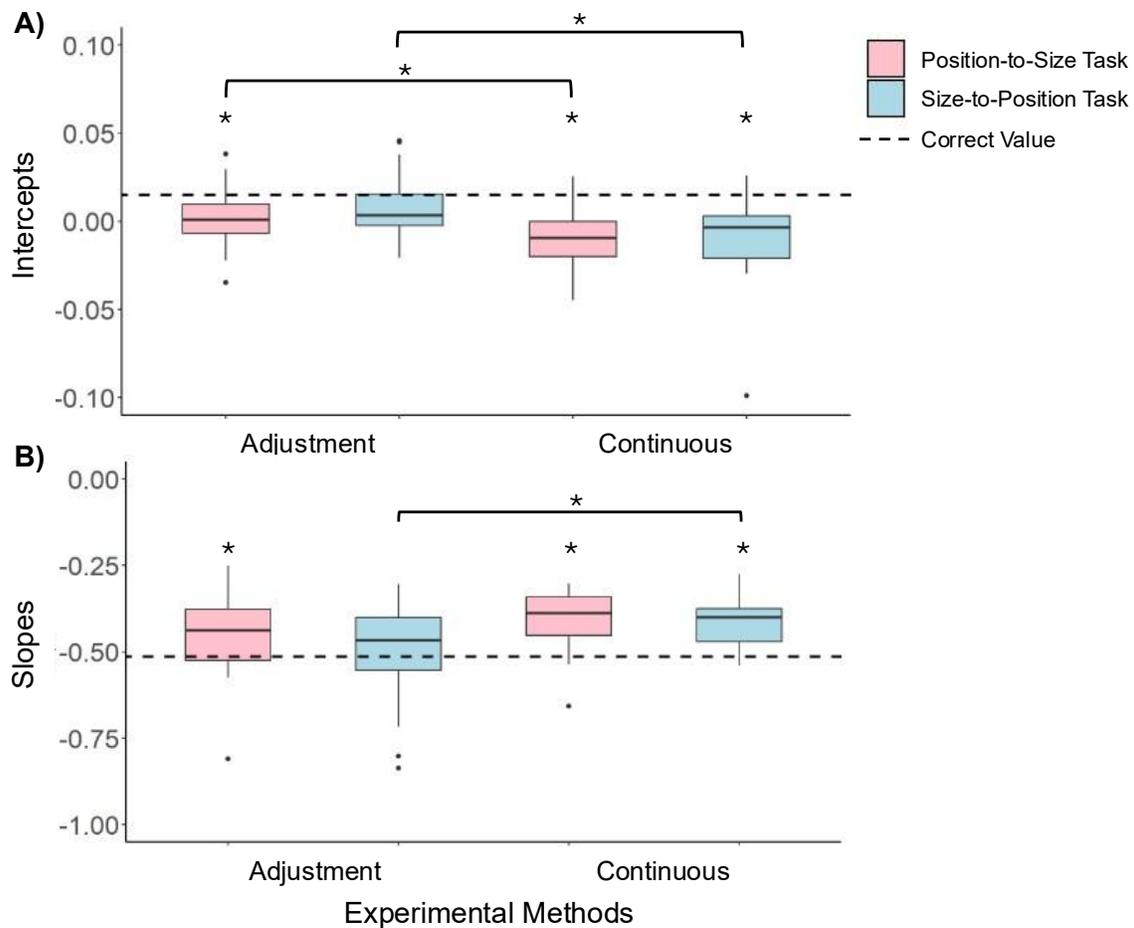


Figure 4.12: Exp 2 - Intercepts and Slopes. The top row (A) shows the intercepts for Position-to-Size task (pink) and Size-to-Position task (blue) compared to the correct horizon on screen (0.015; see section 4.4.2.3). The bottom row (B) shows the slopes for Position-to-Size task (pink)

and Size-to-Position task (blue) compared to the correct slope (0.515; see section 4.4.2.3). The intercepts and slopes were also compared between the two experimental methods: the method of adjustment and continuous tracking.

4.5.2.2.3. Comparing Correlations between Experimental Methods and Tasks.

To compare correlations in participant responses between conditions, they were normalized using Fisher's Z-transform (Fisher, 1915), transforming Pearson correlation coefficient into Z-scores to allow hypothesis testing. A repeated-measures ANOVA comparing the correlations between test methods and tasks revealed a significant main effect of experimental method, $F(1, 25) = 53.388, p < .001, \eta^2 = .442$, but no effect of task, $F(1, 25) = 1.142, p = .295, \eta^2 = .009$. No significant interaction between test method and task was found.

4.5.2.3. Comparing Reaction Times from Continuous Tracking Method.

A paired sample t-test comparing the time lags from continuous tracking method with highest correlation between tasks revealed no effect of task, $t(25) = -.816, p = .422, \text{Cohen's } d = -.160$. The mean time lag for Position-to-Size task was 0.83 seconds (SD = 0.34) and for Size-to-Position task it was 0.76 seconds (SD = 0.34).

4.5.3. Discussion of Experiment 2

Before comparing the two methods, I evaluated participant responses measured using the method of adjustment (using the same statistical analysis as in experiment 1) to verify that participants in experiment 2 responded the same way as those in experiment 1. Participant response sizes in Position-to-Size task were made too big compared to the correct sizes. The positions in Size-to-Position task were placed too high compared to the correct positions for shorter simulated distances (4 and 8 m) and too low for the further simulated distances (16 and 20 m), which was different from experiment 1 where the response positions did not differ from the geometrically correct positions. Such results suggest that the participants in experiment 2

judged the targets with larger visual angles (i.e., simulated to be at closer distances; $D_{simulated} \leq 8$ m) as further away and the targets with small visual angles (i.e., simulated to be at further distances; $D_{simulated} \geq 12$ m) as closer when using the method of adjustment.

I hypothesized that there would be no difference in intercepts and slopes, computed from the stimuli and participants responses, between the experimental methods. The intercepts in participant responses were lower when doing the task using continuous tracking compared to the results obtained from the same participants with the method of adjustment. The slopes were shallower when doing the task using continuous tracking compared to when using the method of adjustment in the Position-to-Size task, but not in the Size-to-Position task. These results did not support my first hypothesis.

My second hypothesis was also not supported as the results did not show evidence of differences in the intercepts between tasks in either experimental method. When compared to the geometrically correct horizon position on screen, the intercepts in participants responses were too low in Position-to-Size task but not in Size-to-Position task when doing the task using method of adjustment. However, they were too low in both tasks when using continuous tracking. Estimating the perceived horizon from the intercepts and the perceived eye height from the slopes, why were participants' perceived horizon and eye height dependent on the experimental method (method of adjustment vs. continuous tracking) when the experimental set up and the scene used were the same?

4.5.3.1. Difference Between the Two Experimental Methods.

The intercepts results aligned with the horizon being lower than the correct horizon by ~ 5.4 cm ($\sim 1.5^\circ$) with continuous tracking and lower by ~ 2.3 cm ($\sim .7^\circ$) with method of adjustment. The difference in corresponding slopes were even larger which aligned with the eye

height being ~ 22.2 cm ($\sim 6.3^\circ$) shorter than the correct eye height with continuous tracking compared to it being shorter by ~ 8.7 cm ($\sim 2.5^\circ$) with method of adjustment. These results suggest that participants were less accurate at using the horizon and their own eye height when doing the tasks using the continuous tracking than when using the method of adjustment. Continuous psychophysics introduced by Bonnen et al. (2015) was originally designed to measure and compare sensitivity to change of a stimulus in different conditions, such as change in target position with different target widths or change in heading direction with difference amount of optic flow (Jörges et al., 2024), using a rapidly changing stimulus. In the present experiment, using this method, I evaluated the accuracy of participants responses comparing them to the correct value. Could the use of the method of continuous tracking have led participants judge the horizon in the scene and their eye height less accurately than the traditional method of adjustment when doing the tasks?

Observer can misjudge their eye level when viewing a scene with an indeterminate horizon (Dixon et al., 2000). Although the hallway scene I used in the present study provided many visual cues that can be used to compute the correct elevation of horizon, the actual horizon was still hidden by the end wall of my finite-length hallway. Using the method of adjustment, participants had as much time as they needed to make the target appear correct for its visual size or its position in the scene. In comparison, with continuous tracking method participants had to continuously adjust target size or position for the target changing every 0.5 seconds which required much quicker responses from them. The bottom of the wall at the end of the hallway in the scene was ~ 5.3 cm ($\sim 1.5^\circ$) below the horizon on the screen which was very similar to the perceived horizon deduced from the continuous tracking responses of ~ 5.4 cm ($\sim 1.5^\circ$) below the correct horizon. Even when using the method of adjustment, participants thought the horizon was

lower by ~ 2.3 cm ($\sim 7^\circ$) which further suggest their responses may have been influenced by the end of the visible corridor. The average lag in participant responses using the mouse wheel was approximately 0.8 seconds which was longer than the typical reaction time of approximately 0.5 seconds (reaction to sound; forced choice; finger press task) found in a past study (Klapp, 1995). Motor reaction time increases with complexity of the task (Christina & Rose, 1985; Klapp, 1995) and the longer reaction time found in the present study may have been due to more cognitive load required to mentally compute target's size/position to adjust it. The NASA Task Load Index score with equal weight for each question, interpreted from the scale in Prabaswari et al. (2019), the task required 'High' mental workload ($M = 63$, $SD = 11.1$) which was higher than the average score found in experiment 1 ($M = 40$, $SD = 10$). The quick responses required in the method of continuous tracking (stimuli changed every 0.5 seconds) may have resulted in participants relying more on what was more visually apparent in the scene, such as the bottom of the wall at the end of the hallway, resulting in biased responses aligning an apparent 'horizon' with it.

Knowing the correct eye height requires the observer to know the correct eye level and the distance to the ground plane. The intercept results suggest that participants misjudged the horizon and could partially be explained by the shorter estimation of their eye height, but not to the extent to the shortened perceived eye height suggested by the slope results. With the horizon perceived to be ~ 5.4 cm ($\sim 1.5^\circ$) lower than the correct horizon and the eye height perceived to be ~ 22.2 cm ($\sim 6.3^\circ$) shorter than the correct eye height with continuous tracking, the difference leaves about 16.8 cm unexplained by misjudged horizon. The horizon was lined up with participants' eye level, and if participants correctly perceived them to be aligned, the difference between the perceived horizon and eye height can only be explained by the misperceived

distance to the ground plane. Similar errors were found with method of adjustment, but they were smaller at about 6.4 cm. People are quite accurate at judging their eye height standing in a real corridor (Mark, 1987). In the present experiment, however, participants were seated during the task, and they may not have been as accurate at judging their eye height seated.

Wu et al. (2008) suggested that space perception becomes less accurate when near-ground-surface is ignored. To keep track of the rapidly changing target with continuous tracking, participants would have had to fixate on the target all the time with no time to look around at other visual cues in the scene. This may have resulted in a distorted perception of space, likely perceived as compressed as has been shown in the past (Roscoe, 1984), thinking the ground was closer to their eyes than it actually was. This line of thought however, cannot explain the difference in the between tasks where the perceived eye heights were significantly different between experimental methods only in the Size-to-Position task and not in the Position-to-Size task (See Figure 4.12B).

4.5.3.2. Difference Between the Tasks.

The intercept and slope results from the Size-to-Position task aligned with the horizon and eye height being perceived correctly when using the method of adjustment, but erroneous (lower horizons and shorter eye heights) when using continuous tracking method. In the Position-to-Size task, however, the results suggest horizons and eye heights were perceived erroneously for both experimental methods. Comparing the time lags and correlations in their responses to the changing stimuli using the method of continuous tracking, I found no difference in their reaction time (time lags) or precision (correlations) between the tasks which suggest that participants did not find one task more difficult than the other. Therefore, the difference in their performance between the tasks must come from a different source.

When looking at a vertical object on the floor, the ‘horizon ratio’, the ratio between the visual height (size) and the distance below the horizon of the object (Sedgwick, 2021) can be used to determine its relative size without estimating its absolute distance. Therefore, if elevation of horizon in the scene was misperceived, the object size judged based on this ‘horizon ratio’ would also be erroneous. Ghasemi et al. (2023) also found that varying eye height affected people’s perceived size: an artificially raised eye height in VR resulting in a larger underestimation of size when sitting down compared to when standing. The present experiment was conducted with participants sitting down which may have amplified the underestimation of target size with larger errors in their perceived horizon using method of continuous tracking compared to method of adjustment. The underestimated size would explain the targets placed closer to the horizon (i.e., further away) in the Size-to-Position task hence resulting in shallower slopes in their responses. The difference in slopes between the tasks here may indicate that misperceiving the horizon impacted the perceived size of the target (not the eye height) in the Size-to-Position task, but not the perceived distance obtained from the target position in the scene in the Position-to-Size task. These results further support the idea that the two tasks rely on different cues.

4.6. General Discussion

The absolute distance of an object can be determined using various sources of distance information as reviewed in Chapter 1. The present study demonstrated that people may be quite accurate at determining the size of objects, leading to them correctly judging their distance when presented in a correctly scaled 2D scene with rich distance cues. To judge distance correctly from the size, however, the size of the object must be correctly perceived first, which may be influenced by various factors. Knowing the correct horizon position in the scene is important in

judging relative sizes of objects on the floor as it is required to compute correct ratio between the visual size of the object and its distance below the horizon (Sedgwick, 2021). This may be even more important when people are sitting down as the impact of a misperceived horizon height can amplify the errors in perceived size when viewing from a sitting down position (Ghasemi et al., 2023). Errors in the perceived horizon seem to reduce when there was enough time to explore the visual cues available, but horizon estimates could still be erroneous even with the correctly scaled scene providing the correct geometry needed to compute the correct horizon height which can lead to misperceived size and hence misperceived distance. When the real horizon is not explicitly visible in a 2D scene, the perceived horizon location can be influenced by the implied vanishing point and the boundary of the visible surface (Ozkan & Braunstein, 2010). A clearly visible horizon in the scene may reduce these problems, but that is usually not the case even in the real world where the horizon is often occluded by other objects such as trees or buildings.

Judging distance using the angular declination from eye level (equivalent to the horizon), as in the Position-to-Size task in the present study, erroneously made the target appear too big indicating that the target was judged to be closer than simulated. Misperceived horizon in a 2D scene may have also contributed to the erroneous perception of absolute distance, but larger errors in the perceived horizon in experiment 2 did not make their performance worse which makes any link between the apparent position of the horizon and perceived absolute distance inconclusive. Features used to improve people's judgement of size and distance in the present study such as large screen, easily discernable target sizes, familiar objects in the scene, and rich ground plane information all seem to improve perception of size and distance, but they were still not enough for them to infer absolute distances correctly from object positions in the 2D scene.

Overall, the results show clear differences between the two tasks (Size-to-Position and Position-to-Size) independent of the methodologies used to assess them. Results diverge in how different sources of error impact them, further supporting the idea that size and distance perception use separate brain mechanisms. However, it is still unclear how much weight each of the contributing factors plays in human distance perception. Further studies are needed to evaluate how each cue influence people's size and distance perception differently.

5.0. General Discussions

Judging the spatial relationship between the objects in the environment and oneself is a necessary process for human navigation. Despite its importance in our daily lives people make systematic errors, generally judging object positions as less eccentric than as shown in the past studies (Kim, 2019; Prablanc et al., 1979; Vercher et al., 1994) and absolute distances as farther for close objects (<4 m) and closer for far objects (>4 m) (Foley, 1980; Z. Li et al., 2011). Furthermore, the egocentric position of these objects change following self-motion, and the process of updating the object positions produces additional updating errors (Gutteling & Medendorp, 2016; Kim, 2019; Klier et al., 2008). Investigating the source of these errors is important for obtaining a better understanding of human spatial perception. The purpose of this dissertation was to investigate the visual cues used in human visual spatial perception and updating of the object positions and their sizes. See Section 1.4 for the brief overview of each experimental chapter, their goals, and hypotheses.

5.1. Misperceived Self-motion Leads to Updating Errors

In Chapter 2, I confirmed that people make updating errors when keeping track of an object following a visually-simulated passive lateral translation, without physical motion cues. The updating errors were in the direction of the movement which indicated underestimated travel distance. The errors increased with longer travel distance (H1 supported) suggesting that people underestimated the extent (or speed) of their self-motion. What might be the source of this underestimation?

5.1.1. *Lack of physical motion cues*

When moving, both visual and physical motion cues provide people with information about their self-motion and integrating these different sensory inputs leads to more accurate

gauging of their self-motion (Bayramova et al., 2021; Hlavacka et al., 1996). Comparing the updating error results from Chapter 2 to the past study conducted using a MOOG motion platform, where people made almost no updating errors following a lateral translation (Kim, 2019), demonstrated that physical motion cues provided by the motion platform resulted in more accurate updating. These results suggest that the lack of physical motion cues in Chapter 2 is one of the reasons for the misperceived travel distance, leading to updating errors. However, when sitting on a motion platform, updating errors were reduced even when people are only visually moved in VR (Kim, 2019), i.e., no physical motion involved, which suggest that lack of physical motion cues may not be the only source of errors found in Chapter 2.

5.1.2. Possibility of Physical Motion

In my Masters thesis (Kim, 2019), I showed that people did not make updating errors when they were passively moved while sitting on the MOOG motion platform whether the motion platform moved (visual and physical cues) or not (visual cues only). In Chapter 2, however, people made updating errors with the identical visual translation, i.e., without physical cues, (H2 supported) which suggests that just sitting on a motion platform may have improved people's updating accuracy. Past study have shown that sitting on a chair that is known to be able to rotate enhances the feeling of rotationalvection compared to sitting on a fixed chair (D'Amour et al., 2021). Therefore, it appears that the mere possibility of physical motion led to more accurate perception of self-motion and reduced updating errors. These results suggest that anticipating physical motion may make people more sensitive to motion cues, even just visual cues, enhancing their perception of self-motion and hence resulting in better updating.

5.2. Visual Cues Influencing Size and Distance Perception

People use various visual distance cues to determine distance and artists, whether on canvases or on computer screens, implement some of these cues to give the viewers a perception of distance on flat 2D surfaces. When judging distance of an object in a 2D scene, people rely on monocular cues such as its visual size and position (i.e., elevation) in the scene. In Chapter 3, I showed that people cannot infer absolute distance of an object correctly from its size or position when viewing it on small computer screens (H3 not supported), but Chapter 4 showed that they can use the size of the object to determine its absolute distance when it is part of a correctly scaled large display (H5 supported). People could, however, successfully segregate objects into different distances according to their size and position in both Chapter 3 and 4, which demonstrates that they understand and use the geometrical relationship between objects' visual sizes and their positions in a 2D scene (detailed in Figure 4.4) to judge distance. Which visual distance cues influenced people's accuracy in judging an object's absolute distance, resulting in different performances in Chapter 3 and 4?

5.2.1. *Familiar Objects in the Scene and Ground Information*

Past studies have shown that familiar objects in the environment can help in judging the size and distance of other objects (e.g., Park et al., 2021). However, having familiar objects in the 2D scene in Chapter 3 only improved people's distance perception when they were judging an object's distance from its size, not from its position (H4 only partially supported). Similarly, in Chapter 4 where the scene contained familiar objects, people judged distance correctly when judging an object's distance from its size, but not from its position. To correctly infer an object's distance from its size, the size must first be correctly perceived. The results from Chapter 3

suggest that familiar objects in the scene improved people's size judgment leading to more accurate distance judgement but did not help in judging the object's position in the scene.

In Chapter 4, the ground in the 2D scene was lined up with a real carpet with the identical pattern, in addition to familiar objects, providing the near-ground-surface (i.e., ground from the observer to the target object) information. The improved accuracy in people's performance in Chapter 4 compared to Chapter 3, where the object was displayed on regular small computer screens; i.e., with no near-ground information (H5 supported), demonstrate that the ground information helped judging absolute distance to an object as previously shown (Wu et al., 2004). Although people were more accurate overall, they still made errors when judging distance from an object's position in the 2D scene.

If familiar objects only help in size judgement, as indicated in Chapter 3 results, the Chapter 4 results suggest that ground information may have helped in judging both the size and position of an object in the 2D scene or only the position. Because both familiar objects and ground information were always available in Chapter 4, it is inconclusive which was the case and further studies are needed to segregate the effect of these different cues.

5.2.2. The Horizon in the Scene and Eye-height

When looking at a scene, an observer uses perspective to judge the size and position of the objects in the environment in relation to each other. The horizon, where the sky meets the ground, is important element for perspective as when visible it aligns with the vanishing point (see Figure 4.5A). The horizon also corresponds to eye level which aligns with the eye-height in the real world and it has been shown that people are quite accurate at judging their own eye height (Mark, 1987). When looking at a 2D scene in a picture, for example, the eye height of an observer may not align with the depicted horizon (See Figure 4.5B) and past literature suggest

that people generally cannot scale their perception of the picture to their eye level correctly when viewing a non-immersive display (Dixon et al., 2000). People use the horizon as the reference to determine the angular declination of an object relative to the eye level (Ooi et al., 2001), i.e., the position of an object in a 2D scene used in Chapter 3 and 4. People also use ‘horizon ratio’, the ratio between the visual height of the objects and its distance below the horizon, to determine relative size between objects on the floor (Sedgwick, 2021). Therefore, gauging the correct horizon is important in judging the size and position of an object in a 2D scene.

In Chapter 4, the horizon and eye height that correspond with the size and position results (represented by the average intercepts and slopes computed by fitting the size and position with a linear model; see Section 4.4.1.5 for details) differed between tasks (i.e., judging distance from object size or from position). There are two possible explanations: 1) observers made different errors when judging the location of horizon and their eye height in the scene, or 2) the misjudged location of horizon and/or eye height influenced observer’s size/position judgement differently. The 2D scene presented was identical for all experimental tasks and the horizon was aligned with the observer’s eye height, therefore it is unlikely that horizon in the scene and the eye height were perceived differently between tasks. What is more likely is that the misperceived horizon influenced size and position judgement differently, resulting in different results in the two tasks. Observers can determine the relative distance to objects and ignore other erroneous information in a picture when the horizon is clearly defined (Gardner et al., 2010). In the scene used in Chapter 4, the horizon was not explicitly visible to the observer as the hallway ended at ~25 m as shown in Figure 4.5A. When horizon is not visible, location of horizon is determined by combining the implied vanishing point and the boundary of the visible surface in the scene (Ozkan & Braunstein, 2010). People may have confused the visible end of floor with the horizon

and used that as reference instead of the true horizon that was slightly above this line and was not visible. The height of the true horizon needed to be deduced from the geometry of the visible features in the room such as the bottom of the walls and tops of the doors. Participants would consequently have scaled the position of an object in the scene incorrectly, especially when they had to make a speeded judgement in the continuous tracking task (Exp 2 in Chapter 4).

5.2.3. Judging Size and Distance may use Different Visual Cues

Past literature suggests that size and distance perception may be processed independently (Haber & Levin, 2001; Kim, 2018; Mon-Williams & Tresilian, 1999; Ozkan & Braunstein, 2010). The visual cues included in the scene (e.g., familiar objects, ground information, and possibly the horizon) in Chapter 3 and 4 influencing the two tasks differently imply that people relied on different cues which may support this idea of independent processes. For example, people can judge an object's size using familiar objects in the scene as references without determining its distance. Similarly, people can judge an object's distance from the near-ground-surface information, such as texture, without determining its size. However, some cues (e.g., ground information) seem to influence both size and distance judgement as suggested by the Chapter 4 results. How much each cue is weighted for different tasks need to be further investigated.

5.3. Immersive and Familiar Test Setting Improves Spatial Perception

In Chapter 2, I compared updating errors people made in VR while sitting on a fixed chair in a lab to the errors made while sitting on a moving platform in a room identical to the VR environment (i.e., the MOOG room) in the past study (Kim, 2019). The comparison revealed that this difference in test setting influenced people's updating accuracy where people make smaller updating errors when tested in the MOOG room. I interpreted that this result was from more

accurate self-motion perception due to enhancedvection from sitting on a chair that can move. However, other spatial judgements may have also improved people's updating such as distance perception.

Experience of interacting with (Interrante et al., 2006) or even just seeing (Kelly et al., 2018) a real-world environment can improve people's judgement of distance in a matching VR environment. In a past study (Kim, 2019) conducted in the MOOG room, people had to walk through the real room and climb up onto the moving platform in order to participate in the experiment. This process, when compared to sitting in a regular lab room in Chapter 2, could have resulted in them being more familiar with the environment with better judgment of the room's spatial layout. In Chapter 4, when viewing an object in a scene while sitting in a semi-immersive setting with a large screen, people were more accurate at judging distance from a 2D scene than they were in Chapter 3 where they viewed the scene on a computer screen. Although there were more visual distance cues available in the experiments described in Chapter 4 (i.e., the ground information) than in those described in Chapter 3, the effect of viewing the scene on computer screens cannot be ignored: past studies have shown, for example, that a real-world scene is perceived as smaller when viewed on a screen (Roscoe, 1984). These results suggest that more immersive and familiar test environment result in a more accurate perception of self-motion and distance, leading to more accurate spatial perception.

5.4. Limitations and Future Directions

Although the results reported in Chapter 2 suggest that updating errors may be due to underestimated self-motion due to insufficient motion cues, those experiments were limited to using passive lateral motion. As introduced in Figure 1.5, people can rotate or translate in other directions which all require updating when keeping track of their environment. Various past

studies have already shown that people make different errors when updating following different self-motions (Gutteling & Medendorp, 2016; Klier et al., 2008; Riecke et al., 2005) for which they utilize very different methods. To fully understand the human spatial updating process and the factors that contribute to the errors people make, a series of studies with varying types of movement using the identical method would be beneficial. For example, spatial updating experiments conducted for fore/aft, up/down, and lateral translation using the same VR environment would reveal whether the lack of physical motion cues influence people's updating different, or the same, amount for different movement directions.

While the results from the experiments explained in this dissertation provide evidence of people making errors in judging their self-motion (in Chapter 2), horizon and eye height (in Chapter 3 and 4), I did not measure these parameters directly. Rather, they were interpreted from the misperceived object positions observed in the experiments. For example, the horizons and eye heights were estimated by fitting the responses into a linear model (See Section 4.4.1.5). Therefore, it is difficult to make a conclusive verdict as to whether people actually misjudged them. Future studies measuring the perceived horizon and eye height judged by people when they look at a 2D image of a scene would help in establishing clearer conclusion of these findings, i.e., whether the errors found in Chapter 4 were from misperceived horizon/eye height or from another source.

5.5. General Conclusions

Visual spatial perception is a complex process where people rely on different cues to perceive and keep track of the surrounding environment for various purposes. This spatial perception is especially important for people's navigation because it allows them to effectively

avoid or interact with the objects in the environment. Despite the importance, people make errors where they judge the position of surrounding objects less eccentric and closer in distance.

During navigation, people move around the environment and as they move, the positions of objects need to be updated accordingly if the person is going to be able interact with them quickly using their updated position in their memory. In addition to the initially misperceived objects positions, people also make updating errors when they move, which differ depending on the type of movement and direction (i.e., rotation or translation – fore/aft, up/down, or lateral). The updating errors following a lateral translation are partially produced by misperceived self-motion and can be reduced with more motion cues (e.g., physical motion cues), or knowing that physical motion is possible (e.g., sitting on a moving platform).

Misperceived absolute distance can also produce updating errors as the magnitude of the change in an object's position depends on its distance following a lateral translation. People's distance judgement improves with more visual distance cues available in the scene. When binocular cues are not available, e.g., when viewing a 2D image of a scene on a surface, people can use objects' visual sizes and positions in the scene (i.e., distance below the horizon; see Figure 3.1) to determine relative distances of the objects. However, people still sometimes make errors in judging absolute distances depending on the visual cues available. For example, familiar objects in the scene help in judging the size of the objects and can lead to correctly judging their absolute distances, but it does not help in judging the correct position in the scene resulting in errors even when the person is looking at the same scene when making those judgements. Near-ground-surface (ground from the observer's feet to the object) information, the person's familiarity and the immersiveness of the environment all improve people's distance

perception, but in the experiments reported in this thesis, they were not effective enough for people to correctly judge absolute distances.

Overall, different visual cues seem to affect people's spatial perception differently, some more effective than the others in certain situations (or tasks). More cues available improve people's judgment but people are also influenced by whether what they experience is real or not. Further studies looking into the weight of each visual cue playing a role in people's spatial perception would be beneficial to understand people's spatial perception ability more clearly.

6.0. References

- Abe, K. (2013). *The Effect of Physical Load in Cognitive Process of Estimation*. 35(35).
- Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, 9(1), 10–10. <https://doi.org/10.1167/9.1.10>
- Angelaki, D. E., Shaikh, A. G., Green, A. M., & Dickman, J. D. (2004). Neurons compute internal models of the physical laws of motion. *Nature*, 430(6999), 560–564. <https://doi.org/10.1038/nature02754>
- Bayramova, R., Valori, I., McKenna-Plumley, P. E., Callegher, C. Z., & Farroni, T. (2021). The role of vision and proprioception in self-motion encoding: An immersive virtual reality study. *Attention, Perception, & Psychophysics*, 83, 2865–2878.
- Bertamini, M., Yang, T. L., & Proffitt, D. R. (1998). Relative size perception at a distance is best at eye level. *Perception & Psychophysics*, 60(4), 673–682. <https://doi.org/10.3758/BF03206054>
- Bertenthal, B. I., & Bai, D. L. (1989). Infants' sensitivity to optical flow for controlling posture. *Developmental Psychology*, 25(6), 936.
- Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision*, 15(3), 14–14.
- Boring, E. G. (1964). Size-Constancy in a Picture. *The American Journal of Psychology*, 77(3), 494. <https://doi.org/10.2307/1421027>
- Brenner, E., & Smeets, J. B. (2018). Depth perception. In *Stevens' handbook of experimental psychology and cognitive neuroscience: Sensation, perception, and attention* (pp. 385–414). Wiley.
- Britten, K. H. (2008). Mechanisms of self-motion perception. *Annu. Rev. Neurosci.*, 31(1), 389–410.
- Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*, 10(12), 551–557. <https://doi.org/10.1016/j.tics.2006.10.005>
- Campos, J. L., Butler, J. S., & Bühlhoff, H. H. (2014). Contributions of visual and proprioceptive information to travelled distance estimation during changing sensory congruencies. *Experimental Brain Research*, 232, 3277–3289.
- Carlson, V., & Tassone, E. (1971). *Familiar versus unfamiliar size: A theoretical derivation and test*.
- Chen, J., McManus, M., Valsecchi, M., Harris, L. R., & Gegenfurtner, K. R. (2019). Steady-state visually evoked potentials reveal partial size constancy in early visual cortex. *Journal of Vision*, 19(6), 8. <https://doi.org/10.1167/19.6.8>
- Cheng, Z., & Gu, Y. (2018). Vestibular system and self-motion. *Frontiers in Cellular Neuroscience*, 12, 456.
- Christina, R. W., & Rose, D. J. (1985). Premotor and motor reaction time as a function of response complexity. *Research Quarterly for Exercise and Sport*, 56(4), 306–315.
- Crane, B. T. (2014). The influence of head and body tilt on human fore-aft translation perception. *Experimental Brain Research*, 232(12), 3897–3905. <https://doi.org/10.1007/s00221-014-4060-4>

- Cruikshank, R. M. (1941). The Development of Visual Size Constancy in Early Infancy. *The Pedagogical Seminary and Journal of Genetic Psychology*, *58*(2), 327–351. <https://doi.org/10.1080/08856559.1941.10534574>
- Cullen, K. E. (2011). The neural encoding of self-motion. *Current Opinion in Neurobiology*, *21*(4), 587–595.
- Cullen, K. E. (2019). Vestibular processing during natural self-motion: Implications for perception and action. *Nature Reviews Neuroscience*, *20*(6), 346–363.
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of space and motion* (pp. 69–117). Elsevier.
- D’Amour, S., Harris, L. R., Berti, S., & Keshavarz, B. (2021). The role of cognitive factors and personality traits in the perception of illusory self-motion (vection). *Attention, Perception, & Psychophysics*, *83*(4), 1804–1817. <https://doi.org/10.3758/s13414-020-02228-3>
- Day, B. L., & Fitzpatrick, R. C. (2005). The vestibular system. *Current Biology*, *15*(15), R583–R586.
- De Moraes, C. G. (2013). Anatomy of the visual pathways. *Journal of Glaucoma*, *22*, S2–S7.
- DeAngelis, G. C., & Angelaki, D. E. (2012). *Visual–vestibular integration for self-motion perception*.
- Dixon, M. W., Wraga, M., Proffitt, D. R., & Williams, G. C. (2000). Eye height scaling of absolute size in immersive and nonimmersive displays. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 582–593. <https://doi.org/10.1037/0096-1523.26.2.582>
- Ebner, T. J., & Pasalar, S. (2008). Cerebellum predicts the future motor state. *The Cerebellum*, *7*, 583–588.
- Erkelens, C. J., & van Ee, R. (2002). The role of the cyclopean eye in vision: Sometimes inappropriate, always irrelevant. *Vision Research*, *42*(9), 1157–1163. [https://doi.org/10.1016/S0042-6989\(01\)00280-2](https://doi.org/10.1016/S0042-6989(01)00280-2)
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. <https://doi.org/10.1038/415429a>
- Fajen, B. R., & Matthis, J. S. (2013). Visual and non-visual contributions to the perception of object motion during self-motion. *PLoS One*, *8*(2), e55446.
- Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2010). Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory: Mechanisms of visual-vestibular cue integration. *European Journal of Neuroscience*, *31*(10), 1721–1729. <https://doi.org/10.1111/j.1460-9568.2010.07207.x>
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, *39*, e229. <https://doi.org/10.1017/S0140525X15000965>
- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*(4), 507–521.
- Fitzpatrick, V., Pasnak, R., & Tyer, Z. E. (1982). The Effect of Familiar Size at Familiar Distances. *Perception*, *11*(1), 85–91. <https://doi.org/10.1068/p110085>

- Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, 87(5), 411–434. <https://doi.org/10.1037/0033-295X.87.5.411>
- Gardner, J. S., Austerweil, J. L., & Palmer, S. E. (2010). Vertical position as a cue to pictorial depth: Height in the picture plane versus distance to the horizon. *Attention, Perception, & Psychophysics*, 72(2), 445–453. <https://doi.org/10.3758/APP.72.2.445>
- Ghasemi, F., Harris, L. R., & Jörges, B. (2023). Simulated eye height impacts size perception differently depending on real-world posture. *Scientific Reports*, 13(1), 20075.
- Gibson, J. J. (1950). *The perception of the visual world*.
- Gibson, J. J. (1954). A theory of pictorial perception. *Educational Technology Research and Development*, 2(1), 3–23. <https://doi.org/10.1007/BF02713318>
- Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological Review*, 58(6), 460–482. <https://doi.org/10.1037/h0061505>
- Gogel, W. C., & Mertens, H. W. (1967). Perceived Size and Distance of Familiar Objects. *Perceptual and Motor Skills*, 25(1), 213–225. <https://doi.org/10.2466/pms.1967.25.1.213>
- Gogel, W. C., Wist, E. R., & Harker, G. S. (1963). A Test of the Invariance of the Ratio of Perceived Size to Perceived Distance. *The American Journal of Psychology*, 76(4), 537. <https://doi.org/10.2307/1419704>
- Granrud, C. E. (2009). Development of size constancy in children: A test of the metacognitive theory. *Attention, Perception & Psychophysics*, 71(3), 644–654. <https://doi.org/10.3758/APP.71.3.644>
- Greenlee, M. W., Frank, S. M., Kaliuzhna, M., Blanke, O., Bremmer, F., Churan, J., Cuturi, L. F., MacNeilage, P. R., & Smith, A. T. (2016). Multisensory Integration in Self Motion Perception. *Multisensory Research*, 29(6–7), 525–556. <https://doi.org/10.1163/22134808-00002527>
- Gutteling, T. P., & Medendorp, W. P. (2016). Role of Alpha-Band Oscillations in Spatial Updating across Whole Body Motion. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00671>
- Gutteling, T. P., Selen, L. P. J., & Medendorp, W. P. (2015). Parallax-sensitive remapping of visual space in occipito-parietal alpha-band activity during whole-body motion. *Journal of Neurophysiology*, 113, 1574–1584. <https://doi.org/10.1152/jn.00477.2014>
- 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>
- Harris, L. R., Jenkin, M. R., Zikovitz, D., Redlick, F., Jaekl, P., Jasiobedzka, U. T., Jenkin, H. L., & Allison, R. S. (2002). Simulating Self-Motion I: Cues for the Perception of Motion. *Virtual Reality*, 6(2), 75–85. <https://doi.org/10.1007/s100550200008>
- Harris, L. R., Jenkin, M., & Zikovitz, D. C. (2000). Visual and non-visual cues in the perception of linear self motion. *Experimental Brain Research*, 135(1), 12–21. <https://doi.org/10.1007/s002210000504>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)

- Harwerth, R. S., Moeller, M. C., & Wensveen, J. M. (1998). Effects of cue context on the perception of depth from combined disparity and perspective cues. *Optometry and Vision Science*, 75(6), 434–444.
- Heiser, L. M., & Colby, C. L. (2006). Spatial updating in area LIP is independent of saccade direction. *Journal of Neurophysiology*, 95(5), 2751–2767.
- Higashiyama, A. (1983). *A Variety of size and Distance Judgments under Monocular Observation: Instructions and Individual Difference*. 大阪府立大学人間科学研究会.
<https://doi.org/10.24729/00004796>
- Higashiyama, A. (1984). The effects of familiar size on judgments of size and distance: An interaction of viewing attitude with spatial cues. *Perception & Psychophysics*, 35(4), 305–312.
<https://doi.org/10.3758/BF03206333>
- Hlavacka, F., Mergner, Th., & Bolha, B. (1996). Human self-motion perception during translatory vestibular and proprioceptive stimulation. *Neuroscience Letters*, 210(2), 83–86.
[https://doi.org/10.1016/0304-3940\(96\)12667-7](https://doi.org/10.1016/0304-3940(96)12667-7)
- Hochberg, C. B., & Hochberg, J. E. (1952). Familiar Size and the Perception of Depth. *The Journal of Psychology*, 34(1), 107–114. <https://doi.org/10.1080/00223980.1952.9916110>
- Horie, A., Nagano, H., Konyo, M., & Tadokoro, S. (2018). *Buttock skin stretch: Inducing shear force perception and acceleration illusion on self-motion perception*. 135–147.
- Hornsey, R. L., & Hibbard, P. B. (2021). Contributions of pictorial and binocular cues to the perception of distance in virtual reality. *Virtual Reality*, 25(4), 1087–1103.
<https://doi.org/10.1007/s10055-021-00500-x>
- Howard, I. P., & Howard, A. (1994). Vection: The Contributions of Absolute and Relative Visual Motion. *Perception*, 23(7), 745–751. <https://doi.org/10.1068/p230745>
- Hubel, D. H., & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculate-cortical fibers in the macaque monkey. *Journal of Comparative Neurology*, 146(4), 421–450.
- Interrante, V., Ries, B., & Anderson, L. (2006). Distance Perception in Immersive Virtual Environments, Revisited. *IEEE Virtual Reality Conference (VR 2006)*, 3–10.
<https://doi.org/10.1109/VR.2006.52>
- Irwin, R. J. (1969). Emmert's Law as a Consequence of Size Constancy. *Perceptual and Motor Skills*, 28(1), 69–70. <https://doi.org/10.2466/pms.1969.28.1.69>
- Ittelson, W. H. (1951). Size as a cue to distance: Static localization. *The American Journal of Psychology*, 64(1), 54–67.
- Jitsumori, M. (2010). Do animals recognize pictures as representations of 3D objects? *Comparative Cognition & Behavior Reviews*, 5.
- Jörges, B., Bansal, A., & Harris, L. R. (2024). Precision and temporal dynamics in heading perception assessed by continuous psychophysics. *PloS One*, 19(10), e0311992.
- Jung, E., Takahashi, K., Watanabe, K., de la Rosa, S., Butz, M. V., Bühlhoff, H. H., & Meilinger, T. (2016). The Influence of Human Body Orientation on Distance Judgments. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00217>
- Kappers, A. M., & Koenderink, J. J. (1999). Haptic perception of spatial relations. *Perception*, 28(6), 781–795.

- Kelly, J. W., Cherep, L. A., Klesel, B., Siegel, Z. D., & George, S. (2018). Comparison of Two Methods for Improving Distance Perception in Virtual Reality. *ACM Transactions on Applied Perception, 15*(2), 1–11. <https://doi.org/10.1145/3165285>
- Kim, J., Chung, C. Y. L., Nakamura, S., Palmisano, S., & Khoo, S. K. (2015). The Oculus Rift: A cost-effective tool for studying visual-vestibular interactions in self-motion perception. *Frontiers in Psychology, 6*. <https://doi.org/10.3389/fpsyg.2015.00248>
- Kim, J. J.-J., McManus, M. E., & Harris, L. R. (2022). Body Orientation Affects the Perceived Size of Objects. *Perception, 51*(1), 25–36. <https://doi.org/10.1177/03010066211065673>
- Kim, J.-J. (2019). *Updating During Lateral Movement Using Visual and Non-Visual Motion Cues*.
- Kim, N.-G. (2018). Independence of Size and Distance in Binocular Vision. *Frontiers in Psychology, 9*, 988. <https://doi.org/10.3389/fpsyg.2018.00988>
- Klapp, S. T. (1995). Motor response programming during simple choice reaction time: The role of practice. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 1015.
- Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science, 9*(4), 293–298. <https://doi.org/10.1111/1467-9280.00058>
- Klier, E. M., & Angelaki, D. E. (2008). Spatial updating and the maintenance of visual constancy. *Neuroscience, 156*(4), 801–818. <https://doi.org/10.1016/j.neuroscience.2008.07.079>
- Klier, E. M., Hess, B. J. M., & Angelaki, D. E. (2008). Human Visuospatial Updating After Passive Translations in Three-Dimensional Space. *Journal of Neurophysiology, 99*(4), 1799–1809. <https://doi.org/10.1152/jn.01091.2007>
- Kunnapas, T. (1968). Distance perception as a function of available visual cues. *Journal of Experimental Psychology, 77*(4), 523.
- Lappe, M., Bremmer, F., & van den Berg, A. V. (1999). Perception of self-motion from visual flow. *Trends in Cognitive Sciences, 3*(9), 329–336. [https://doi.org/10.1016/S1364-6613\(99\)01364-9](https://doi.org/10.1016/S1364-6613(99)01364-9)
- Lewald, J., & Ehrenstein, W. H. (2000). Visual and proprioceptive shifts in perceived egocentric direction induced by eye-position. *Vision Research, 40*(5), 539–547.
- Leyrer, M., Linkenauger, S. A., Bühlhoff, H. H., Kloos, U., & Mohler, B. (2011). The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization - APGV '11*, 67. <https://doi.org/10.1145/2077451.2077464>
- Li, N., & Angelaki, D. E. (2005). Updating visual space during motion in depth. *Neuron, 48*(1), 149–158. <https://doi.org/10.1016/j.neuron.2005.08.021>
- Li, Z., Phillips, J., & Durgin, F. H. (2011). The underestimation of egocentric distance: Evidence from frontal matching tasks. *Attention, Perception, & Psychophysics, 73*(7), 2205–2217. <https://doi.org/10.3758/s13414-011-0170-2>
- MacNeilage, P. R., Banks, M. S., DeAngelis, G. C., & Angelaki, D. E. (2010). Vestibular Heading Discrimination and Sensitivity to Linear Acceleration in Head and World Coordinates.

- Journal of Neuroscience*, 30(27), 9084–9094. <https://doi.org/10.1523/JNEUROSCI.1304-10.2010>
- Maier, A., Cox, M., Westerberg, J., & Dougherty, K. (2022). Binocular integration in the primate primary visual cortex. *Annual Review of Vision Science*, 8(1), 345–360.
- Maltz, M. V., Stubbs, K. M., Quinlan, D. J., Rzepka, A. M., Martin, J. R., & Culham, J. C. (2021). Familiar size affects the perceived size and distance of real objects even with binocular vision. *Journal of Vision*, 21(10), 21. <https://doi.org/10.1167/jov.21.10.21>
- Mark, L. S. (1987). Eyeheight-scaled information about affordances: A study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 361.
- Mather, G., & Smith, D. R. (2000). Depth cue integration: Stereopsis and image blur. *Vision Research*, 40(25), 3501–3506.
- McKenzie, B. E., Tootell, H. E., & Day, R. H. (1980). Development of visual size constancy during the 1st year of human infancy. *Developmental Psychology*, 16(3), 163–174. <https://doi.org/10.1037/0012-1649.16.3.163>
- McManus, M., & Harris, L. R. (2023). Enhancement of visual cues to self-motion during a visual/vestibular conflict. *PLOS ONE*, 18(3), e0282975. <https://doi.org/10.1371/journal.pone.0282975>
- Medendorp, W. P. (2011). Spatial constancy mechanisms in motor control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1564), 476–491. <https://doi.org/10.1098/rstb.2010.0089>
- Meehan, J. W., & Triggs, T. J. (1992). Apparent Size and Distance in an Imaging Display. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(3), 303–311. <https://doi.org/10.1177/001872089203400305>
- Merriam, E. P., Genovese, C. R., & Colby, C. L. (2003). Spatial Updating in Human Parietal Cortex. *Neuron*, 39(2), 361–373. [https://doi.org/10.1016/S0896-6273\(03\)00393-3](https://doi.org/10.1016/S0896-6273(03)00393-3)
- Mon-Williams, M., & Tresilian, J. R. (1999). The size-distance paradox is a cognitive phenomenon. *Experimental Brain Research*, 126, 578–582.
- Morgan, C. L. (1978). Constancy of egocentric visual direction. *Perception & Psychophysics*, 23(1), 61–68.
- Murata, K., Seno, T., Ozawa, Y., & Ichihara, S. (2014). Self-motion perception induced by cutaneous sensation caused by constant wind. *Psychology*, 5(15), 1777.
- Ono, H., Mapp, A. P., & Howard, I. P. (2002). The cyclopean eye in vision: The new and old data continue to hit you right between the eyes. *Vision Research*, 42(10), 1307–1324.
- Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature*, 414(6860), 197–200. <https://doi.org/10.1038/35102562>
- Ozkan, K., & Braunstein, M. L. (2010). Background surface and horizon effects in the perception of relative size and distance. *Visual Cognition*, 18(2), 229–254.
- Palmisano, S., Gillam, B., Govan, D. G., Allison, R. S., & Harris, J. M. (2010). Stereoscopic perception of real depths at large distances. *Journal of Vision*, 10(6), 19–19. <https://doi.org/10.1167/10.6.19>
- Palmisano, S., & Riecke, B. E. (2018). The search for instantaneous vection: An oscillating visual prime reduces vection onset latency. *PloS One*, 13(5), e0195886.

- Park, H., Faghihi, N., Dixit, M., Vaid, J., & McNamara, A. (2021). Judgments of Object Size and Distance across Different Virtual Reality Environments: A Preliminary Study. *Applied Sciences*, *11*(23), 11510. <https://doi.org/10.3390/app112311510>
- Pearson, K. (1976). The control of walking. *Scientific American*, *235*(6), 72–87.
- Peer, A. (2017). On exploring the mitigation of distance misperception in virtual reality. *2017 IEEE Virtual Reality (VR)*, 415–416. <https://doi.org/10.1109/VR.2017.7892353>
- Polat, L., Harpaz, T., & Zaidel, A. (2024). Rats rely on airflow cues for self-motion perception. *Current Biology*, *34*(18), 4248–4260.
- Prabaswari, A. D., Basumerda, C., & Utomo, B. W. (2019). *The mental workload analysis of staff in study program of private educational organization*. *528*(1), 012018.
- Prablanc, C., Echallier, J. F., Komilis, E., & Jeannerod, M. (1979). Optimal response of eye and hand motor systems in pointing at a visual target: I. Spatio-temporal characteristics of eye and hand movements and their relationships when varying the amount of visual information. *Biological Cybernetics*, *35*(2), 113–124. <https://doi.org/10.1007/BF00337436>
- Predebon, J. (1992). The role of instructions and familiar size in absolute judgments of size and distance. *Perception & Psychophysics*, *51*(4), 344–354. <https://doi.org/10.3758/BF03211628>
- Predebon, J. (1993). The Familiar-Size Cue to Distance and Stereoscopic Depth Perception. *Perception*, *22*(8), 985–995. <https://doi.org/10.1068/p220985>
- Probst, T., Brandt, T., & Degner, D. (1986). Object-motion detection affected by concurrent self-motion perception: Psychophysics of a new phenomenon. *Behavioural Brain Research*, *22*(1), 1–11.
- Probst, T., Straube, A., & Bles, W. (1985). Differential effects of ambivalent visual-vestibular-somatosensory stimulation on the perception of self-motion. *Behavioural Brain Research*, *16*(1), 71–79.
- Prsa, M., Gale, S., & Blanke, O. (2012). Self-motion leads to mandatory cue fusion across sensory modalities. *Journal of Neurophysiology*, *108*(8), 2282–2291.
- Reuten, A., Smeets, J., Martens, M., & Bos, J. (2022). Self-motion perception without sensory motion. *Experimental Brain Research*, *240*(10), 2677–2685.
- Riecke, B. E., Cunningham, D. W., & Bühlhoff, H. H. (2007). Spatial updating in virtual reality: The sufficiency of visual information. *Psychological Research*, *71*(3), 298–313. <https://doi.org/10.1007/s00426-006-0085-z>
- Riecke, B. E., Feuereissen, D., Rieser, J. J., & McNamara, T. P. (2012). *Self-motion illusions (vection) in VR—Are they good for anything?* 35–38.
- Riecke, B. E., Heyde, M. V. D., & Bühlhoff, H. H. (2005). Visual cues can be sufficient for triggering automatic, reflexlike spatial updating. *ACM Transactions on Applied Perception*, *2*(3), 183–215. <https://doi.org/10.1145/1077399.1077401>
- Riecke, B. E., Våljamäe, A., & Schulte-Pelkum, J. (2009). Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception (TAP)*, *6*(2), 1–27.
- Rogers, B. (2021). Optic Flow: Perceiving and Acting in a 3-D World. *I-Perception*, *12*(1), 1–25. <https://doi.org/10.1177/2041669520987257>

- Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*(2), 125–134.
- Rogers, S. (1996). The horizon-ratio relation as information for relative size in pictures. *Perception & Psychophysics*, *58*(1), 142–152. <https://doi.org/10.3758/BF03205483>
- Roscoe, S. N. (1984). Judgments of Size and Distance with Imaging Displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(6), 617–629. <https://doi.org/10.1177/001872088402600601>
- Saunders, J. A. (2014). Reliability and relative weighting of visual and nonvisual information for perceiving direction of self-motion during walking. *Journal of Vision*, *14*(3), 24–24.
- Schor, C. M., & Flom, M. C. (1969). The relative value of stereopsis as a function of viewing distance. *Optometry and Vision Science*, *46*(11), 805–809.
- Sedgwick, H. (2005). Visual space perception. *Blackwell Handbook of Sensation and Perception*, 128–167.
- Sedgwick, H. A. (2021). J. J. Gibson’s “Ground Theory of Space Perception.” *I-Perception*, *12*(3), 204166952110211. <https://doi.org/10.1177/20416695211021111>
- Seno, T., Hasuo, E., Ito, H., & Nakajima, Y. (2012). Perceptually plausible sounds facilitate visually induced self-motion perception (vection). *Perception*, *41*(5), 577–593.
- Silk, T. J., Bellgrove, M. A., Wrafter, P., Mattingley, J. B., & Cunnington, R. (2010). Spatial working memory and spatial attention rely on common neural processes in the intraparietal sulcus. *Neuroimage*, *53*(2), 718–724.
- Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth: Newborn infants’ responses to retinal and real size. *Journal of Experimental Child Psychology*, *49*(2), 314–322. [https://doi.org/10.1016/0022-0965\(90\)90061-C](https://doi.org/10.1016/0022-0965(90)90061-C)
- Sousa, R., Brenner, E., & Smeets, J. B. (2011). Judging an unfamiliar object’s distance from its retinal image size. *Journal of Vision*, *11*(9), 10–10.
- Sperandio, I., Sperandio, I., & Chouinard, P. A. (2015). The Mechanisms of Size Constancy. *Multisensory Research*, *28*(3–4), 253–283. <https://doi.org/10.1163/22134808-00002483>
- Steinbach, M. J. (1987). Proprioceptive knowledge of eye position. *Vision Research*, *27*(10), 1737–1744.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181. <https://doi.org/10.1037/h0046162>
- Takeda, T., Hashimoto, K., Hiruma, N., & Fukui, Y. (1999). Characteristics of accommodation toward apparent depth. *Vision Research*, *39*(12), 2087–2097.
- ter Horst, A. C., Koppen, M., Selen, L. P. J., & Medendorp, W. P. (2015). Reliability-Based Weighting of Visual and Vestibular Cues in Displacement Estimation. *PLOS ONE*, *10*(12), e0145015. <https://doi.org/10.1371/journal.pone.0145015>
- Tresilian, J. R., Mon-Williams, M., & Kelly, B. M. (1999). Increasing confidence in vergence as a cue to distance. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1414), 39–44.
- Väljamäe, A., Larsson, P., Västfjäll, D., & Kleiner, M. (2008). Sound representing self-motion in virtual environments enhances linear vection. *Presence: Teleoperators and Virtual Environments*, *17*(1), 43–56.

- Vercher, J. L., Magenes, G., Prablanc, C., & Gauthier, G. M. (1994). Eye-head-hand coordination in pointing at visual targets: Spatial and temporal analysis. *Experimental Brain Research*, *99*(3). <https://doi.org/10.1007/BF00228987>
- Viguiier, A., Clément, G., & Trotter, Y. (2001). Distance Perception within near Visual Space. *Perception*, *30*(1), 115–124. <https://doi.org/10.1068/p3119>
- Voss, P. (2016). Auditory Spatial Perception without Vision. *Frontiers in Psychology*, *07*. <https://doi.org/10.3389/fpsyg.2016.01960>
- Wang, R. F. (2017). Spatial updating and common misinterpretations of spatial reference frames. *Spatial Cognition & Computation*, *17*(3), 222–249. <https://doi.org/10.1080/13875868.2017.1304394>
- Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*(3), 215–250. [https://doi.org/10.1016/S0010-0277\(00\)00105-0](https://doi.org/10.1016/S0010-0277(00)00105-0)
- Warren Jr, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature*, *336*(6195), 162–163.
- Wheatstone, C. (1838). XVIII. Contributions to the physiology of vision. —Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *128*, 371–394. <https://doi.org/10.1098/rstl.1838.0019>
- White III, R. L., & Snyder, L. H. (2007). Spatial constancy and the brain: Insights from neural networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1479), 375–382.
- Witt, J. K., Proffitt, D. R., & Epstein, W. (2004). Perceiving Distance: A Role of Effort and Intent. *Perception*, *33*(5), 577–590. <https://doi.org/10.1068/p5090>
- Wolbers, T., Hegarty, M., Büchel, C., & Loomis, J. M. (2008). Spatial updating: How the brain keeps track of changing object locations during observer motion. *Nature Neuroscience*, *11*(10), 1223–1230. <https://doi.org/10.1038/nn.2189>
- Woods, A. J., Philbeck, J. W., & Danoff, J. V. (2009). The various perceptions of distance: An alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1104.
- Wu, B., Ooi, T. L., & He, Z. J. (2004). Perceiving distance accurately by a directional process of integrating ground information. *Nature*, *428*(6978), 73–77. <https://doi.org/10.1038/nature02350>
- Wu, J., He, Z. J., & Ooi, T. L. (2008). Perceived relative distance on the ground affected by the selection of depth information. *Perception & Psychophysics*, *70*, 707–713.
- Yoo, S.-A., Lee, S., & Joo, S. J. (2023). Monocular cues are superior to binocular cues for size perception when they are in conflict in virtual reality. *Cortex*, *166*, 80–90. <https://doi.org/10.1016/j.cortex.2023.05.010>
- Zhou, L., & Gu, Y. (2023). Cortical mechanisms of multisensory linear self-motion perception. *Neuroscience Bulletin*, *39*(1), 125–137.

Appendix A

Table A.1: Ch 3 - Exp 1 – Repeated Measures ANOVA – comparing target sizes

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Simulated Distance	1.796	^a 4	^a 0.449	^a 130.388	^a < .001	^a 0.124
Residuals	1.116	324	0.003			
Order (initial vs. final)	2.997	1	2.997	33.034	< .001	0.207
Residuals	7.348	81	0.091			
Intended Distance * Order	0.096	^a 4	^a 0.024	^a 6.958	^a < .001	^a 0.007
Residuals	1.116	324	0.003			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table A.2: Ch 3 - Exp 1 – Repeated Measures ANOVA – comparing target positions

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Simulated Distance	68699.343	^a 4	^a 17174.836	^a 143.443	^a < .001	^a 0.639
Residuals	38793.319	324	119.732			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table A.3: Ch 3 - Exp 2 – Repeated Measures ANOVA – comparing target positions

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Simulated Distance	149698.623	^a 4	^a 37424.656	^a 385.818	^a < .001	^a 0.664
Residuals	19788.166	204	97.001			
Order (initial vs. final)	2279.815	1	2279.815	6.582	0.013	0.010
Residuals	17665.312	51	346.379			
Intended Distance * Order	17324.548	^a 4	^a 4331.137	^a 46.905	^a < .001	^a 0.077
Residuals	18837.248	204	92.339			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table A.4: Ch 3 - Exp 2 – Repeated Measures ANOVA – comparing target sizes

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Intended Distance	2.390	^a 4	^a 0.597	^a 68.352	^a < .001	^a 0.573
Residuals	1.783	204	0.009			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table A.5: Ch 3 - Exp 3: Group 1 – Repeated Measures ANOVA – comparing target positions

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Order (initial vs. final)	9757.387	1	9757.387	29.855	< .001	0.054
Residuals	13399.824	41	326.825			
Simulated Distance	118578.018	^a 4	^a 29644.505	^a 422.301	^a < .001	^a 0.658
Residuals	11512.402	164	70.198			
Intended Distance * Order	16233.517	^a 4	^a 4058.379	^a 61.639	^a < .001	^a 0.090
Residuals	11512.402	164	70.198			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table A.6: Ch 3 - Exp 3: Group 1 – Repeated Measures ANOVA – comparing target sizes

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Simulated Distance	0.976 ^a	4 ^a	0.244 ^a	40.487 ^a	< .001 ^a	0.497
Residuals	0.988	164	0.006			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table A.7: Ch 3 - Exp 3: Group 2 – Repeated Measures ANOVA – comparing target sizes

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Order (initial vs. final)	2.018	1	2.018	12.123	0.001	0.193
Residuals	6.826	41	0.166			
Simulated Distance	0.696	^a 4	^a 0.174	^a 64.329	^a < .001	^a 0.067
Residuals	0.444	164	0.003			
Simulated Distance * Order	0.010	^a 4	^a 0.002	^a 0.916	^a 0.456	^a 9.502e -4
Residuals	0.444	164	0.003			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Table A.8: Ch 3 - Exp 3: Group 2 – Repeated Measures ANOVA – comparing target positions

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Intended Distance	46643.277	^a 4	^a 11660.819	^a 77.179	^a < .001	^a 0.653
Residuals	24778.311	164	151.087			

Note. Type III Sum of Squares

^a Mauchly's test of sphericity indicates that the assumption of sphericity is violated ($p < .05$).

Appendix B

On each scale, write the number that best reflects your experience with the task.

Question 1

How much mental load was required for understanding and processing information (e.g. thinking, deciding, calculating, remembering, looking, searching ...)? Was the task easy or demanding, simple or complex, does it require high accuracy or is it fault-tolerant?

Mental requirements

1. Low ... 7. High

Question 2

How much physical activity was required (e.g. pulling, pressing, turning, controlling, activating ...)? Was the task easy or difficult, easy or exhausting, restful or tedious?

Physical demands

1. Low ... 7. High

Question 3

How much time pressure did you feel in terms of the frequency or rhythm with which the tasks or task elements occurred? Was the task slow and leisurely or fast and hectic?

Time requirement

1. Low ... 7. High

Question 4

In your opinion, how successfully have you achieved the goals set by the experimenter (or yourself)? How satisfied were you with your performance in pursuing these goals?

Achievement

1. Good ... 7. Bad

Question 5

How hard did you have to work to achieve your level of task accomplishment?

Effort

1. Low ... 7. High

Question 6

How insecure, discouraged, irritated, stressed and upset (versus safe, confirmed, satisfied, relaxed and satisfied with yourself) did you feel during the task?

Frustration

1. Low ... 7. High