

Investigating the Relationship between Binocular Disparity, Viewer Discomfort,  
and Depth Task Performance on Stereoscopic 3D Displays

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

BY

JOHN PAUL MCINTIRE

B.A., University of Dayton, 2005

M.S., Wright State University, 2007

---

2014

Wright State University

WRIGHT STATE UNIVERSITY  
GRADUATE SCHOOL

April 10, 2014

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY John Paul McIntire ENTITLED: Investigating the Relationship between Binocular Disparity, Viewer Discomfort, and Depth Task Performance on Stereoscopic 3D Displays BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

---

Scott N.J. Watamaniuk, Ph.D.  
Dissertation Director

---

Scott N.J. Watamaniuk, Ph.D.  
Graduate Program Director

---

Debra Steele-Johnson, Ph.D.  
Chair, Department of Psychology

---

Robert E.W. Fyffe, Ph.D.  
Vice President for Research and Dean, Graduate School

Committee on Final Examination

---

Scott N.J. Watamaniuk, Ph.D. – Wright State University

---

Paul R. Havig, Ph.D. – US Air Force Research Laboratory

---

Robert H. Gilkey, Ph.D. – Wright State University

---

Kevin B. Bennett, Ph.D. – Wright State University

## ABSTRACT

McIntire, John Paul. Ph.D., Department of Psychology, Human Factors Psychology Program, Wright State University, 2014. Investigating the Relationship between Binocular Disparity, Viewer Discomfort, and Depth Task Performance on Stereoscopic 3D Displays.

Stereoscopic 3D (S3D) displays offer the capability to enhance user performance on a variety of depth tasks. However, the benefit derived from viewing S3D depends in part on the magnitude of binocular disparity that is displayed. Surprisingly few studies have directly investigated the relationship between disparity and depth task performance. The studies that have been conducted suggest that a minimum amount of disparity (10-50 arc min) may be needed to improve performance over conditions in which no S3D is present, but it is unclear the extent to which performance might improve with increases in disparity beyond this range.

From a human factors perspective, there are compelling reasons for using binocular disparities that are smaller than a strict geometrical interpretation of the scene would require (i.e., microstereopsis); one reason is to make the viewing experience more comfortable. This is important because S3D displays appear to cause a variety of simulator sickness-type problems for as many as 25-50% of users (including eye strain, headache, nausea, etc.). Preliminary evidence on the use of microstereopsis suggests that it does indeed result in a more comfortable and less fatiguing depth percept, particularly if binocular disparity is limited to a maximum of about 60 to 70 arc min (the One Degree Rule). But does microstereopsis also negate the performance benefits of stereopsis? How much can disparities be reduced before performance decrements are noticeable, and how comfortable are these disparities? Is there a stereo “sweet spot” in which both performance and comfort are high? And is this sweet-spot dependent on the particular depth task being tested?

Results from a simple 2 degree-of-freedom (DOF) virtual precision object alignment task showed that when averaged across participants, maximum performance was achieved when

disparity was limited to +/- 80 or 100 arc min of disparity during a 30 minute session.

Performance with S3D cues improved alignment accuracy by up to 80% compared to no stereo cues, though several participants received an inconsistent benefit, and in a few cases, S3D resulted in detrimental performance. The tested magnitudes of disparity limits were also generally comfortable, although a significant correlation between increasing disparity and decreasing comfort was confirmed. Several optometric measures (e.g. stereoacuity, fusion ranges) predicted performance, but not comfort, on S3D displays.

Results from a more complex 5 DOF virtual precision object alignment task showed that the best performance was achieved with disparity limits from +/- 60 to 100 arc min of disparity. Again, the tested magnitudes of disparity limits were generally comfortable, and several optometric measurements predicted performance but not comfort.

Overall, the results suggest that the One Degree Rule for stereoscopic disparity limits can be expanded for near-viewing desktop applications. The results also suggest that while camera separations resulting in microstereopsis showed improved performance over no-stereopsis conditions, best performance is achieved with orthostereoscopic or near-orthostereoscopic levels of camera separation. The findings provide little support for Postural Instability Theory, but some support for Cue Conflict Theory, as useful guides for studying and mitigating viewer discomfort and simulator sickness symptoms on stereoscopic 3D display applications.

*Keywords:* stereoscopic display, depth perception, binocular disparity, camera separation, 3D

## TABLE OF CONTENTS

	Page
I. INTRODUCTION.....	1
II. GENERAL METHODS.....	20
III. EXPERIMENT ONE – METHOD.....	28
IV. EXPERIMENT ONE – RESULTS AND DISCUSSION.....	32
V. EXPERIMENT ONE – FOLLOW-UP TESTING.....	46
VI. EXPERIMENT TWO – METHOD.....	59
VII. EXPERIMENT TWO – RESULTS AND DISCUSSION.....	63
VIII. EXPERIMENTS ONE AND TWO – COMBINED ANALYSIS.....	96
IX. GENERAL RESULTS AND DISCUSSION.....	103
X. CONCLUSIONS AND FUTURE WORK.....	111
XI. TABLES.....	117
XII. ATTACHMENTS.....	135
XIII. APPENDICES.....	139
XIV. REFERENCES.....	149

## LIST OF FIGURES

	Page
1-1. An illustration of the Vergence-Accommodation (VA) conflict.....	5
1-2. Mean alignment error along the Z-axis (camera separation).....	10
1-3. Binocular disparity.....	11
1-4. Mean alignment error along the Z-axis (binocular disparity).....	11
1-5. Task completion times (camera separation).....	12
1-6. Task completion times (binocular disparity).....	13
1-7. Viewing comfort (binocular disparity).....	14
1-8. The comfortable stereoscopic viewing zone limits.....	15
3-1. Rosenberg's virtual object positioning task.....	28
3-2. A screenshot and schematic of the virtual objects.....	29
4-1. The main effect of disparity limit on placement error.....	33
4-2. The main effect of disparity limit on placement error, by participants.....	36
4-3. The main effect of disparity limit on placement error, by minority subgroup.....	37
4-4. The main effect of disparity limit on SSQ scores.....	39
5-1. The main effect of disparity type on placement error.....	51
6-1. A side-view diagram of the experimental set-up.....	61
6-2. The TUI spatial input device.....	62
7-1. The main effect of disparity limit on positional error.....	65
7-2. The main effect of disparity limit on rotational error.....	66
7-3. The effect of disparity limit on positional error, per participant.....	69
7-4. The virtual objects (positional versus shape disparity).....	71
7-5. The effect of disparity limit on rotational error, per participant.....	74

7-6. The effect of disparity limit on positional error, per spatial dimension (x, y, z).....	77
7-7. The effect of disparity limit on positional error, per spatial dimension (x, y).....	78
7-8. The effect of target orientation in depth on positional error.....	79
7-9. The effect of lateral target orientation on positional error.....	81
7-10. The effect of lateral target orientation on rotational error.....	82
7-11. The effect of target orientation in depth on rotational error.....	84
7-12. The effect of target location in depth on positional error.....	87
8-1. Topological space showing relationships between clusters of data points.....	98
8-2. The topological clusterings color-coded by optometric measures.....	99
8-3. The topological clusterings color-coded by discomfort and performance.....	101

## LIST OF TABLES

	Page
1. Different environmental situations and the implications of Cue Conflict Theory....	117
2. Summary of experimental studies that have manipulated binocular disparity.....	118
3. Binocular disparities, camera separations, virtual IPDs, and viewing distances.....	119
4. Participants' performance with 2D versus S3D display in Experiment One.....	120
5. Partial correlations between disparity limits and SSQ in Experiment One.....	121
6. SSQ results across disparity limits in Experiment One.....	122
7. Partial correlations between eyestrain, postural instability, SSQ, and disparity limits in Experiment One.....	123
8. Correlations between USAFSAM pre-screening optometric data and S3D performance in Experiment One.....	124
9. Correlations between the pre-session optometric data and S3D performance in Experiment One.....	125
10. Stereoacuity threshold measurements for Experiment One.....	126
11. Partial correlations between disparity limits and SSQ in Experiment Two.....	127
12. SSQ results across disparity limits in Experiment Two.....	128
13. Partial correlations between eyestrain, postural instability, SSQ, and disparity limits in Experiment Two.....	129
14. Participants' performance with 2D versus S3D display in Experiment Two.....	130
15. Correlations between USAFSAM pre-screening optometric data and S3D performance in Experiment Two.....	131
16. Correlations between the pre-session optometric data and S3D performance in Experiment Two.....	132
17. Distinguishing optometric measurements in the topological data analysis.....	133
18. Comparing exophoria, exophoria, and orthophoria groups.....	134

## ACKNOWLEDGMENTS

I am deeply indebted for the help, guidance, support, and collaboration provided by my dissertation committee members: Dr. Paul Havig (AFRL), Dr. Scott Watamaniuk (WSU), Dr. Robert Gilkey (WSU), and Dr. Kevin Bennett (WSU). Special thanks are due to Dr. Havig and Dr. Lawrence “Kent” Harrington, for all of their guidance, advice, and many fruitful discussions of this work. Thanks for the many years of support from my branch, the Battlespace Visualization branch (RHCV) in the 711<sup>th</sup> Human Performance Wing of the United States Air Force Research Laboratory. Thanks specifically to Eric Heft for programming the experiment and for his help in working through much of the binocular imaging geometry with me, and to Simon Su for help with utilizing the electronic balance board and other VR technology.

Thanks also to Dr. Steven Wright, Alex Van Atta, Marc Winterbottom, Dr. Darrell Rouse, and other members of the US Air Force School of Aerospace Medicine’s OBVA Lab for their support in collecting data, finding volunteers, conducting optometric testing, and for the many helpful reviews of my work. And thank you to Dr. Leslie Blaha (AFRL) and Dr. Dustin Arendt (AFRL) for their assistance with data analysis and visualization.

Thank you to all of the participants who volunteered to contribute their valuable time.

Most of all, from the bottom of my heart, I owe a deep thanks to the two people who matter most to me in the whole world, Lindsey and Megan, for their eternal love and support.

A handwritten signature in black ink, appearing to read "John Paul McIntire". The signature is fluid and cursive, with a large loop on the first letter 'J' and a long horizontal stroke at the end.

John Paul McIntire

April 10, 2014

## I. INTRODUCTION

*“The stereoscope is not a toy: it is a divine gift, placed in our hands nominally by science, really by that inspiration which is revealing the Almighty through the lips of the humble students of Nature.”*

- Oliver Wendell Holmes (as quoted in Rowlands & Killian, 1937)

Great thinkers as early as Euclid, Galen, and Leonardo da Vinci had realized that the two eyes receive slightly different images, but all had failed to notice what this ‘binocular disparity’ implied for visual depth perception (Wade, 1987). Not until the 1800’s did Sir Charles Wheatstone provide the modern explanation for stereoscopic vision (binocular parallax) and subsequently invented the first stereoscopic 3D (S3D) display, the “stereoscope,” (Wheatstone, 1838). The physicist David Brewster improved upon and popularized stereoscopes, and S3D viewing technology was further developed by the famous German scientists Helmholtz and Pulfrich (Waack, 1987). The novelty of the technology and the compelling perceptual experience of depth elicited by stereoscopic displays have fascinated viewers ever since their invention. Despite waxing and then waning in popularity several times over the last 175 years, the current resurgence of S3D may suggest it is here to stay, as 3D is now finding wide interest and application in entertainment (e.g., especially in movies and in games), medicine, industrial design, education and training, and in the military (McIntire, Havig, & Geiselman, 2012; 2014).

There are a surprising number of examples of stereoscopic 3D display technologies being applied to real-world problems, both historical and modern. During World War II, British aerial surveillance photography was viewed stereoscopically to help image analysts detect and identify camouflaged German ground targets that would have been nearly-impossible to see when viewed in non-stereo 2D. This use of stereoscopic technology helped the Allies identify and attack

hidden Nazi rocket bases threatening Britain, and even helped ensure the success of the D-Day landings (NOVA, 2012).

The medical community is another field where S3D display technologies have found many real-world applications. A 1937 review article of stereoscopic displays noted the use of 3D X-rays, for which the technology provided “incalculable value” to doctors attempting to determine the location of foreign objects or defects inside human bodies; the authors also noted its important use in fluoroscopy (Rowlands & Killian, 1937). The best-known modern medical example of stereo 3D is the da Vinci surgical robot system, which received FDA approval in 2000. This system utilizes a stereoscopic 3D display for surgeons to teleoperate remotely controlled robotics on surgical patients, and for a wide variety of different types of surgeries (Intuitive Surgical, Inc., 2013). About 200,000 surgeries a year are now conducted using the da Vinci system.

In the modern military, stereoscopic 3D displays are being investigated for their possibility of improving performance on visually-demanding and depth-related tasks, such as image analysis and intelligence tasks (e.g., Peinsipp-Byma, Rehfeld, & Eck, 2009), air traffic control (e.g., Russi, 2013), teleoperative robotics and indirect-vision driving (e.g., Chen, Oden, Drexler, & Merritt, 2010; Chen, Oden, & Merritt, 2014). They are being utilized in the next-generation of aerial refueling tanker aircraft (the Airbus MRTT, Boeing KC-767, and Boeing KC-46A), in which the refueling boom operators will, for the first time ever, view *only* an S3D image of the boom arm and the refueling aircraft during operations. There will be no out-of-the-glass boom window view for the operator; there will be no boom pod as there has been for over 60 years; there will be only a high-definition stereoscopic 3D display and its associated eyeglasses (Remote Vision System, 2011).

Relative to traditional 2D displays, stereoscopic displays have been shown to enhance performance on a variety of depth-related tasks. These tasks include judging distances, finding and identifying objects by breaking camouflage and eliciting perceptual “pop-out,” performing spatial manipulations of objects such as positioning and tracking, and navigating difficult or complex terrain. More cognitively, stereoscopic displays can improve the spatial understanding of 3D scenes or complex objects, improve recall of scenes or objects, and improve learning of spatial relationships. A review of the performance benefits offered by stereoscopic 3D displays can be found in McIntire, Havig, & Geiselman (2012, 2014). Despite the many benefits offered on depth-related tasks, there are also some particular disadvantages when one considers the human factors issues surrounding these displays. These will be discussed next.

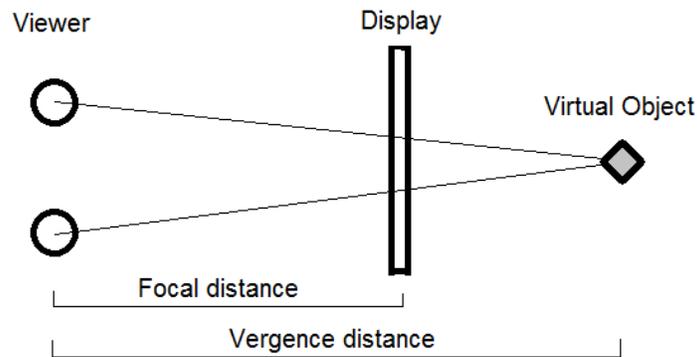
### **Simulator Sickness and the Accommodation-Convergence Conflict**

Stereoscopic displays offer a host of unique human factors issues and potential problems for users. The reported problems include headaches, eyestrain, fatigue, perceptual disturbances, discomfort, disorientation, malaise, nausea, and even in some cases flashbacks (Singer et al., 1995). In other words, virtual environment (VE) displays in general and S3D displays in particular seem to elicit visual discomfort and related sickness symptoms. The patterns of symptoms for VE displays are sometimes called *cybersickness* (cybersickness definitions and reviews given in Stanney, Kennedy, & Drexler, 1997; LaViola, 2000), *virtual-reality induced symptoms and effects* (VRISE review given in Cobb, Nichols, Ramsey, & Wilson, 1999), and *simulator sickness* or *simulator adaptation syndrome* (simulator sickness reviews in Kolasinski, 1995; Lawson et al., 2002; Pausch et al., 1992; Mollenhauer, 2004).

What causes simulator sickness or cybersickness-like symptoms in S3D displays? Binocular rivalry has been suggested as a major contributing factor (Stanney, Mourant, &

Kennedy, 1998). Kooi and Toet (2004) also cite binocular rivalry, which can be caused by a variety of possible “stereo imperfections” including interocular cross-talk, imaging distortions (e.g., keystone via vertical parallax, camera alignment error, etc.), and luster as contributing causes to viewer discomfort. Besides binocular rivalry, a major contributing factor of interest to this study is the *Vergence-Accommodation (VA) conflict* (Ehrlich, 1997; Lambooi et al., 2007; Häkkinen et al., 2006; Yano et al., 2004; Hoffman et al., 2008; Banks et al., 2008; Wöpking, 1995; Wann, Rushton, & Mon-Williams, 1995; Shibata et al., 2011a/b; Yang & Sheedy, 2011; Inoue & Ohzu, 1997). The VA conflict is a sensory/perceptual conflict in which the visual system’s accommodative response (focus) is in conflict with the alignment of the eyes (vergence), disrupting a process that is normally neurologically linked to respond in synchrony (Howard, 2002).

In fact, these two systems (accommodation and vergence) are two of the three interactive neural systems that make up the *Near Triad* or the *Near Vision Complex*, which are optometric terms to describe the coordinated behaviors of the visual system when a viewer is looking at a near object: focus of the crystalline lens, alignment of the eyes, and change of pupil size (von Noorden & Campos, 2002). Unfortunately, the VA conflict is inherent in stereoscopic 3D displays, because these systems indicate depth through binocular disparity (requiring eye vergence) but with no accompanying change in the focus of objects, since they are displayed from two flat 2D images at a fixed distance, as shown in Figure 1-1. In other words, the alignment of the eyes may be signaling depth far in front of or behind the screen while the accommodation system is *always* signaling depth at the surface of the display, resulting in perceptual conflict. Thus, to properly view a 3D image on a stereoscopic display requires the unnatural decoupling of vergence eye movements from the accommodative response.



**Figure 1-1.** An illustration of the Vergence-Accommodation (VA) conflict. On a stereoscopic 3D display, the depth of a virtual object requires vergence eye movements for a particular distance that is different from the focal distance of the display, creating conflict between the two neurologically-linked systems underlying accommodation and vergence.

It is difficult to engineer simple solutions for alleviating the VA conflict, due to the very nature of stereoscopic display systems which require this unnatural sensory system decoupling. We next look at some of the major theories on simulator sickness and their relation to the VA conflict and S3D displays.

### **Theories of Simulator Sickness**

*Cue Conflict Theory* states that when sensory or perceptual systems present conflicting information to the brain, sickness results. This theory traditionally grew out of the motion-sickness literature. The Simulator Sickness Questionnaire or SSQ, which is used to measure simulator sickness, was adapted from the MSQ – the Motion Sickness Questionnaire (Kennedy et al., 1993). Cue Conflict theory is by far the dominant explanation for simulator sickness and motion sickness. Other names for this theory include sensory conflict theory, sensory

rearrangement theory, sensory incongruity theory, sensorimotor conflict theory, perceptual conflict theory, perceptual decorrelation theory, neural mismatch theory, mismatch theory, and incongruity theory (Johnson, 2005). In motion sickness, there is typically a sensory/perceptual mismatch in that the vestibular system is being stimulated due to physical motion, but not in a way that is consistent with the visual scene (see Table 1, preceding the attachments). Think of being below deck of a boat during choppy seas, where there are vestibular cues to motion but not visual cues. This inconsistency creates conflict between two or more sensory/perceptual systems and often results in sickness symptoms, according to the *Poison Theory* (Treisman, 1977), because such conflicts are potentially indicative of poisoning, and we evolved to become sick (and vomit or otherwise expel the poison) when encountering strange sensory/perceptual conflicts. In simpler terms, sensory or perceptual conflict can make us ill.

In simulator/cyber sickness, the opposite pattern of conflict often occurs (see Table 1, preceding the attachments). For instance, in a fixed-base driving simulator, there are clear visual cues to self-motion (e.g., optic flow, motion parallax) but no corresponding vestibular cues as there would normally be when driving a vehicle. If there are vestibular cues, such as in a motion-based simulator, the cues are often inaccurate or inadequate. This mismatch presumably causes conflict which cannot be resolved, again leading to sickness symptoms (Reason, 1970; Mollenhauer, 2004; Johnson, 2005; Kolasinski, 1995). However, this opposite pattern seems to result in a slightly different but related symptomology to motion sickness, leading Hettinger and Riccio (1992) to hypothesize that simulator sickness is a unique and specific type of motion sickness that is primarily visually-induced. Despite the compelling story and high face validity, traditional Cue Conflict + Poison theory does not seem to account for some sickness symptoms

(such as eye strain or blurry vision), does not predict when or how individuals will elicit a response, or explain why some people are more affected than others (Mollenhauer, 2004).

*Postural Instability Theory* is an alternative (or addition) to Cue Conflict theory based on ecological psychology. This theory states that sensory systems are constantly trying to maintain postural stability (or balance) in the environment. Postural stability is achieved when uncontrolled movements to achieve stability are minimized (Riccio & Stoffgren, 1991). Sickness symptoms arise when an individual is attempting to achieve stability but either does not know how (they have yet to learn effective coping strategies) or they are unable to achieve it (due to the unique/unnatural environmental conditions). Postural Instability theory works best as an explanation for motion sickness or motion-based simulator sickness, and the idea that users adopt coping strategies might explain why symptoms seem to diminish over time (e.g., Regan, 1995). However, it does not explain why instability should result in sickness symptoms *per se*; why symptoms appear to be elicited in situations where postural instability does not appear to be an issue (e.g., Lampton et al., 1996); nor does it seem to account for the particular types of eyestrain-related discomfort commonly experienced in simulators, especially when S3D is incorporated.

### **Implications of Cue Conflict Theory for 3D Displays: Microstereopsis**

Sickness, fatigue, and/or other visual discomfort are reported by a significant number of S3D display users. In fact, it had been noted as early as 1869 that incorrectly or poorly photographed stereograms caused headaches for viewers (Waack, 1987). An online survey conducted by the American Optometric Association reported that at least a quarter of people who watched S3D films, TV or videogames experienced symptoms of eyestrain, blurred vision, dizziness, headaches, or nausea (AOA.org, 2010). An informal online survey by

HomeTheater.com found that 53% of people who have viewed 3D content have experienced sickness symptoms (Wilkinson, 2011). There's even popular websites like 3Dsick.com and Motion-Sickness-Guru.com that provide public forums for sharing cures and advice on dealing with the sickness caused by S3D viewing. Simulator or cyber sickness can obviously be a widespread problem when using S3D displays, and has serious implications for the use of S3D in training or operational settings. So how can symptoms be controlled, minimized, or eliminated?

Cue Conflict theory suggests that reducing sensory/perceptual conflicts should help alleviate sickness. Since the VA conflict is caused by a mismatch between oculo-vergence cues to depth and focal cues to depth, then Cue Conflict theory would predict that sickness symptoms might be reduced by decreasing this mismatch. Indeed, in a series of papers, Siegel and colleagues reported that using interocular separations of only a few percent of the normal human inter-pupillary distance (IPD), i.e., *microstereopsis* or *hypostereopsis*, results in “just enough reality” to elicit the subjective impression of stereoscopic depth with less discomfort, which they called “kinder, gentler stereo” (Siegel, 1999; Siegel et al., 1999; Siegel & Nagata, 2000). However, they did not conduct a formal, rigorous experimental evaluation of this question and relied instead on subjective reports.

Some experimental evidence does support the idea that using little or no stereo will lessen sickness. Singer et al. (1995) found that the removal of stereoscopy on a variety of depth tasks resulted in lower simulator sickness scores. They were unsure as to why, although the VA conflict was not considered and seems a reasonable explanation. Wöpking (1995) tested the subjective viewing comfort of binocular imagery while manipulating the VA conflict (via binocular disparity and depth of focus) and showed that as disparity was decreased, subjective comfort increased. Ehrlich (1997) found that using an HMD biocularly was less nauseogenic

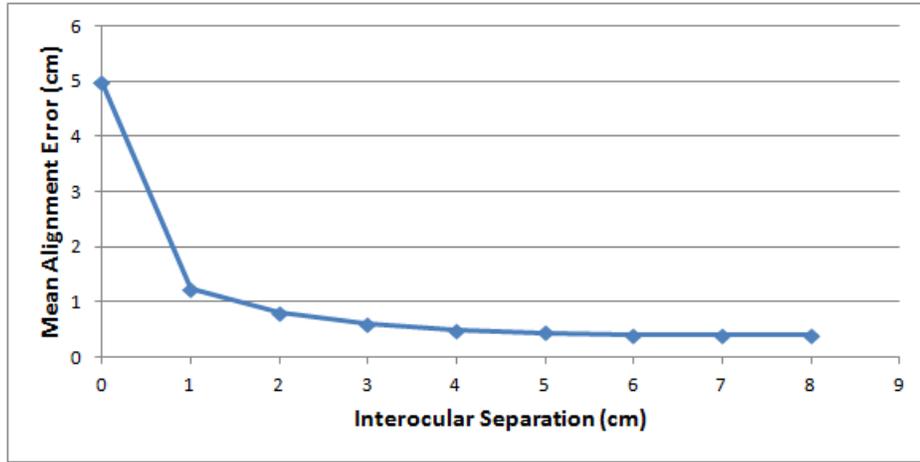
than using one stereoscopically on a variety of depth tasks. Yano, Emoto, and Mitsuhashi (2004) showed that displaying virtual objects inside the viewers' measured depths of field (removing the VA conflict) resulted in lower visual discomfort and fatigue. Häkkinen et al. (2006) found that visual discomfort/disorientation decreased when stereoscopy was removed from a driving simulator. Hoffman et al. (2008) and Banks et al. (2008) showed that when the magnitude of the VA conflict is decreased (by manipulating focus), viewer fatigue and discomfort are reduced while visual performance is increased (as measured by the response times for identifying a stereoscopic stimulus, stereoacuity on a time-limited task, and accuracy for judging perceived depth). As Cue Conflict theory would predict, the preliminary subjective reports and some experimental data support the notion that lessening disparity will result in a more comfortable S3D viewing experience.

### **Microstereopsis and Performance**

This assertion that microstereopsis is “good enough 3D” has important implications for the design and use of S3D displays, and thus needs clear empirical validation. However, a vitally important question in the use of microstereopsis is what happens to depth-task performance with lower levels of disparity? Perhaps sickness symptoms are improved, but does performance suffer? Is there a comfortable middle-ground where both performance and comfort are high? Can we characterize this trade-off and its range, and is it task-dependent? These are some of the interesting experimental questions that are in need of an answer.

Rosenberg (1993) specifically manipulated camera separation to investigate its effects on depth-task performance. He used a simple virtual object alignment task and measured placement accuracy. He found that varying the virtual camera separation between 0 and 8 cm resulted in a

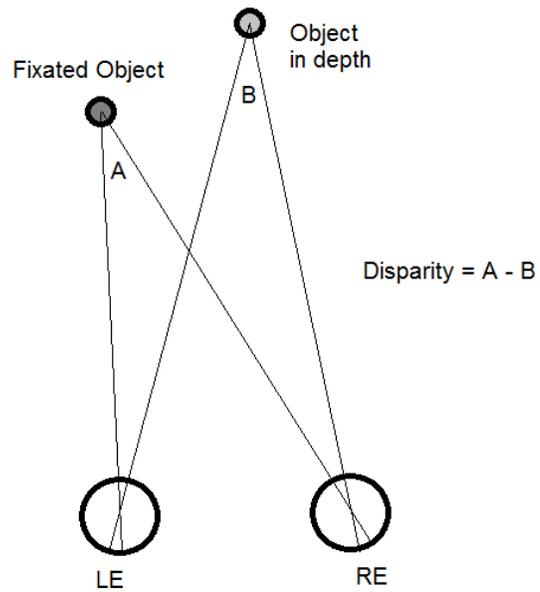
logarithmic relationship between separation and performance as measured by mean object alignment error in depth (see Figure 1-2 below).



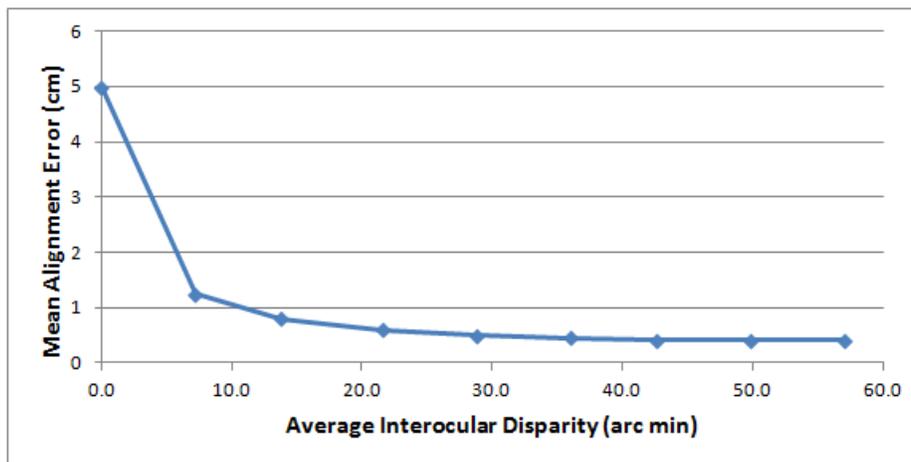
**Figure 1-2.** Mean alignment error along the Z-axis, as a function of binocular (interocular) camera separation distance. Adapted from Rosenberg (1993).

As binocular separation was increased from zero, performance climbed dramatically (alignment errors dropped) before starting to level off at around 2 cm of separation, and with no additional performance benefit being provided from separations greater than 3 cm. Instead of looking at this data in terms of eye separation, we can convert Rosenberg’s stated inter-camera distances into average binocular disparities. Binocular disparity is calculated as the difference between the angle of the fixation point and the angle of the object off the fixation plane (see Figure 1-3; also see Appendix 1 for further information on this calculation).

Given Rosenberg’s (1993) display setup (viewing distance = 80 cm, effective display volume =  $\pm 20$  cm, and camera separations from 0 to 8 cm) and that virtual target positions were randomly determined within the volume, then average binocular disparities can be estimated for each given camera separation. Using this information, the data from Figure 1-2 is re-drawn in Figure 1-4.



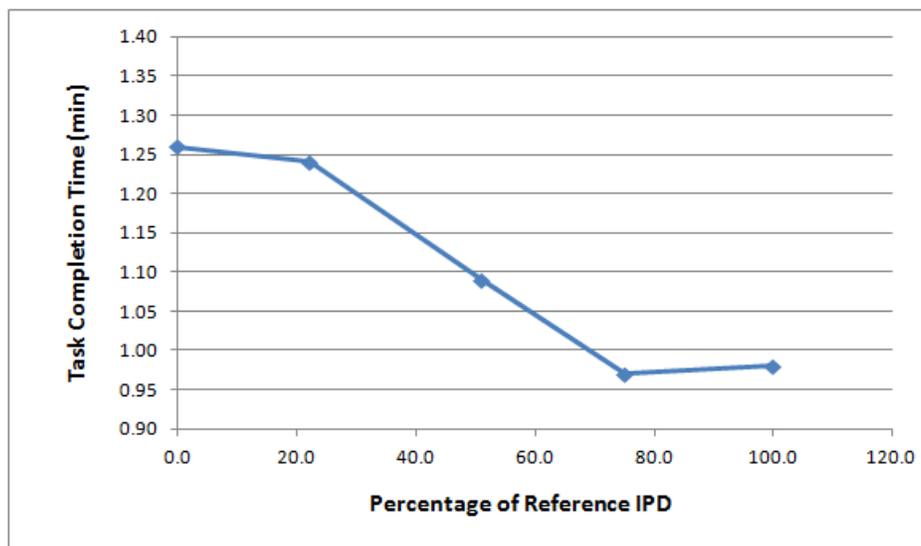
**Figure 1-3.** Binocular disparity is the difference between the angle formed by the fixated object and the eyes, and the angle formed by the object in depth.



**Figure 1-4.** Mean alignment error along the Z-axis, as a function of average binocular (interocular) distance. Adapted from Rosenberg (1993).

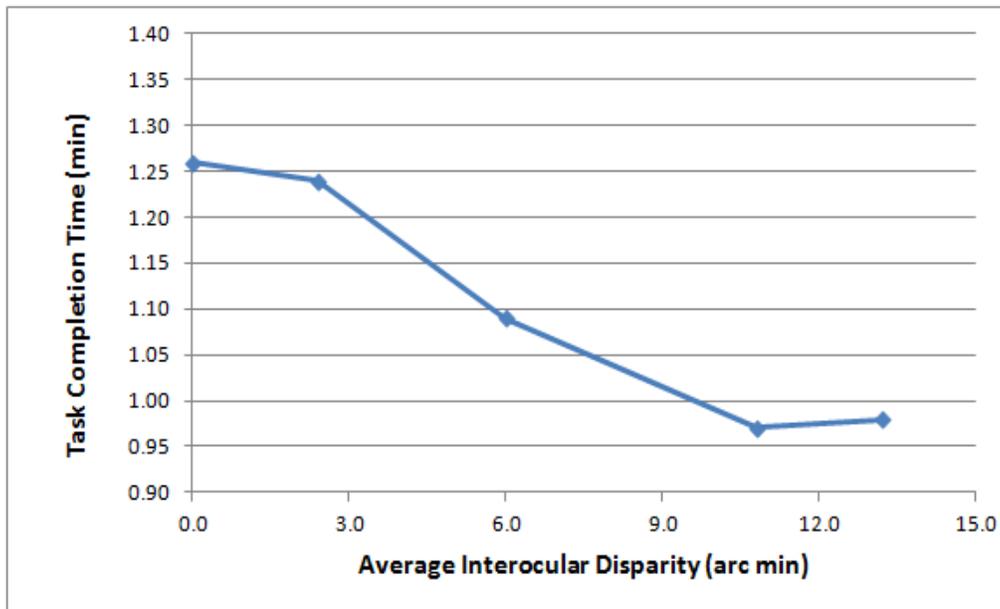
The important note from this re-plotting of the data is that performance appears to level off at about 20 arc min of binocular disparity, with no discernable increase in depth placement accuracy despite increases in disparity beyond this point. Notice, too, that the traditional rule-of-thumb advice for comfortable stereoscopic viewing is to not exceed 70 arc min of binocular disparity (e.g., Wöpking, 1995; see *The One Degree Rule* section below). Rosenberg's results thus suggest that there is a reasonably wide range (from about 25 to 70 arc min) in which both performance and visual comfort can be high when using stereoscopic displays.

Ellis et al. (2005) also varied camera separation to study its effect on performance. They used an object placement task requiring a series of precise ring placements with a surgical telerobot. Performance was measured via total completion time. Their results suggested that as camera separation was decreased from 100% of IPD to 75%, performance stayed high (response times were still fast). At 50% of IPD, performance dropped (completion time increased) but was still significantly better than at 25% or zero IPD. These results are shown in Figure 1-5.



**Figure 1-5.** Task completion times as a function of camera separation, in units of reference IPD. Adapted from Ellis et al. (2005).

Again, we can re-plot these data in terms of approximate binocular disparity, instead of in terms of IPD (see Figure 1-6). Here we find that performance seems to plateau at disparities of about 10 arc min and larger. This is about half the disparity needed for maximum performance according to Rosenberg's (1993) results, which conceivably widens the "sweet spot" between good depth task performance and user discomfort (which is reached, according to previous research, as binocular disparity approaches 70 arc min).

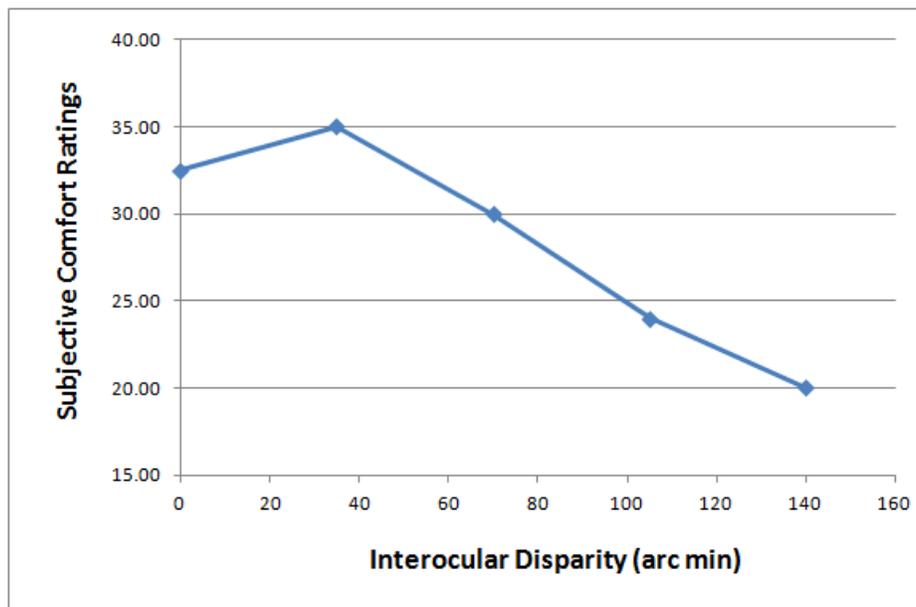


**Figure 1-6.** Task completion times as a function of binocular (interocular) disparity. Adapted from Ellis et al. (2005).

### Microstereopsis and Comfort

What about the magnitude of binocular disparities and visual comfort? Wöpking (1995) also manipulated camera separation to study viewer comfort. Shown in Figure 1-7 are the subjective visual comfort ratings for moderate to high-resolution stereoscopic imagery (above 5 cycles/deg in both foreground and background). We can see that lower disparities ( $\leq 40$  arc min) result in the highest comfort, with a linear decline in comfort following further increases in

disparity; and comfort dropped considerably at 70 arc min and beyond. We should additionally note Wöpking's results that showed imagery with lower resolution backgrounds (less than 5 cycles/deg) were much more comfortable to view at higher disparities, providing support for the idea that the VA conflict contributes to viewer discomfort, since this conflict is lessened by using lower spatial frequency background stimuli (i.e., this is roughly analogous to blurring the backgrounds, which dampens conflict with the accommodative system).



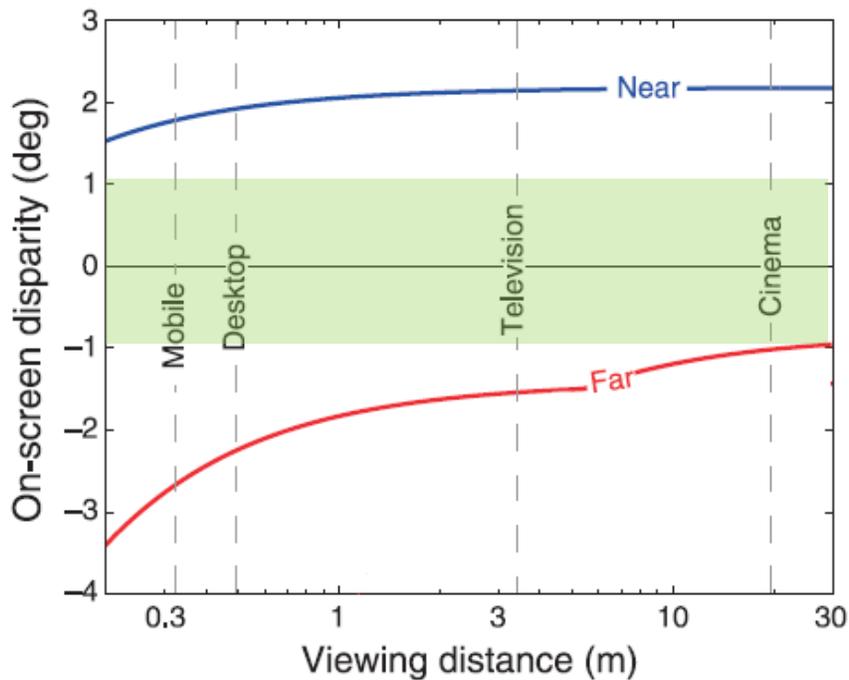
**Figure 1-7.** Viewing comfort as a function of binocular (interocular) disparity.

Adapted from Wöpking (1995).

### **The One Degree Rule and the Comfort Threshold**

The general rule-of-thumb in the literature supported by theory and some experimental evidence is that no more than  $\pm 1$  degree of binocular disparity (or about 60-70 arc min) should be displayed to avoid inducing visual discomfort (e.g., Wöpking, 1995; Lambooj, IJsselsteijn, & Heynderickx, 2007; Shibata et al., 2011a/b; Nojiri et al., 2004; Pastoor, 1993), which we shall refer to hereafter as the *One Degree Rule*. The Japanese 3D Consortium has adopted this rule as

well in their official Safety Guidelines, citing a fusional limit of 2 deg and a practical disparity limitation of 60 arc min (Kim, Choi, & Sohn, 2011). This suggestion places a firm limit on the amount of binocular disparity that can be used comfortably in S3D display space, particularly for desktop systems. Binocular disparity limits are also sometimes referred to as a “depth budget” or “disparity range” (e.g., Kytö, Hakala, Oittinen, & Häkkinen, 2012). The experimental work of Shibata et al. (2011b) in determining the “zone of comfort” for stereoscopic displays provides clear support for the One Degree Rule, as shown in Figure 1-8 below. Implementing a One Degree Rule for both crossed (near) and uncrossed (far) disparity ensures that all on-screen disparities are kept within the comfortable viewing zone, and this is true across various display devices and viewing distances.



**Figure 1-8.** The comfortable stereoscopic viewing zone limits based on human factors results (near = crossed; far = uncrossed disparities), as a function of viewing distance. The One Degree Rule is overlaying these results: the shaded

green region indicates the spatial region that lies within  $\pm 1$  degree of disparity.

Adapted from Shibata et al. (2011b) Figure 23D.

An experiment by van Beurden, IJsselsteijn, and de Kort (2011) studied depth task performance, workload, and visual comfort while manipulating disparity (ranging from 0 to 50 arc min) and task difficulty. This is the only study we have found that measured both performance and comfort while manipulating disparity. On a visual path-tracing task of a 3D wireframe structure, accuracy of responses peaked at about 25 arc min of disparity. Response times were fastest at about 25 arc min under the easier condition but in the more difficult condition, responses were fastest at 50 arc min. Somewhat surprisingly, discomfort did not noticeably increase with disparity (although there may be some methodological issues in the way they measured discomfort; i.e., participants rated discomfort using a single 20-point rating scale only at the end, and the before/after discomfort change was not measured). Task difficulty interacted with most of these results: discomfort was higher on the more difficult task; accuracy declined on the harder task; and completion times were longer. Generally, we see that the manipulation of disparity had a more pronounced or magnified impact on the more difficult task relative to the easier task. These results suggest that the amount of disparity needed for best performance is on the order of 25 to 50 arc min, at least on their task (visual path-tracing).

In summary, experimental evidence shows that the perception of stereoscopic depth might be comfortably elicited using smaller binocular disparities (or camera separations) than is required for strict geometric accuracy (i.e., microstereopsis versus orthostereopsis), and the evidence also suggests that perhaps depth task performance can still be high, even when using these smaller disparities.

## **Binocular Status Index: Predicting Performance and Comfort on S3D Displays**

Lambooij, Fortuin, IJsselsteijn, Evans, and Heynderickx (2011) cite a binocular status index developed by Evans that relies on a brief questionnaire and a rapid optometric screen (including one-eye cover tests, associated phoria, dissociated phoria, binocular suppression, etc.). The index classifies users into either “moderate” or “good” binocular status groups (MBS versus GBS). In a previous study, these same researchers (Lambooij, Fortuin, IJsselsteijn, Evans, & Heynderickx, 2010) found that 18% out of 39 participants were classifiable as MBS, and that the MBS group had more visual discomfort and objective signs of visual fatigue after a short stereoscopic reading task. In their more recent replication (Lambooij et al., 2011), these researchers again found 18% of participants from a sample of 33 could be classified into an MBS group. The MBS group had significantly more complaints of headache, pain, strain, and discomfort, especially with larger disparities ( $\pm 0.75$  and larger) when using the stereo 3D display. However, there were no differences between the groups’ visual performance on accommodative facility and vergence facility measurements, nor on the 3D reading task.

There are a few possible issues to note with the classification scheme for the index. One, the most heavily weighted aspect of Evans’ binocular status index is a subjective questionnaire regarding binocular vision-related complaints (e.g., symptomatic heterophoria). This adds an air of circularity to some of their findings, since people that complain about having binocular vision-related problems also happen to complain more about doing tasks on stereoscopic viewing devices, which requires utilizing the binocular visual system. Two, it’s not clear what the real-world utility of the index might be, since the index was not able to discriminate between objective measures of visual performance between the groups (again, there were no differences between the groups’ visual performance on accommodative facility and vergence facility

measurements, nor on the 3D reading task). Third, instead of classifying into two groups, these researchers might have correlated participants' binocular status index measures with their performance and subjective measures, as this seems like it could have yielded more informative and predictive results. In any case, finding a robust index that captures individual differences in performance and comfort with stereo 3D displays is certainly a worthy goal and deserves further experimental work. The present work touches on some of this.

### **Purpose of This Research**

Only one experiment in the literature appears to have specifically investigated the trade-off between the seemingly conflicting goals of minimizing discomfort while maximizing performance on stereoscopic depth tasks via the manipulation of binocular disparity (van Beurden, IJsselsteijn, & de Kort, 2011). However, they did not explicitly set out to study this trade-off, and their results were somewhat mixed about what levels of disparity are best for performance, workload, or comfort. Additionally, only a few other experiments have manipulated disparity to specifically study its effects on performance (a summary of these is given in Table 2, preceding Attachment 1).

Siegel and colleagues (Siegel, 1999; Siegel et al., 1999; Siegel & Nagata, 2000) and others (e.g., Aitsiselmi & Holliman, 2009) have noted the need for further research and formal experimental work on microstereopsis. The present research is an attempt at simultaneously investigating the effect of manipulating binocular disparity (by effectively introducing various levels of microstereopsis, and one level of orthostereopsis) on depth-task performance and simulator sickness symptoms. Various possible predictors and objective indicators of both discomfort and/or performance were also investigated. And the experiments serve as a test of Cue Conflict Theory versus Postural Instability Theory, since the two theories make different

predictions regarding the relationships between perceptual conflict, the extent of sickness symptoms, and postural instabilities.

## II. GENERAL METHODS

### Apparatus

**Stereoscopic 3D Display.** A high-resolution temporally-multiplexed 120 Hz stereoscopic 3D display was used to present the imagery to the participants (NVIDIA Personal GeForce 3D Vision Active Shutter Glasses, and Samsung® SyncMaster™ 2233RZ). This display was a 22-inch diagonal LCD display with a refresh rate of 120 Hz with native resolution of 1680 (horizontal) x 1050 (vertical). This display system required the wearing of electro-optical active shutter glasses that rapidly oscillated between translucence and opacity in synchrony with the display's oscillation between each eye's imagery (at 60 Hz per eye). For the purpose of this study, observers viewed this display at a distance of approximately 24 inches.

**Wii® Balance Board.** This apparatus is an electronic balance board developed by Nintendo® that wirelessly communicates with Nintendo's Wii® console system via Bluetooth short-wave radio. The board can accurately measure up to 330 lbs and is used to record center of pressure displacement and its variability. The board has been shown to be both highly reliable (on test, re-test reliability) and valid for posture testing in VR environments, comparable to other laboratory-grade force platforms (Clark, et al., 2010).

**Keyboard, Mouse, and Tangible User Interface (TUI).** A standard QWERTY keyboard and mouse were used for Experiment One. A Tangible User Interface spatial input device was used for Experiment Two. Further detail on the interaction devices is provided in the Methods section of the relevant experiment.

**Keystone View™ Telebinocular®.** This telebinocular visual screening apparatus by Keystone View™ is a general purpose, rapid optometric screening tool that tests normal visual function via test slides presented on a stereoscopic viewing device. The test slides can be adjusted along the z-dimension (in depth) so that different ranges of focus (from near to far

point) can be tested. The tests are quick, accurate, dependable, and standardized. The Professional Performance series of test cards were utilized.

## **Participants**

Each experiment tested 12 different volunteers, for a total of 24 participants. In the first experiment, we excluded any volunteers over the age of 40 years old to avoid the possibly complicating issue of presbyopia (inability to accommodate to a near distance) which often onsets at this age. Due to further pilot testing which suggested presbyopia to be a non-issue in relation to this research, and to aid in the collection of adequate numbers of volunteers, we relaxed this age restriction in the second experiment. From both experiments, we excluded any volunteers with deficient monocular or stereovision, as determined by standard Snellen acuity test charts, the standard Titmus clinical stereotest, the Keystone Telebinocular stereotest card, and a specialized clinical binocular/stereoscopic vision screening process administered by an optometrist in the US Air Force School of Aerospace Medicine (USAFSAM) Operational Based Visual Assessment (OBVA) Lab. The OBVA lab recorded participants' refractive error, fusion ranges, phorias, and fixation disparities. Their data collection efforts lasted approximately 15-20 min per participant. A sample recruitment correspondence is available in Attachment 3.

## **Binocular Disparity Limits**

Binocular disparity limits were manipulated in all three experiments, across sessions, to correspond to a range between 0 and 100 arc min, in 20 arc min steps. This manipulation was analogous to fixing virtual camera separation in each session to a single value, which differed across sessions. Another analogous way to think about this manipulation is that the virtual IPD ranged from 0 to 100% (assuming an average IPD of 2.6 inches, or 66 mm) in 20% steps of "microstereopsis", with 0% corresponding to a no stereopsis condition, and 100% corresponding

to an orthostereoscopic condition. See the table below for comparisons between these equivalent formulations. Each experimental session presented only one limit/range per session. The order in which disparity limits were presented (one per session) was randomized across participants via a Latin Square design.

<b><u>Stereopsis Level :</u></b>	<b><u>none</u></b>	<b><u>micro-stereopsis</u></b>			<b><u>ortho</u></b>	
Binocular Disparity Limit (arc min)	0	$\pm 20$	$\pm 40$	$\pm 60$	$\pm 80$	$\pm 100$
Virtual camera separation (vIPD%)	0	20	40	60	80	100
Virtual camera separation (mm)	0.0	13.2	26.4	39.6	52.8	66.0

An example should help illustrate these concepts. Every participant would have a binocular disparity limit of  $\pm 60$  arc min in one of their sessions, meaning that no trials during this session presented the virtual objects at a location with a disparity larger than  $\pm 60$  arc min (relative to the display surface). But any given trial within this session could present targets anywhere within the range of  $-60$  to  $+60$  arc min. The actual location of the virtual objects within the viewing volume, in conjunction with the fixed virtual camera separation, determined the ultimate binocular disparity values that a viewer experienced on any given trial.

As an alternative example, the zero disparity condition meant that within this session, there was no virtual camera separation (only one camera’s view shown to both eyes; also referred to as a “bi-ocular” as opposed to a “binocular” view) and hence there were no stereopsis cues to objects’ depths on any trials within this session, regardless of where the objects appeared within the virtual viewing volume. Note, however, that monocular cues to depth such as size and

texture cues would still be visible. Thus, this session served as a baseline of performance with which to compare the other sessions which additionally utilized stereopsis cues to depth. See Table 3 for more details on this design and setup.

This manipulation was meant to test the concept of using disparity limits (the One Degree Rule), different camera separations, and/or ortho- versus micro-stereopsis when displaying stereoscopic 3D stimuli to a viewer over a given time period, relative to a 2D baseline condition.

### **Pre-experimental Data Collection**

An initial participant Demographic Questionnaire was issued before data collection, and gathered the following information: age, gender, video gaming experience, 3D displays/movies experience, motion sickness history, inter-pupillary distance (IPD), and dominant eye and hand (see Attachment 2). As suggested by Ukai and Howarth (2008), we asked participants about their personal history with migraines, as this may be related to motion sickness and possibly simulator sickness on 3D displays. Similarly, in regards to Cue Conflict + Poison theory, each individual's personal history with food poisoning was also recorded (see Demographic Questionnaire, Attachment 2).

### **Viewer Discomfort, Eyestrain, and Fatigue**

This aspect was measured using multiple methods/instruments, both objective and subjective, as described below.

**Simulator Sickness.** Simulator sickness was measured using the standardized Simulator Sickness Questionnaire (SSQ; see attachment 1). The SSQ was administered both pre- and post-test for each experimental session. The SSQ measures three main components of sickness using subscales: Nausea, Oculomotor, and Disorientation. The traditional SSQ as used in simulators uses a weighted-scoring scheme to derive the scores for the subscales and a Total Severity score,

but we used the un-weighted scale and computed our own Total score for several reasons. According to the developers of the SSQ, “there is no particular interpretive meaning to the values in the conversion formulas” (Kennedy et al., 1993, p. 212) and the recommended scoring procedures are meant for only post-exposure ratings (p. 211). But since we are primarily interested in the pre/post changes anyway, we used the raw ratings data for the subscales and simply added them together for a Total score. The SSQ was administered electronically on a separate computer system.

**Phoria.** Both lateral (horizontal) and vertical phorias are measurements of the alignment of the eyes in the absence of binocular stimulation of the vergence system. Essentially, if the eyes do not have binocular stimuli that bring the foveae into alignment, phoria measures the resting state of this system and may be sensitive to manipulations that can stress or fatigue the eyes (see Appendix 2 for relevant citations). Lateral phoria was measured at both near and far focal distances, while vertical phoria was measured only at the far distance. These measurements were taken before and after each experimental session to investigate any oculomotor fatigue/eye strain effects introduced by the S3D display, to supplement the subjective discomfort and sickness measured by the SSQ. Using the KeystoneView™ Telebinocular visual screening apparatus, a standardized optometric device, both near and far lateral phoria measurements were quickly obtained using their included Professional Performance test card set.

**Fusional Range.** Another supplemental objective measure of eye strain/fatigue, fusional range was measured using the S3D display. A stimulus at the plane of the display was moved inward using crossed disparity (towards the viewer in depth) until the image either became blurry or broke into a double image, at which point the viewer signaled with a button press. The image was reset at the depth plane of the screen and moved in the opposite direction, again until the

image either became blurry or broke into two, and again the viewer signaled this event with a button press. This gave measures of the near point of convergence and the far point of convergence (i.e., divergence); also called the fusion near and far points. The measurement of the distance between these two points is the fusion range, which is potentially sensitive to manipulations of eye fatigue (e.g., Nojiri, Yamanoue, Hanazato, Emoto, & Okano, 2004; see also Appendix 2). This measure was taken before and after each experimental session to investigate any oculomotor fatigue/eye strain effects that might be introduced by the 3D display, and also to supplement the subjective discomfort/sickness measured by the SSQ.

### **Postural Instability Task**

We used an electronic balance board (Wii® Balance Board) to record center of balance variability over time. The test required standing in a static posture, upright, with arms to the sides and eyes closed for 30 seconds. Ehrlich (1997) had found that simulator sickness symptoms increase when switching from bi-ocular to binocular displays on VE tasks (in support of what Cue Conflict Theory predicts), but found no effect on postural stability when switching to stereo (against what Postural Instability Theory predicts). We expect to find a similar result. Kolasinski (1994) found that people who were generally less posturally stable (score lower on the pre-test) were more likely to experience motion sickness, so this task will also seek to confirm whether initial instability might predict simulator/cyber-sickness on S3D displays.

### **Depth-Task Performance**

Performance was measured using two depth tasks related to the use of stereoscopic imagery in applied settings. The tasks were meant to be representative of performance tasks that users might conduct on an S3D display, including teleoperative robotics, aerial refueling, and/or surgical tasks. These depth tasks were:

- *Experiment One*: object positioning
- *Experiment Two*: object docking (positioning plus orienting)

## **General Procedure**

The independent variable under manipulation was the binocular disparity limit. The dependent variables under study were depth-task performance, simulator sickness, eye strain, and postural instability. The Simulator Sickness Questionnaire (SSQ; see Attachment 1) was administered pre- and post-testing for each experimental session. Also, objective measures of eye-strain were taken pre- and post-testing: near and far lateral phoria (via the KeystoneView Telebinocular optometric system), and phoria range (via the S3D computer display). The postural instability test was also administered pre- and post-testing for each experimental session. The order in which these tests occurred (within both pre- and post-sessions) was randomized via a Latin Square design.

Each experimental session only tested a single binocular disparity limit, and the number of trials per session were adjusted for each task so that total participation time per session was 30 minutes to ensure adequate time for potentially inducing mild discomfort/eyestrain; most studies on eyestrain, fatigue, or discomfort last anywhere from 30 minutes to a few hours of display exposure and generally only cause mild to moderate eyestrain, if any. Participation was limited to a maximum of two sessions per day, although all participants opted for separate days for all sessions. A summary of the general procedure is shown in Attachment 4.

## **Training**

Participants underwent brief training sessions on each task before formal experimental data collection. For training, the experimental task of interest was performed under the 60 arc

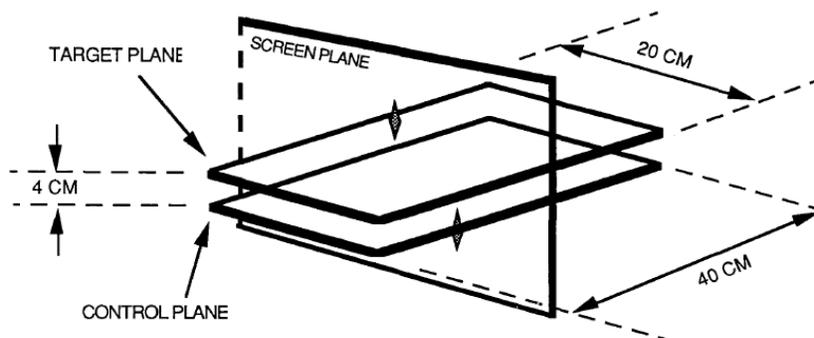
min disparity condition until participants reported being comfortable with the task set-up, the visual stimuli, and the performance demands, and had no more questions.

### III. EXPERIMENT ONE – METHOD

#### 2-DOF Object Positioning Task

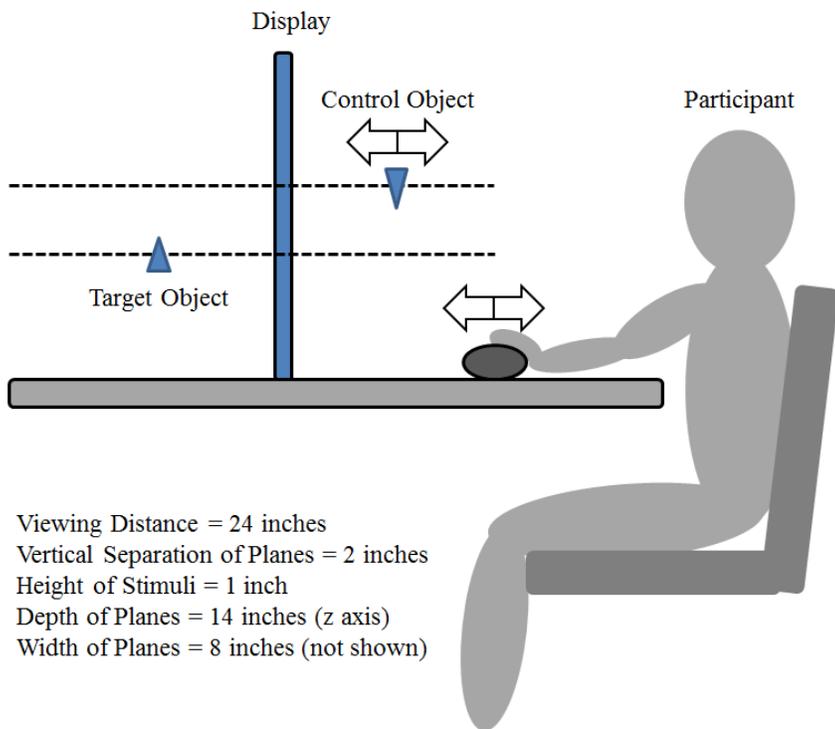
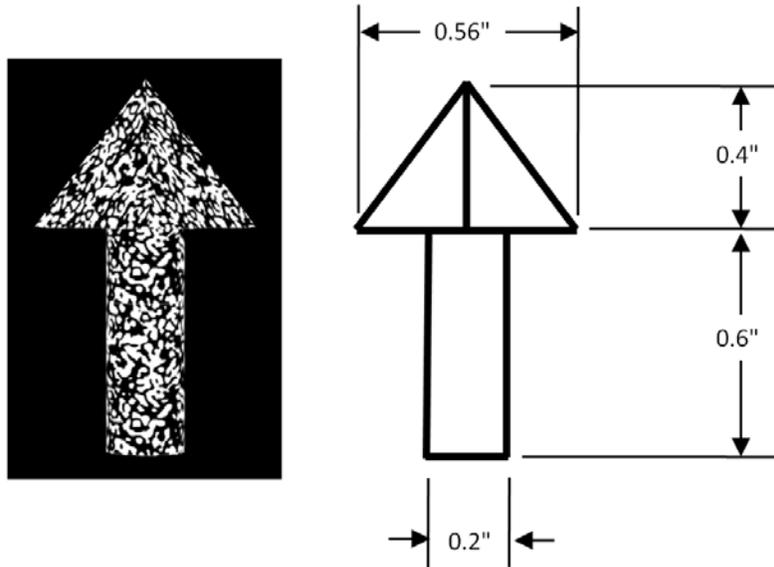
For each trial, the participant used their right hand to control a computer mouse to position a virtual object (e.g., small textured diamond, the “control” arrow) at an indicated depth on the display, matching the depth and vertical positioning of a reference or “target” arrow. This task served as a replication-and-extension of previous work by Rosenberg (1993) who tested a similar virtual object positioning task (using pegs instead of arrows) and measured alignment accuracy, as represented in Figure 3-1.

On each trial, the starting position of the target peg was a randomly chosen point in the 3D space (more specifically, the target could appear anywhere along the x-dimension of the x-y plane, but appeared in depth on the z axis at one of eleven possible distances corresponding to five crossed disparities, five uncrossed disparities, and zero; see Table 3). The control peg always started at the intersection of the control plane and the screen plane, centered along the x-axis. Movement of the control peg was limited to the horizontal (x-z) control plane. The target peg remained stationary at all times.



**Figure 3-1.** Rosenberg’s virtual object positioning task. The participant used a mouse to control the virtual peg’s 2D (x, z) position on the control plane. The task required the vertical alignment of the control peg with the fixed target peg. A

similar task was utilized in Experiment One as a replication of Rosenberg's experiment. This graphic was adapted from Rosenberg (1993).



**Figure 3-2.** A screenshot (top left) and schematic (top right) of the virtual objects that are used in the virtual object alignment task. A side-view diagram of the experimental set-up is also provided (bottom).

As mentioned, the viewer and display arrangements we used were very similar to Rosenberg's (1993) setup, see Figure 3-2. The viewer sat approximately 24 inches from the display. The target and control planes were vertically separated by a gap of 2 inches, and measured 8 inches wide by 14 inches deep. The two planes both extended in the z-dimension of virtual space 5.1 inches coming out of the screen, towards the viewer, and 8.8 inches behind the screen away from the viewer. The planes were not strictly centered in depth, along the z-axis, so that both of the near and far limits would correspond to a maximum disparity of  $\pm 100$  arc min in the condition with the largest virtual eye separation (or virtual camera separation; see Table 3).

Participants clicked the space bar on the keyboard with their left hand when satisfied with their positioning. Performance measures included completion times and positional (placement) error. Positional error was defined as the difference between the optimal placement and the actual placement of the control object in x-z virtual space (absolute value of 2D Euclidean distance). Accuracy, not response time, was emphasized as the primary measure of interest.

The geometry of the virtual space was designed to match, as precisely as possible, the geometry of the real-world viewing space, so that the positional error in "virtual inches" would be approximately equivalent to inches as measured in the real-world viewer/display space. For instance, distance from the virtual camera to a virtual object at the virtual display plane can be equated to viewing distance from the viewer to the virtual object image on the surface of the display. In other words, the virtual camera/imaging/scene space was designed to geometrically

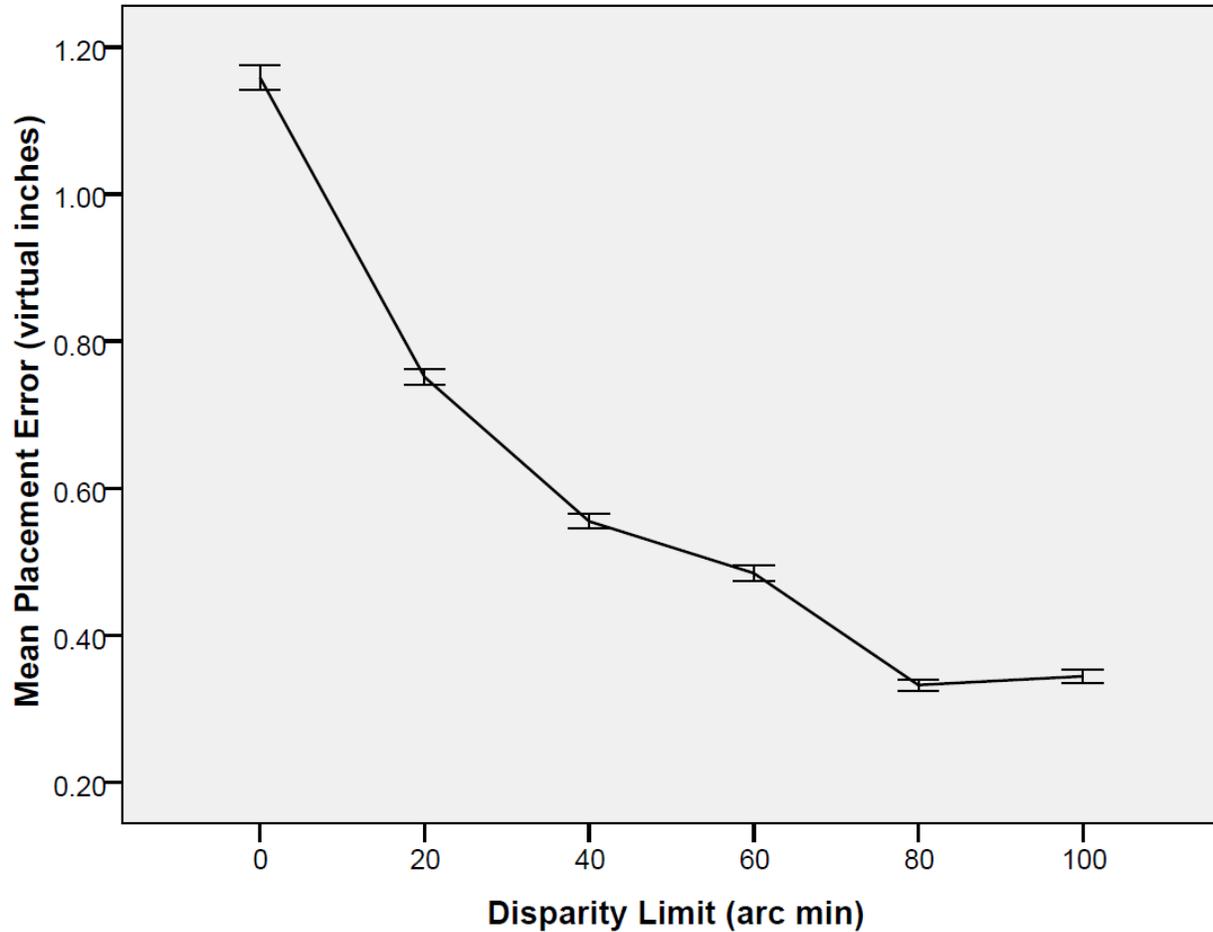
correspond to the viewer/display space so that virtually-measured magnitudes would closely represent the real-world magnitudes in cases where direct measurement of these values would be impossible (virtual inches in depth along z-axis from two 2D computer-generated images). It should also be noted that this measure represents error in 3D but is not a psychological or perceptual measure of “depth perception magnitude” or some such metric; instead it is a virtual analogue of a physical error value.

#### IV. EXPERIMENT ONE – RESULTS AND DISCUSSION

Full factorial repeated measures analysis of variance (ANOVA) tests were conducted when possible, in which participants were treated as random effects using Type III Sum of Squares (i.e., error terms for main effects used main effect x subject interactions). Post-hoc multiple comparisons utilized the Games-Howell (GH) test assuming unequal variances. Reported correlations are either standard Pearson correlation coefficients, or when possible, partial correlations controlling for participant and session (practice) effects. Significance for correlations was tested using one-tailed *t*-tests when a specific direction of influence was suspected (otherwise two-tailed tests were utilized), and paired-sample *t*-tests were used where appropriate. Significance levels for all tests were  $\alpha = .05$ . For Experiment One, two out of 14 possible volunteers were excluded due to atypical or deficient stereoscopic vision.

##### **Disparity Limits and Placement Error**

The primary result of our study, using a repeated measures ANOVA, was that the magnitude of the disparity limit (the range in which disparities were allowed to vary within a session) had a statistically significant effect on placement accuracy [ $F(5,55)=17.44, p<.001$ ]. There was also a significant correlation between the magnitude of the disparity limit and placement error ( $r=-.36, p=.001$ ). These results are not surprising, as it is generally observed that giving larger disparity cues allows for better depth-related task performance (see Introduction and Table 2), and also that S3D is usually helpful on depth-related tasks compared to non-stereo 2D (e.g., McIntire, Havig, & Geiselman, 2012/2014). Figure 4-1 shows this main effect, confirming a beneficial effect of stereopsis cues on alignment performance.



**Figure 4-1.** The main effect of disparity limit on placement error. Error bars represent +/- 1 SEM. All disparity limit levels were significantly different from one another except for the 80 and 100 arc min limits.

We see that in the non-stereopsis condition, when the disparity limit was fixed to zero within a session, average placement errors were on the order of 1.2 virtual inches. By increasing the disparity limit in steps of 20 arc min, performance improved almost linearly until leveling off between 80 and 100 arc min, at just below 0.4 inches of error. This represents a reduction in placement error magnitude of approximately 70%. The GH tests showed that the 80 and 100 arc min comparison was the only one in which the two conditions were not significantly different

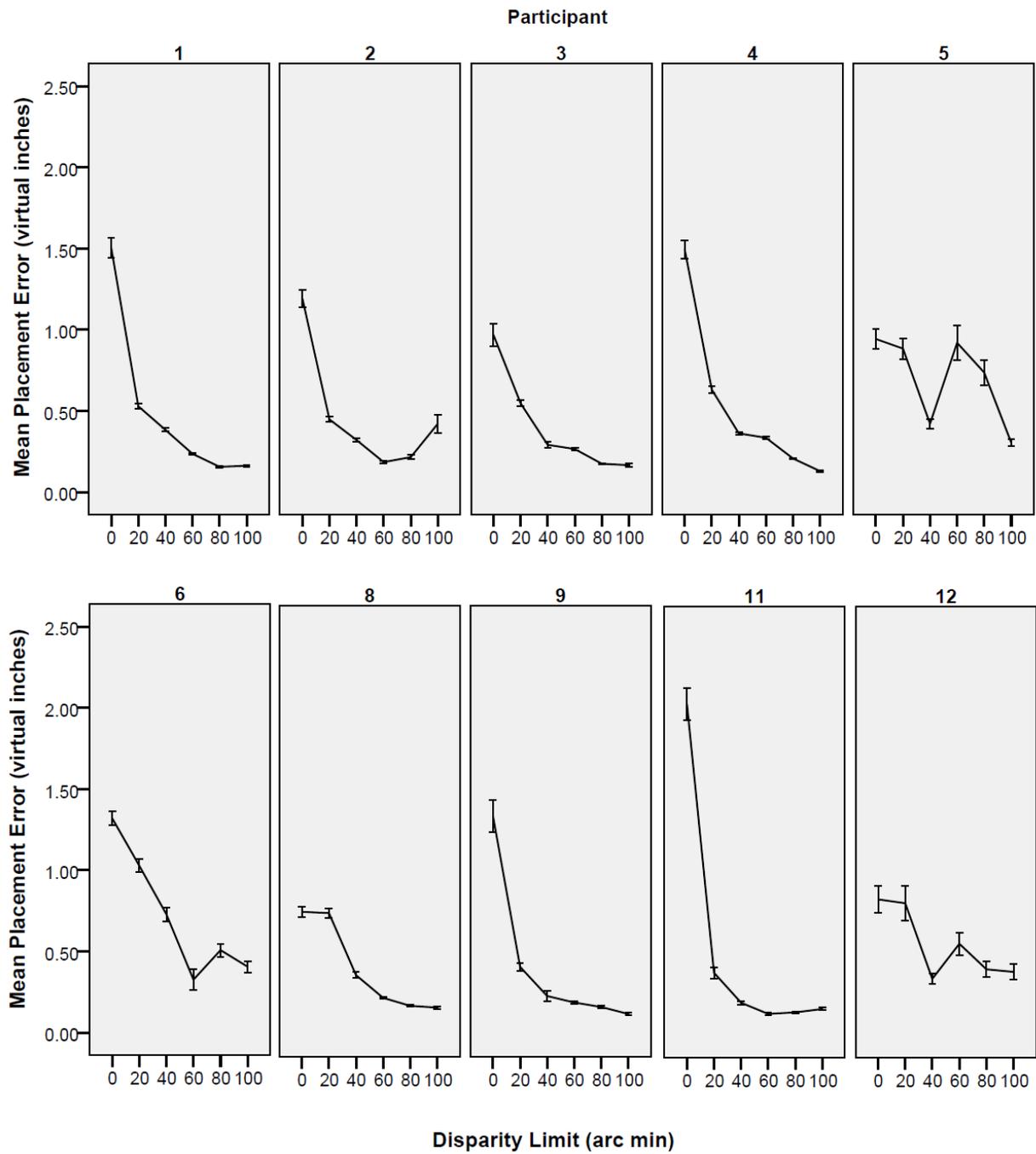
from one another ( $p=.931$ ). Not only did accuracy improve with S3D cues, but the variability of that accuracy improved, with the standard deviation of placement error shrinking by about 50% from the zero disparity condition to the 80 arc min disparity condition. So the provision of S3D cues generally improved both precision and accuracy of virtual object placement.

It is interesting to note that performance seemed to be improved by grossly violating the One Degree Rule for S3D displays, as the 80 and 100 arc min sessions produced significantly better performance than when disparity was limited to 60 arc min (one degree) or lower. This result seems to contradict earlier experimental studies in which increases in disparities beyond about 25 to 50 arc min did not result in noticeable performance advantages (see Table 2). The previous results had suggested that smaller disparity limits (under ~60 arc min) could be just as useful and more comfortable than orthostereoscopic display setups in which larger disparities were present.

It is possible that our experimental design might explain this difference. We used a fair number of participants ( $n=12$ ) and collected many repeated trials: the average number of trials completed by each participant was 300 in each of the six sessions, with a mean completion time of 6 seconds per trial. The larger sample size and repeated measures design (which permits increased statistical power) seems to have allowed us to detect smaller-but-significant effects than have otherwise been reported. This is not necessarily a trivial finding regarding just a few arc seconds of difference between studies, since we found that the difference in performance between the 40 and the 80 arc min limit conditions, for example, is an average reduction in alignment error magnitudes of about one third. For depth-related manual tasks in which precision placements are absolutely critical, e.g., surgery or bomb disposal, it seems that providing viewers with disparities larger than 60 arc min could noticeably improve performance.

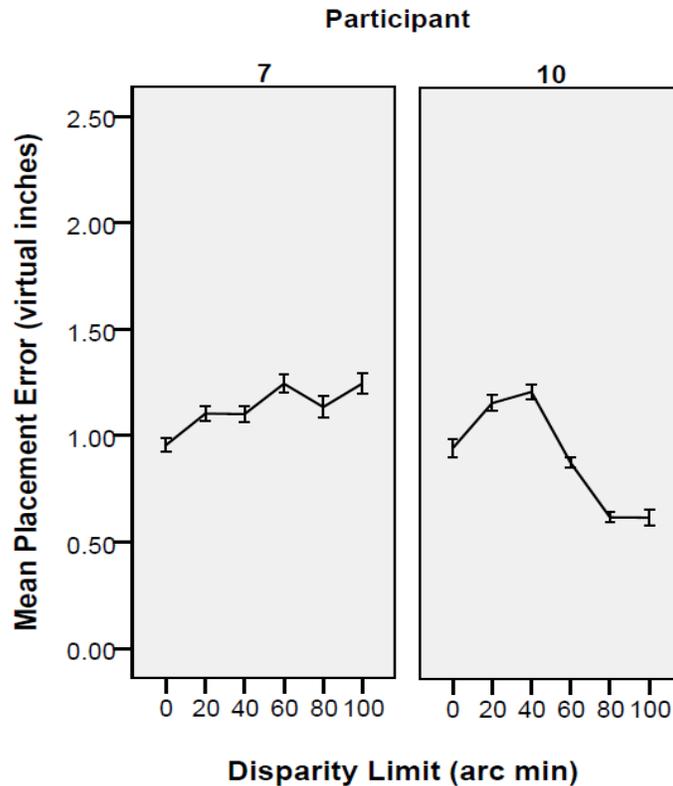
**Individual Performance Results.** Despite the apparently large benefit to performance of enlarging disparity limits on S3D displays, there were some individual differences in performance across the limits that are worth exploring. Figures 4-2 and 4-3 show the effect of disparity limits on placement errors, for each of the 12 participants. There are two sub-groups within our data which we classified based upon the benefit they received from S3D cues. Table 4 lists the relative performance of the 12 participants, and shows that Participants 7 and 10 clearly have atypical performance when using S3D cues (getting no overall benefit, or a detriment from S3D).

The results from the ‘typical’ subgroup show what we might expect to see if the average effect of disparity on performance held across individuals, in which increasing disparity helped depth task performance (see Figure 4-2). This subgroup demonstrates a huge benefit with stereo 3D, ranging from 30% to 91% reductions in placement errors when provided with S3D cues. Some participants’ performance either peaked or plateaued at 60 or 80 arc min of disparity, suggesting no further benefit from enlarging disparities beyond these levels. Interestingly, two participants in this subgroup never peaked {4 and 9}, and might have continued to benefit from even larger increases in disparity beyond the range we tested (up to 100 arc min or 1.67 degrees). We would expect all participants’ performance curves, if given a large enough range of test disparities, to eventually ‘peak’ due to reaching and then exceeding their own limits for binocular fusion which are individual specific. A fusion break results in diplopic and/or blurry vision, binocular rivalry, suppression, discomfort, and associated visual problems, which would cause performance to suffer.



**Figure 4-2.** The main effect of disparity limit on placement error, by participants.

These data represent a majority subgroup that clearly benefited from having S3D cues. Error bars represent +/- 1 SEM.



**Figure 4-3.** The main effect of disparity limit on placement error, by the minority subgroup of participants whose benefit from S3D cues was overall neutral (Participant 10) or negative (Participant 7). Error bars represent +/- 1 SEM.

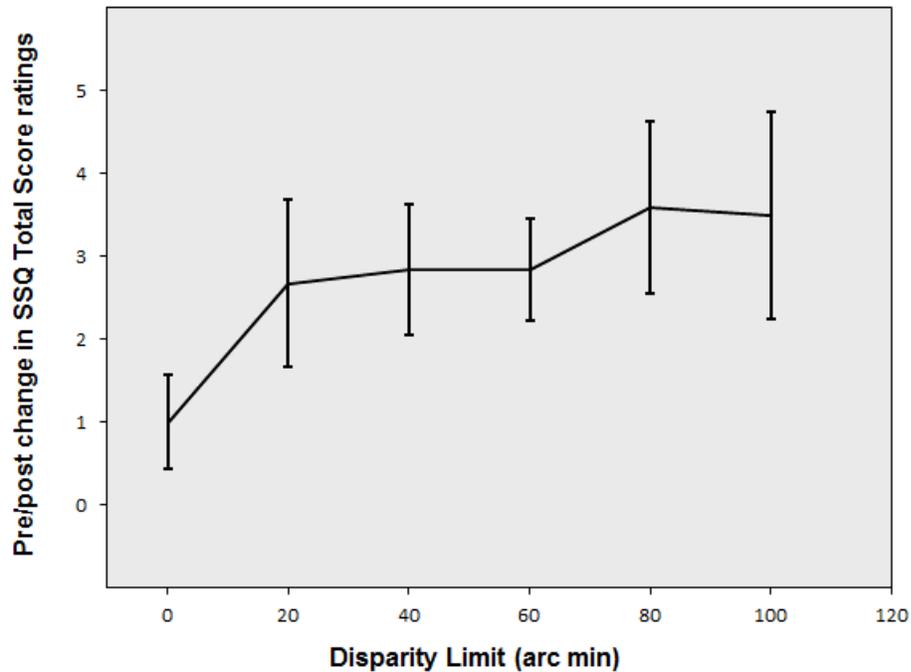
The two participants that demonstrated an obviously atypical pattern of unexpected results {7 and 10} are shown in Figure 4-3. Participant 10 *on average* received no benefit from the S3D cues; but interestingly, performance was helped with the large (80 and 100 arc min) disparity limits, whereas performance was *worse than baseline* with the small (20 and 40 arc min) disparity limits. Participant 7 received absolutely no benefit from S3D, and in fact *appeared to get worse for any disparity levels above zero*. The presentation order of the sessions were randomized across participants to minimize any possible practice or training effects, and the session orders did not appear to coincide in any way with the atypical results (Participant 7

showed no practice effects across sessions, and Participant 10 actually seemed to get less accurate across sessions). These results seemed very odd and unexpected, and this is a topic we will return to after looking at some further results.

### **Disparity Limits and Sickness, Discomfort, Eye Strain, and Balance**

**Simulator Sickness Questionnaire.** We found somewhat surprisingly that the disparity limit did not have a significant effect on the Simulator Sickness Questionnaire Total Score pre/post change, using a repeated measures ANOVA [ $F(5,55)=1.127, p=.357$ ], nor on the SSQ's Oculomotor subscale [ $F(5,55)=1.314, p=.271$ ], as one might have expected. Both of these results fail to support the Cue Conflict Theory of simulator sickness as applied to S3D displays. But there was a trend in the SSQ measures in which discomfort increased with larger disparities, which seems to support Cue Conflict Theory, and these will be discussed next.

There were significant partial correlations between the disparity limits and the pre/post changes in SSQ scores, indicating that increases in disparity limits resulted in higher sickness ratings changes (see Figure 4-4). The disparity limit was positively correlated with the SSQ Oculomotor subscale, the SSQ Disorientation subscale, and the SSQ Total score (see results in Table 5).



**Figure 4-4.** The main effect of disparity limit on pre/post session changes in the SSQ Total Score ratings. Error bars represent +/- 1 SEM.

Also, a paired-sample one-tailed *t*-test was conducted between the average SSQ changes in the S3D conditions (sessions with non-zero disparity limits) versus the ratings in the 2D condition (sessions with the zero disparity limit). The result suggests a significant difference between 2D and S3D in terms of discomfort ( $t=2.833, p=.008$ ); indeed, the SSQ changes in each of the S3D sessions were significantly different than the zero-disparity session (see bottom rows of Table 5).

These findings moderately support the utility of Cue Conflict Theory of simulator-sickness-type symptoms as applied to S3D displays. We found a positive relationship between increasing disparity limits and reported discomfort, as suggested by the significant correlations

and also by the significant 2D versus S3D *t*-tests, but the ANOVA tests failed to reach significance, possibly due to one or more of the following reasons.

There were large differences observed across participants. For example, Participant 10 averaged a pre/post change of zero units per session, while Participant 1 averaged a pre/post change of 6.2 units per session. Range restrictions may also be an issue, as we included only participants with good binocular vision under the age of 40, we limited disparities across all conditions to a maximum of 100 arc min, and we limited S3D exposure time to 30 minutes.

Another potential issue may be the generally small magnitudes of discomfort that appear to be induced by the S3D display (e.g., see Table 6). Of the 72 pre/post observations, 29% were changes in score of zero (or lower), and 49% were changes of less than two. Kennedy et al. (1993) warned that 40% to 75% of the SSQ rating scale items were likely to be zeros, and suggested that the more interesting values to experimenters are probably the non-zero items.

In looking at the non-zero rated items, we found that all pre/post changes in SSQ scores of five or larger (24% of observations) only occurred in the S3D sessions. It may also be interesting to note that the two participants {7 and 10} who received little or no benefit from the S3D cues in terms of performance also reported little or no discomfort induced by the display system (with average rating changes of +0.6 and +0.8 units, respectively). In any case, overall the results suggest only a small relationship between disparity limits and comfort, as the S3D display was fairly comfortable for most viewers, even when sometimes using disparity limits larger than the recommended One Degree Rule.

**Objective Measures of Eyestrain.** The six objective measures of eyestrain recorded both before and after each session included: lateral (horizontal) phorias at both near and far distances, vertical phoria at far distance, near fusion limit, far fusion limit, and the fusion range. Several of

the changes in objective eyestrain measures correlated with the subjective SSQ score changes, as intended.

Six of the 24 partial correlations between the objective and subjective measures were statistically significant (see Table 7). Two of the objective measures, the fusion near point and the fusion range, correlated with the SSQ Nausea subscale, the SSQ Disorientation subscale, and the SSQ Total score. Interestingly, none of the objective eyestrain measurement changes correlated with the Oculomotor subscale of the SSQ.

Also, only one of the objective eyestrain measures significantly correlated with the disparity limit manipulation: the pre/post change in the near lateral phoria measure. These data suggest that the objective measures of eyestrain were not as effective at capturing possible discomfort or physiological strain induced by the S3D display, although they may be related for some reason to nausea and disorientation-related discomfort.

These results do not necessarily imply that eyestrain, fatigue, or viewer discomfort was not objectively occurring, or is not a concern for stereoscopic 3D displays. Longer viewing sessions with larger disparities than were tested in this work are certainly conceivable and probably common: think of feature-length 3D movies which often utilize huge disparities to magnify the 3D experience; or 3D gamers who might spend hours at a time parked in front of a stereo display. The subjective measures of eyestrain (see the SSQ section above) and self-reports by the participants suggest that at least some eye-related discomfort was induced by stereo 3D in this experiment, especially for the larger disparities and for some individuals. But the objective measures of eyestrain tested in this work *generally* do not appear to be strongly sensitive to the disparity limits tested (up to 100 arc min) and/or for viewing durations of 30 minutes or less.

**Postural Instability.** Changes in postural stability before and after each session were measured as a possible alternative objective indicator of simulator sickness, disorientation, fatigue, and/or discomfort. The prediction was that virtual environment-related sickness or discomfort issues would be related to changes in postural stability. But changes in the average center-of-pressure velocity did not significantly correlate with self-reported discomfort as measured by the SSQ, or with the disparity limit manipulation on the S3D display (see Table 7). These results suggest that the Postural Instability Theory of simulator sickness may not apply well to understanding discomfort on S3D displays, since changes in balance do not appear to be related to subjective reports of discomfort/sickness, nor to the disparity limit manipulation.

### **Predicting Performance with Optometric Measures and Pre-Session Measures**

An optometric screening process, conducted by a USAFSAM professional optometrist, was used to ensure all participants had normal binocular vision. But the collected data was also used to investigate possible relationships with performance on S3D displays. These results and discussion should be prefaced with the acknowledgment that with only 12 participants in Experiment One, our statistical findings by themselves should be considered carefully. In any case, some of the patterns of results still seem theoretically interesting and suggestive.

We collected the six optometric screening measures (and several measures derived from these) for each participant, and the five pre-session optometric measures averaged for each participant. Then these measurements were correlated with the average S3D placement error for each individual, which was the mean performance of all trials in which any magnitude of S3D cues were present.

For the six clinical screening measures and their derived measures, only one was significantly correlated with performance (see Table 8): near fusion range ( $r=-.51, p=.045$ ). The

fusion range measure was derived by averaging the break and recovery points for both base-in prism and base-out prism, and adding the magnitudes together to give a functional range for fusion, in units of prism diopter. Participants with a larger near fusion range tended to have smaller errors in the S3D placement task. Conceptually, this finding suggests that viewers with larger fusion ranges for near-focused stimuli were able to properly fuse the entire range of displayed disparities, even the large disparities that might be uncomfortable (or impossible) for others to view, when using a desktop stereo system.

For the five pre-session optometric measures, two were significantly correlated with performance in the suspected direction (see Table 9). The two significant correlations were fusion near point ( $r=-.50$ ,  $p=.049$ ) and fusion range ( $r=-.60$ ,  $p=.020$ ). These pre-session findings were consistent with the pre-experimental screening results. We found viewers with closer fusion near point limits (i.e., convergence near points) and viewers with larger fusion ranges performed better on S3D displays. Again, these relationships suggest that some participants were better able to view the larger disparities on S3D displays before losing fusion, particularly for near stimuli requiring large convergent eye movements (and larger VA conflicts), and might help explain why some viewers gain such a large relative performance advantage from S3D cues when compared to others.

In conducting our follow-up testing of the atypical participants (as will be discussed in a later section), we attempted to measure stereoacuity thresholds for all participants. Two of our participants {5 and 7} had difficulty with the threshold measurements as conducted on the S3D display, but we were able to estimate their stereoacuties using the OVT and/or Randot stereotests. We correlated the twelve stereoacuity thresholds with placement accuracy in all of the S3D trials and found a strong significant correlation ( $r=.76$ ,  $p=.002$ ). If this finding holds

with larger samples and across different task types, it may provide an easy-to-administer clinical measure that is predictive of individual depth task performance on stereoscopic 3D displays.

Related previous research on stereoacuity in regards to S3D is sparse and somewhat conflicted. Hale and Stanney (2006) tested two groups in a S3D virtual environment on locomotion, object manipulation, and reaction time tasks. One group had “low” stereo acuity (higher than 80 arc sec) and the other group had “good” stereoacuity (80 arc sec or lower). The only notable difference between performance of the two groups was that the “good stereoacuity” group made more efficient movements during object manipulation, but the RT’s were comparable between groups.

Apart from performance on S3D displays, a variety of experiments confirm that stereoacuity plays a key role in performance on real-world depth tasks. For instance, O’Connor et al. (2010) showed that viewers with normal stereoacuity (60 to 250 arc sec or better, depending on the clinical test) generally performed better on pegboard, bead, and water-pouring tasks than those with reduced stereoacuity, and those with reduced stereoacuity often performed better than those with no measurable stereoacuity. Unfortunately, as in most studies, stereoacuity was not correlated with performance, and viewers with clinically “normal” stereopsis were simply compared to non-normal groups.

In a review of performance issues and the design of experiments testing stereoscopic 3D displays, Hsu et al. (1996) recommended the consideration of individual differences in stereoacuity, and speculated that “depending on the stereo perception task that is required of the subjects, *stereoacuity tests may or may not be a good predictor of task performance*” (p. 814) [emphasis added]. Our research may be the first to report that even for viewers with clinically normal stereopsis, there is a significant relationship between stereoacuity and performance on an

S3D precision placement task. Future research on the relationship between stereoacuity and S3D performance is warranted.

### **Predicting Discomfort with Optometric Measures and Pre-Session Measures**

In an attempt to potentially predict which viewers might find S3D displays particularly uncomfortable, we correlated the SSQ self-reported changes in discomfort in the S3D display conditions with the pre-session optometric measurements and the USAFSAM clinical measurements (including refractive errors, horizontal and vertical phorias, fusion near and far limits, fusion ranges, and stereoacuity thresholds). We found no statistically significant correlations between optometric measures and reported discomfort as induced by the stereo display. Additionally, as Postural Instability Theory specifically predicts that users with more initial postural instability should have higher levels of discomfort from simulator/virtual environment exposure, we tested this correlation and found it to be non-significant ( $r=-.24$ ,  $p=.774$ ). Further research on these topics using larger sample sizes and other optometric measurements may be warranted.

## V. EXPERIMENT ONE – FOLLOW-UP TESTING

Recall that two participants had atypical and unexpected performance data, in which S3D cues either helped very little overall or were actually detrimental to performance. This was despite the fact that all participants tested normally on the Titmus stereovision clinical test, the Keystone View Telebinocular stereovision test card, and passed a binocular and stereoscopic vision clinical screening by professional optometrists from the USAF School of Aerospace Medicine. This is a potentially disturbing finding for real-world applications of S3D because it suggests the possibility that some viewers with (apparently) perfectly normal binocular and stereoscopic vision, as tested in the lab or verified clinically, might fail to perceive 3D stimuli as intended and could result in performance comparable to 2D (with no benefit of S3D); or even worse, performance may be hampered by S3D. It is important that we at least try to determine the cause of this issue for future researchers, and for a more complete understanding of how the human visual system interfaces with S3D technology applications.

Perusal of the demographic and personal history questionnaires, involving questions of history of viewing 3D movies/games/TV, age, gender, inter-pupillary distance, experience of migraines, motion sickness, etc. revealed no obvious explanations as to why these two participants would have little or no benefit from S3D. Recall that some of the pretrial optometric measures and some of the USAFSAM optometric screening measures were generally correlated with performance; but the two individuals were not noticeable outliers in any of these measurements.

The two atypical participants reported being able to see “3D” and “depth” though perhaps not a strong sensation of it. For instance, Participant 10 noted in a follow-up conversation: “Now that I think about it, when starting a new session I never thought to myself, ‘Oh this is definitely 2D or 3D.’ I only knew for sure it was 3D when I saw double.” This reference to double vision

could have referred either to diplopia due to loss of fusion from large disparities, or it could refer to the subtle interocular crosstalk inherent in the S3D display system in which images intended for one eye “bleed through” to the unintended eye, which perceptually can result in a dim doubling or tripling of the stimulus, even when fusion is occurring as intended.

However, mild magnitudes of crosstalk do not typically have major implications for comfort (e.g., Kooi & Toet, 2004) nor would we expect major implications for performance, unless the crosstalk interfered with the fusional process (by creating false binocular matches) or perhaps by serving as an inadvertent cue to position in depth (by allowing a viewer to align the ghost images of the target with the ghost images of the control object). If this latter explanation were correct, though, we would expect performance to have improved with the larger disparities since the ghost image separations would be more obvious and allow for more precise alignment (this might explain Participant 10, but not 7).

The 3D shutter glasses were verified as working normally. It might also be worth noting that these participants had generally fast response times, though they were not the fastest, and had relatively good performance in the zero-disparity (no stereopsis) condition, though they were not the best. The participants both denied having inadvertently closed one eye due to any visual discomfort induced by the S3D display itself, or due to the glasses. In fact, there was almost no change from pre-to-post in their SSQ ratings, indicating little or no discomfort was induced by the display; indeed we might expect viewers who are not perceiving the S3D effect to find the display perfectly comfortable since there would be no VA conflict contributing to fatigue/eyestrain.

### **Possible Explanations**

The two clearly atypical results seem to be consistent with at least three non-exclusive possibilities: (1) these participants may have traded-off accuracy for speed in an attempt to “rush through” the study (even though it was time-limited to 30 minutes regardless of speed) or due to misunderstanding the instruction to focus on accuracy as the primary measure. This seems unlikely from an examination of their response time data. (2) Participants experienced intermittent stereopsis with unintentional monocular suppression (and didn’t notice), which could have possibly been caused by the unnatural viewing situations introduced by S3D displays (e.g., possibly the VA conflict or some related perceptual/cognitive conflict). Or (3) participants simply ignored or weakly weighted the S3D cues (either on purpose or inadvertently) and primarily or exclusively relied on monocular depth cues like size and/or texture to perform the task.

Again, as far as we could tell, these observers had “normal” binocular and stereoscopic vision, and should have been able to use the disparity cues to vastly improve their placement accuracy, as the ten other participants did. Other experimenters may have noticed a similar phenomenon. Froner (2011) observed that some participants had lower task performance than might be expected based on their pre-screening vision tests, which confirmed 20/20 acuity in each eye individually and binocularly, and stereoscopic acuity of 40 arc sec or better; the author speculated that focus-vergence issues may have been a contributing factor (p. 193). Also, Hoffman, Girshick, Akeley, and Banks (2008) noted in their Discussion section (p. 21):

*Many viewers cannot fuse a binocular stimulus with a vergence-focal conflict....We presented our stimuli to 11 young subjects. Three could not fuse most of the stimuli. Another two could fuse the stimuli but complained that doing so was too fatiguing. Only six subjects could fuse all the stimuli without significant fatigue and discomfort....All of*

*the disqualified subjects had normal binocular vision [verified via the Titmus Test] and could also fuse binocular stimuli in the natural environment across a wide range of distances. They only had problems with the experimental stimuli in which vergence-focal conflicts were present.*

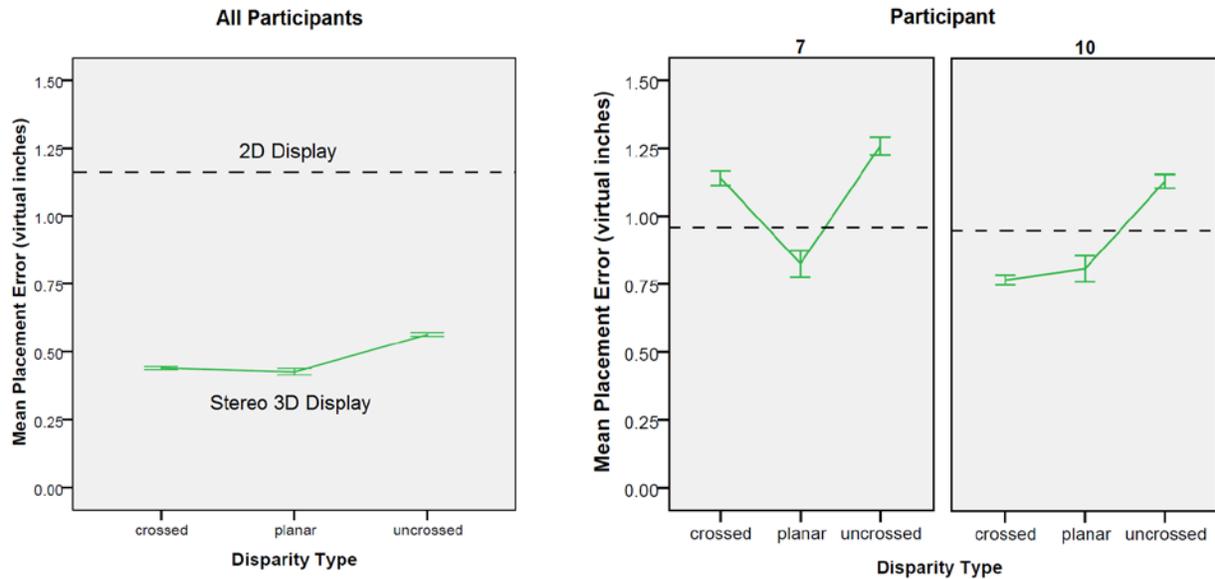
Our results seem consistent with Froner (2011) and Hoffman et al.'s (2008) findings that some viewers, even with normal stereopsis verified clinically, may have problems viewing S3D stimuli, presumably due to the VA conflict. This would certainly mesh well with our interpretation of Cue Conflict Theory as applied to comfort and even performance on S3D displays. But the results of Participant 10 may not be fully explained since this participant was helped with the larger disparities (80 and 100 arc min limit conditions), in which conflicts are larger, while only the smaller disparities degraded performance.

### **Anomalous Stereopsis**

Could this finding be in some way related to the concept of *anomalous stereopsis*, as described by Patterson (2009), who cited Richards (1971)? Richards found that in a sample of 75 observers using several depth judgment tasks, there seemed to be different types of “stereo anomalous” viewers who failed to detect either crossed disparities, uncrossed disparities, or confused the two. On the depth-matching portion of the experiment, Richards only found four out of fourteen observers (29%) to be apparently “normal” while the remaining had stereo-anomalies. He argued that three neuronal “pools” of specialized disparity-tuned neurons (corresponding to crossed, uncrossed, and near-zero disparities) might explain his results; if any one of the three pools is non-functioning or behaving problematically, biased responses could appear in the form described by Richards.

The idea of three (or four) distinct pools of disparity receptors seems to have been more recently abandoned in favor of continuous disparity mechanisms (Blake & Wilson, 2011). But even if this idea were true, it is not clear that Richards' results would have implications for the present study anyway. Richards resorted to using briefly flashed stimuli (80 ms) in order to detect performance lapses in gross depth judgments, while the present study used a more complex precision placement task with virtual objects in which trials were not time-limited (average trial time was 6-8 seconds). In our task, vergence eye movements had plenty of time in which to occur, and should have theoretically allowed even stereo-anomalous viewers to use whatever functioning pool of disparity neurons they possessed to perform the task, so they should still have performed better with stereo 3D cues than without.

Interestingly enough, we found that the type of disparity *did* seem to matter (crossed, uncrossed, or planar), at least for our two anomalous observers. Note that here 'type of disparity' refers to the direction of binocular disparity of the target object on a given trial relative to the display surface. The pattern of effect was different between these two individuals, and both were distinctly different than any of the other individuals' patterns which all appeared very similar to each other (see Figure 4-5). In general, participants received a large beneficial effect of S3D over 2D, regardless of disparity type; and uncrossed disparity was usually slightly less accurate than crossed or planar disparities.



**Figure 5-1.** The main effect of disparity type (crossed, uncrossed, or planar) on placement error, compared to their baseline performance in the 2D display condition. The left panel represents the average of all participants (including 7 and 10) which demonstrates the typical pattern. The panels on the right show the results for the two participants with an atypical pattern. The dashed line represents the corresponding performance in zero disparity conditions. Error bars represent +/- 1 SEM.

Notice the difference between the 2D and S3D display conditions for the group as a whole, versus only Participants 7 and 10. Participant 7 did seem to be slightly helped by S3D cues, but only if the virtual objects were portrayed at the surface of the display (planar); otherwise, crossed or uncrossed disparities hurt performance relative to the 2D condition. Note that “planar” here refers to the location of the virtual target object at the plane of the display surface. Perhaps even stranger, for Participant 10, S3D also slightly helped performance, but

only for crossed or planar disparities. It was uncrossed disparities which made performance worse than baseline 2D for Participant 10.

These data suggest that these two atypical participants were apparently seeing *some* stereoscopic depth from the 3D display, but only if the virtual object was near the surface of the screen where VA conflict cues would be very small. However, even if they were seeing some S3D, the beneficial effect was mild for both viewers, and transformed into a hindrance with either uncrossed disparity (Participant 10) or both crossed and uncrossed disparities (Participant 7).

### **Pseudo-Anomalous Stereopsis**

A recent study by Kihara, Fujisaki, Ohtsuka, Miyao, Shimamura, Arai, and Taniguchi (2013) may hold an explanation for our atypical observers. Kihara et al. (2013) reported the results of 134 participants on a depth rating task in which shading cues to depth and disparity cues to depth were utilized in various combinations. All participants were verified as having some stereopsis ability, with stereoacuties of 4 arc min (240 arc sec) or better on the Titmus stereotest. They discovered that in the young age groups (17-39 years old), a significant portion (11 to 21%) of the participants did not utilize the available stereopsis cues to perform their depth judgments, and instead seemed to rely on shading cues to perform the task. They described this subgroup as demonstrating “pseudo-anomalous stereopsis.”

There are a few notable differences between their work and ours. Kihara et al. (2013) used a subjective depth-rating task and a smaller disparity range (from zero to 8 arc min), while we used an objective performance measure of virtual object precision placement with a much larger disparity range (from zero to 100 arc min). Also, Kihara et al. (2013) included all participants with stereoacuties of 240 arc sec or better, but this would mean viewers only had to

get two or three items correct (out of nine) on the Titmus stereotest, and this seems to conflict with established norms. Fielder and Moseley (1996) report that stereoacuity thresholds of 30-40 arc seconds are regarded as “normal” in clinical practice, and point out that under ideal conditions, stereoacuity thresholds of 2-3 arc seconds are routinely observed in some viewers. Moreover, on the Wirt circles portion of the Titmus stereotest, participants with only one eye open can correctly guess the target down to 140 arc sec, indicating that non-stereoscopic cues may be present (Fricke & Siderov, 1997). In our Experiment One, all participants scored near-perfect on the Titmus stereotest (indicating thresholds of 40-50 arc sec or better).

Despite these differences and concerns, the results of Kihara et al. (2013) are potentially suggestive in the interpretation of our atypical viewers’ results. Kihara et al. (2013) hypothesized that their pseudo-anomalous observers, while being technically able to see stereo3D, seemed to use familiar pictorial cues to depth while ignoring disparity, at least for the gross depth judgment ratings they were performing. The authors also speculated that these pseudo-anomalous observers would abandon this strategy and utilize stereopsis cues when more precise judgments were required by the task.

### **Follow-Up Results and Discussion**

A brief description of the follow-up tests can be found in Appendix 3.

**Stereoacuity thresholds.** The stereoacuities of the atypical observers were the poorest of the group (in the range of 25-30 arc sec), but were still generally good and clinically normal (see Table 10). We had attempted to measure the stereoacuities of all participants using the USAFSAM custom software that uses an adaptive QUEST thresholding technique (e.g., see Watson & Pelli, 1983), and requires depth discrimination ability. However, two participants (which were also two of the three worst performers in S3D) had great difficulty in performing the threshold measure on the S3D display. They both reported difficulty in properly viewing the

stimuli, and described the stimuli as “having depth” but being perceptually unstable. On many trials they could not say for certain whether the depth of a target object was crossed or uncrossed relative to flanking objects, and that their perception of crossed versus uncrossed sometimes “flipped” if given enough viewing time. Both viewers were able to pass the clinical stereopsis test booklets in which odd-one-out depth detection (not discrimination) was required, such as the Titmus test (down to 40 arc sec) and Randot test (down to 20 arc sec). One participant’s stereoacuity was also estimated via USAFSAM optometrists using their OVT test.

This phenomena of “seeing depth” but being unable to identify its direction (crossed versus uncrossed) relative to other objects is consistent with Richards’ (1971) description of *anomalous stereopsis*, as discussed earlier, who made similar observations. Richards suggested the possible explanation that there were individual differences in neuronal pools that are specialized for disparity-detection and that correspond to crossed, uncrossed, and near-zero disparity pools. Research by Jones (1977) confirmed the existence of stereo-anomalies in a different sample of viewers, and suggested that such anomalies were limited to coarse, low resolution, high-disparity stereopsis. Jones also noted that such anomalies were present even in persons “with adequate binocular vision and normal fine stereopsis” (p. 621).

But, as mentioned, Blake and Wilson (2011) have pointed out that the idea of distinct neuronal pools for disparity processing mechanisms has been abandoned due to psychophysical and computation modeling results, gathered over the last 25 years. So while the accepted explanation has been called into question, our results provide confirmation of the existence of *some* type of anomalous stereopsis, in which viewers with apparently normal stereoscopic vision for real-world stimuli may have problems judging depth directions and/or magnitudes, yet still retain some psychological impression of “depth” that is occurring (thus allowing for good

performance on detection but not discrimination tasks). Our results suggest that anomalous stereopsis may not be limited to coarse, large-disparity situations, since the phenomenon occurred even with the thresholding task which is by its very nature involves fine stereopsis with low magnitudes of disparity.

**Other optometric and performance results.** For Participant 10, no monocular suppression was evident on the Randot suppression test, or when tested by the USAFSAM optometrists. When Participant 10 was forced to use binocular disparity cues on the modified placement task (forcing the use of S3D cues by making size cues unreliable), error magnitudes decreased in the S3D condition by 83% (from an average of 6.0 to 1.0 inches) and the standard deviation of errors shrank 75% (from 3.7 to 0.9 inches) while response times remained unchanged. The USAFSAM threshold measurement suggested that Participant 10's stereoacuity was about 26 arc sec.

Participant 7 also demonstrated no monocular suppression. When Participant 7 was forced to use binocular disparity cues for the modified placement task, error magnitudes *increased* in the S3D condition by 5% (from an average of 4.1 to 4.3 inches) and the standard deviation of errors *also increased* by 52% (from 2.9 to 4.4 inches), *and* response times were longer. It seems clear that this participant was simply unable to see stereoscopically on the S3D display as intended, and reported great difficulty in performing the requested tasks when taking the S3D USAFSAM threshold measurement and on the forced binocularity placement task. Participant 7's stereoacuity was instead estimated via the OVT stereopsis test, at around 25-30 arc sec.

In conclusion, in our attempts to determine the cause of the two atypical viewers' performance data with follow-up testing, we can say with confidence: (1) both observers had at

best stereoscopic acuities of around 25-30 arc sec, and were clinically classifiable as having “normal” stereopsis, so they should have been easily able to perceive and utilize the disparity cues used in the present study (up to 100 arc min); (2) neither observer indicated the occurrence of any sort of monocular suppression; (3) one participant was able to use disparity cues to improve performance when forced (Participant 10), while the other (Participant 7) apparently could not see stereoscopically on S3D displays either on the forced 3D task or on the USAFSAM stereoacuity task.

**Participant 10 Discussion.** This leaves us with two different probable (though not definitive) explanations. Participant 10’s atypical performance seems to be due to a different unconscious “strategy” or “depth cue weighting” being deployed in the performance of the S3D placement task. This participant appears to have been more heavily weighting the monocular pictorial cues of size and/or texture over the disparity cues to depth of the objects. Apparently, when the disparity cues were large enough (80 or 100 arc min), this seemed to allow disparity to finally help, perhaps by exceeding some threshold of activity in the disparity signaling mechanism that overtook the signaling of pictorial cues. This interpretation seems consistent with the results of Kihara et al. (2013), who demonstrated individual differences in how viewers utilize and combine various cues in a scene to determine depths of objects. If true, ours may be the first demonstration that individual differences in depth cue weightings effect not only depth judgments in a small disparity range (as in Kihara et al., 2013), but also performance on an active depth placement task across a large disparity range. And that such performance differences may manifest even in a sample of observers with clinically normal stereopsis, and with good (though not excellent) stereoacuity.

Kihara et al. (2013) speculated that their pseudo-anomalous observers would abandon their depth-cue-weighting strategy and utilize stereopsis cues when more precise judgments were required by the task (p. 502). However, our results suggest that Participant 10 *did not* abandon this strategy, even though precise judgments were required by our task (except perhaps when the relatively large disparities of 80 and 100 arc min were presented). Participant 10 only seemed to be definitely helped by S3D cues when the task was specifically designed to *force* the use of stereo cues by eliminating the reliability of the monocular pictorial cues (see Appendix 3).

It is possible that observers with pseudo-anomalous stereopsis may benefit from specialized training in the use of S3D displays. For instance, Fujisaki, Yamashita, Kihara, and Ohtsuka (2012) reported that many pseudo-stereoanomalous viewers benefited from specific stereo training to help estimate object depths, presumably by learning to more heavily weight the available disparity cues. Also, McKee & Taylor (2010) found two observers whose stereoacuity as measured by a stereoscope was many magnitudes worse than their stereoacuity for real objects. Through extensive practice, one of the observers was able to achieve performance comparable to real-world viewing, while the other improved somewhat with random dot stimuli but not with virtual objects in S3D, presumably due to cue conflict interference. Unfortunately, our work did not touch on this interesting possibility of improving anomalous viewers' use of S3D via training, but this seems an area ripe for future research.

**Participant 7 Discussion.** The explanation for Participant 7's results seems to be altogether different. Again, this participant tested "normal" on the clinical stereopsis measures, and had good though not excellent stereoacuity of 25-30 arc min. But whenever this participant was tested using S3D displays, perceptual difficulty was reported, and performance was poor. On the placement task, whenever *any level* of disparity was presented, performance declined relative

to the zero disparity condition. These results seem more consistent with the problems reported by Hoffman, et al. (2008), in which a large subset of participants had clinically “normal” stereovision for real-world stimuli across a wide range of distances and disparities but had “problems with the experimental stimuli in which vergence-focal conflicts were present” (p. 21). Our results support this interpretation of the VA conflict being the culprit. It is of note, too, that Participant 7 had small fusion ranges as measured by the USAFSAM optometrists and when measured on the S3D display (in both instances, this Participant scored the 2<sup>nd</sup> smallest fusion ranges out of the 12 participants). In light of these optometric tests and performance data on the placement task, the results suggest that Participant 7 may have had great difficulty in ‘breaking’ the reflexive neurological link between vergence and accommodation, thus making it difficult or perhaps impossible for this viewer to effectively view stereo imagery that requires vergence eye movements off the plane of the display (crossed or uncrossed disparities).

## VI. EXPERIMENT TWO – METHOD

### Participants

Two volunteers out of a possible 14 were excluded. One exclusion was due to a participant's deficient distance acuity (poorer than 20/20 Snellen), and the other participant dropped out half-way through data collection due to extreme eyestrain and visual discomfort caused by the larger disparity 3D sessions. A total of 12 participants passed the visual screening protocols and completed all data collection. None of the 12 participants in this experiment had previously participated in Experiment One. Ages ranged from 21 to 51 years old (average=38.7), all but one were right-handed, and there were 10 males and 2 females.

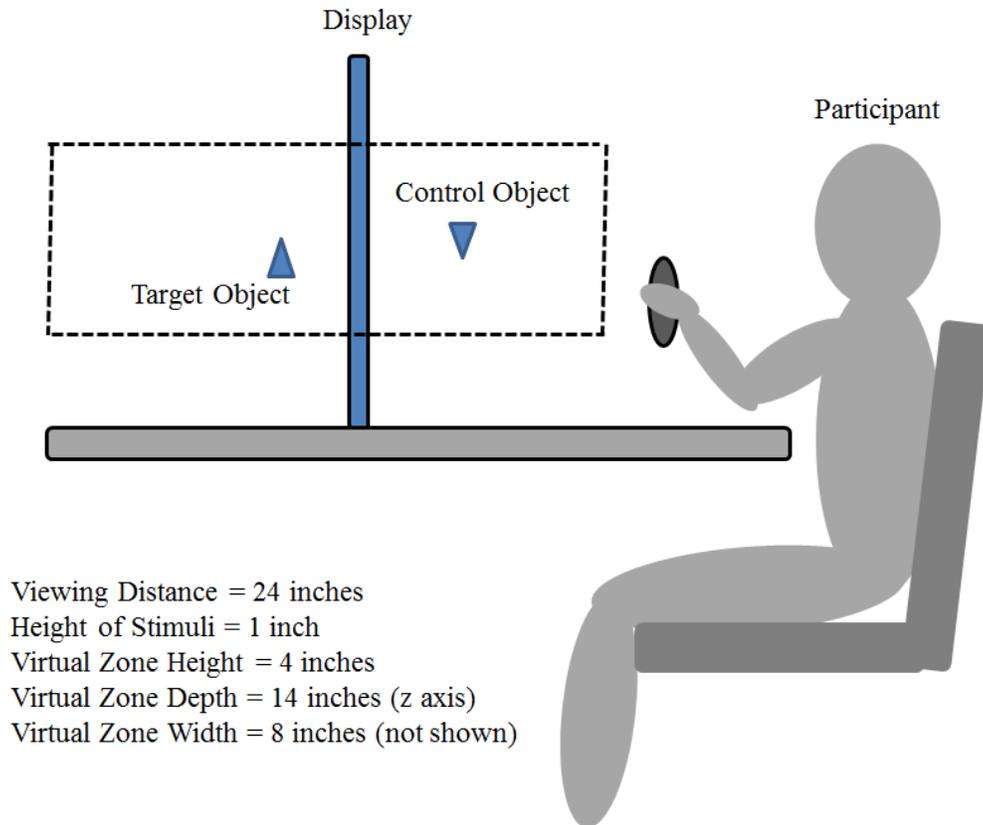
### 5 DOF Object Docking Task

This experiment was largely a replication and extension of Experiment One. The visual stimuli and experimental procedures were essentially identical to Experiment One, with the only notable change being that the task was not a 2 degree-of-freedom (DOF) virtual object positioning task but a 5 DOF virtual object positioning *and* orienting (i.e., docking) task.

For each trial, the participant used a spatial input device to position (3 DOF) and orient (2 DOF) the control peg so that the tip of the control peg touched the tip of the target peg while at the same spatial orientation or alignment in 3D space (i.e., “docking” the pegs). Participants signaled with a button press using their left hands when satisfied with their docking. Performance measures included completion times and accuracy (in both position and orientation). Accuracy (not completion time) was emphasized as the primary measure. Positional accuracy per trial was defined via *positional error*, equivalent to the absolute Euclidean difference between the actual placement and the optimal placement of the control object; or the distance from the (x,y,z) position of the control object to the (x,y,z) position of the target object (in virtual inches). Rotational accuracy per trial was defined via *rotational error*, equivalent to

the absolute angular difference between the actual angular alignment and the optimal angular alignment (in degrees), as determined by the appropriate cross-product multiplication.

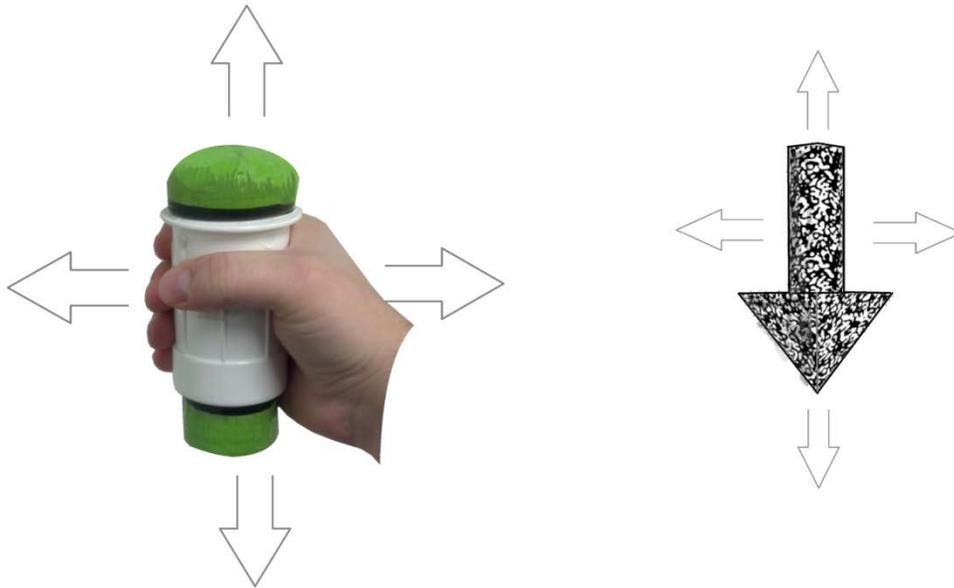
On each trial, the starting positions of the target peg were randomly chosen within a limited volume of virtual space (8 inches wide, 4 inches tall, 14 inches deep; see Figure 6-1). Possible orientations were limited to orientations within +/- 90 degrees left-to-right from vertical (rotation around the z-axis) and within +/- 45 degrees front-to-back from vertical (rotation around the x-axis). These limitations were utilized due to preliminary pilot testing problems of discomfort and armstrain, so that no trial required rotating the object upside down to complete the task; and the reason front-to-back rotation was limited was to ensure that the arrow-tip points for positional alignment would be visible or nearly-visible on all trials, avoiding visual occlusion problems. Movement of the control object was possible throughout all of x, y, z space, and any orientation through 3 DOF was possible (although only 2 DOF of orientation were considered for performance data, as matching the roll or 'twist' of the target was not required). The target object remained stationary at all times.



**Figure 6-1.** A side-view diagram of the experimental set-up.

**Input Device.** A high-accuracy, hand-held magnetic 6-degrees-of-freedom (6 DOF) spatial tracker was used in Experiment Two (Figure 6-2). This Tangible User Interface (TUI) was manipulated primarily by the participant’s right hand, augmented with the use of a keyboard or other signaling device by the left hand when necessary. The TUI used was an Ascension Technology© *Flock-of-Birds* magnetic mid-range tracker. The control sensor was embedded within a custom PVC handheld configuration and tethered by a thin cord emanating from the bottom (cord not shown in Figure 6-2). The specified static accuracy of the system is 0.07 inches RMS (positional) and 0.5 degrees RMS (orientation/angular) with a temporal resolution of up to 144 measurements/sec. The system allows for precision tracking up to 30 inches away from the

main magnetic device which was mounted immediately in front of the viewer on the desk and below the display; this created the magnetic field within which the hand-held sensor could be tracked.



**Figure 6-2.** The TUI spatial input device. The participants used a spatial TUI (left) to control the virtual object's x, y, and z position *and* orientation in the display volume (right). The TUI utilizes an embedded magnetic 6 DOF tracker (tether cord not shown).

## VII. EXPERIMENT TWO – RESULTS AND DISCUSSION

Full factorial repeated measures analysis of variance (ANOVA) tests were conducted when possible, in which participants were treated as random effects using Type III Sum of Squares (i.e., error terms for main effects used main effect x subject interactions). Reported correlations are either standard Pearson correlation coefficients, or when possible, partial correlations controlling for participant and session (practice) effects. Significance for correlations was tested using one-tailed *t*-tests when a specific direction of influence was suspected (otherwise two-tailed tests were utilized), and paired-sample *t*-tests were used where appropriate. Significance levels for all tests were  $\alpha = .05$ .

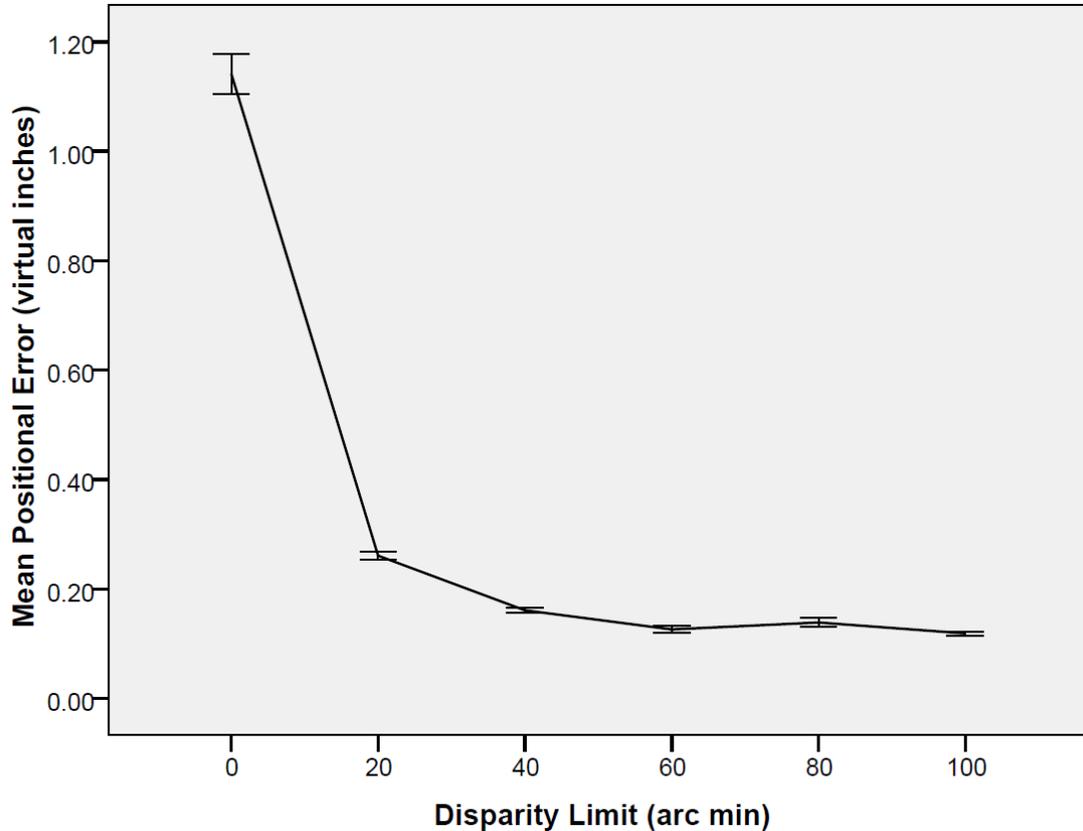
The raw data for this experiment consisted of 9024 total trials. The average trial took 13 seconds, so the average 30-minute session captured nearly 140 trials per participant. Outlier trials were identified and excluded if positional errors were  $\pm 6.0$  virtual inches or more, and/or rotational errors were  $\pm 45$  degrees or larger, with the assumption that these massive error magnitudes indicated inadvertent button presses or perhaps problems in the spatial tracking hardware/software. This excluded only 229 trials (2.5%), leaving 8795 total trials for statistical analysis.

### **Disparity Limits and Placement Error: Position and Orientation**

The primary results of this second study are that providing any level of S3D cues (non-zero disparity limits) greatly improved performance. In terms of positional accuracy, S3D improved performance over non-stereo by 86%. In terms of rotational accuracy, S3D improved performance over non-stereo by 29%. These data suggest that stereo cues can improve both positional and rotational performance, but may provide a larger benefit for positioning in x-y-z

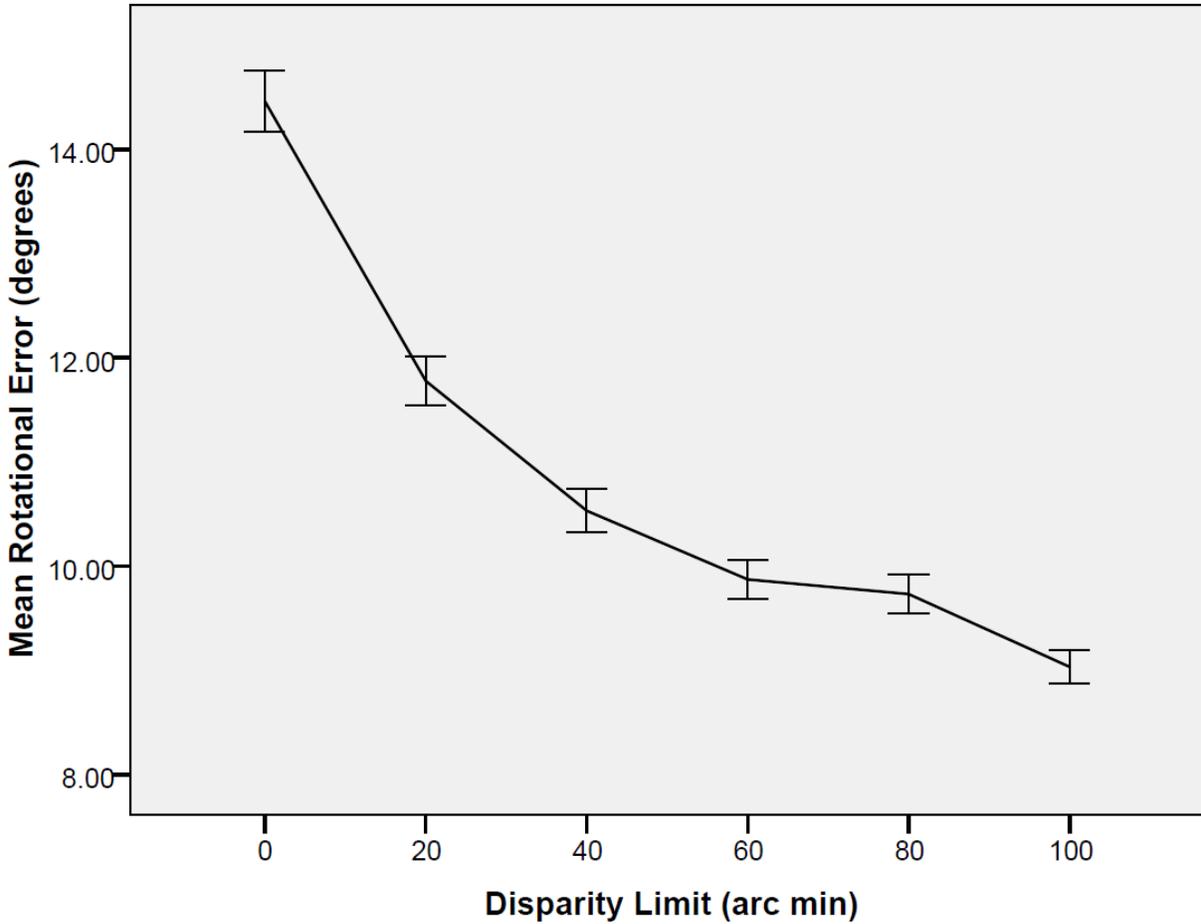
space, as opposed to orienting/rotating. We next explore the effect of manipulating the magnitude of the stereo cues on performance.

We found that the magnitude of the disparity limit (the range in which disparities were allowed to vary within a session) had a statistically significant effect on placement accuracy in terms of both positional error [ $F(5,55)=13.71$ ,  $p<.0001$ ] and also on rotational error [ $F(5,55)=9.35$ ,  $p<.0001$ ]. Figure 7-1 shows the main effect of disparity limits on positional error. For positional error, changing from no stereopsis cues (zero disparity) to only 20 arc min of disparity decreased error magnitudes by approximately 78%. Further small improvements seemed evident going from 20 to 40 to 60 arc min, but accuracy basically reached asymptote at 60 arc min of disparity limits and beyond. The average reduction in positional errors afforded by S3D cues of any magnitude (disparity limits from 20 to 100 arc min) was 86% (see averages in Table 14). These results are in general agreement with Experiment One, as well as with previous related research (see Introduction and Table 2; also performance reviews in McIntire, Havig, & Geiselman, 2012/2014): providing disparity cues generally improved performance in terms of placement accuracy.



**Figure 7-1.** The main effect of disparity limit on positional error. Error bars represent +/- 1 SEM.

Figure 7-2 shows the main effect of disparity limits on rotational error. For this measure, improvements in performance were not as dramatic but were still large. Changing from no stereopsis cues (zero disparity) to 20 arc min of disparity decreased errors by 17%, and unlike positional errors, further increases in the disparity limits provided continual improvements to rotational performance. At the maximum disparity limit of 100 arc min, participants had decreased their rotational error magnitudes by 38% compared to the zero disparity session. The average reduction in rotational errors afforded by S3D cues of any magnitude (disparity limits from 20 to 100 arc min) was 29% (see averages in Table 14).



**Figure 7-2.** The main effect of disparity limit on rotational error. Error bars represent +/- 1 SEM.

In terms of the One Degree Rule for S3D displays, positional accuracy did not improve with increases in disparity limits beyond 60 arc min (Figure 7-1). But rotational accuracy was best at 100 arc min limits and its trend suggests that performance may have even improved beyond these limits (Figure 7-2), had we tested even larger ranges. These results, and the results from Experiment One, contradict earlier experimental studies in which increases in disparities beyond about 25 to 50 arc min did not result in noticeable performance advantages (see Table 2). The previous results had suggested that smaller disparity magnitudes (under ~60 arc min) could

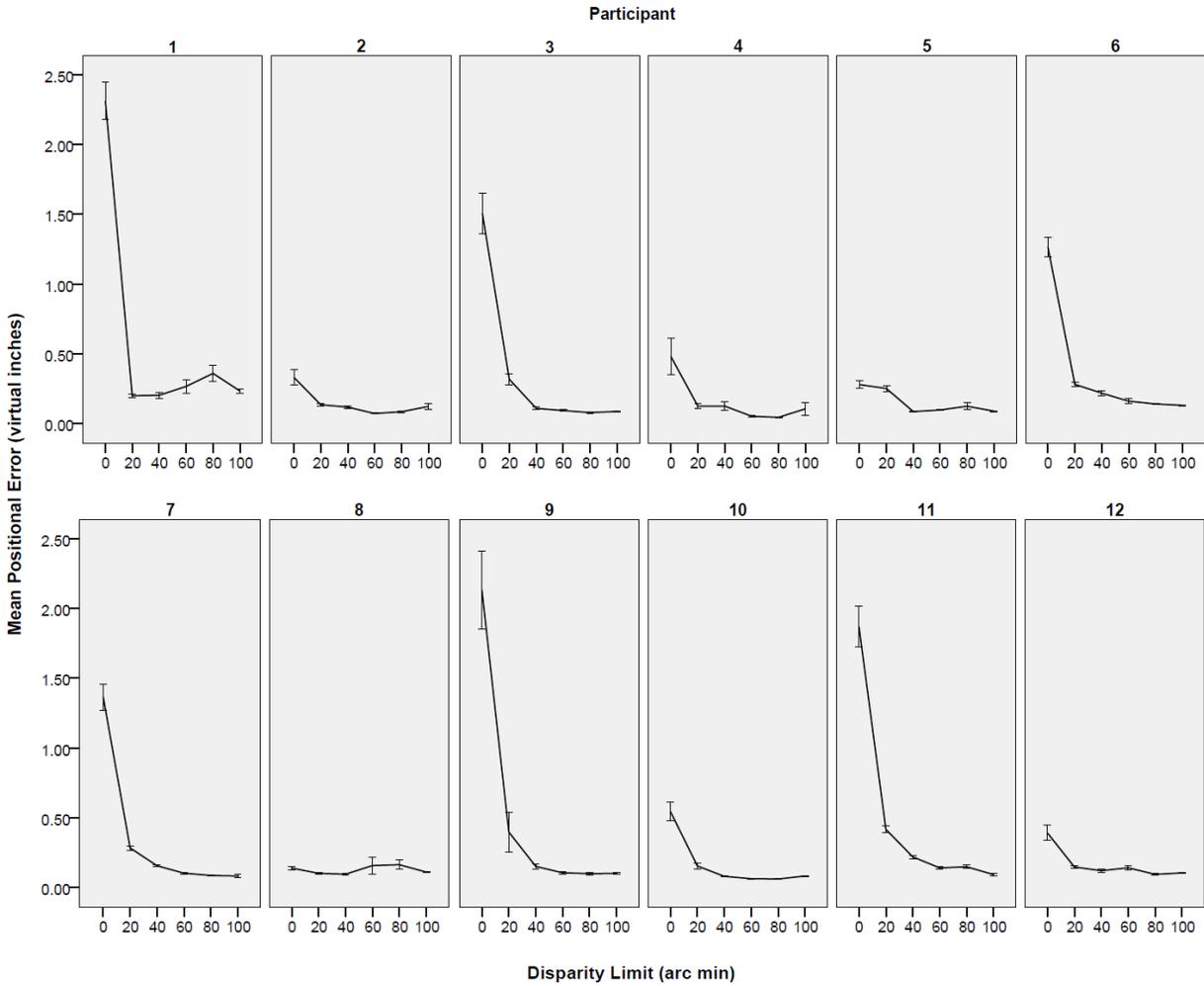
be just as useful (and more comfortable) than orthostereoscopic display setups in which larger disparity magnitudes were present. Our results instead show that large disparity magnitudes, provided by camera separations at or near orthostereoscopic levels, provide for the best performance, at least for tasks similar to our 30-min manual-spatial docking tasks.

The question naturally arises as to why performance may have “plateaued” in the 60-100 arc min range for positional error performance, but not rotational. There are a few possibilities which are not necessarily mutually exclusive. Positioning versus orienting might be thought of as separate tasks, which have separate measures (units) of error, and that just happen to be performed simultaneously due to this experimental design. Conceivably, participants could have traded these off by focusing on one over the other, and so benefited more or less from S3D on one task over another. Or perhaps one of these tasks is just “easier” than the other and differentially benefits from having S3D cues.

For the positional task, it is almost unnecessary that the bulk of the stimuli were even there, as the task simply required accurate spatial positioning of the tips (points) of the arrow objects, as closely as possible to one another. However, for the rotational task, the entire bulk of the stimulus body, for each arrow, could conceivably be useful in visually-determining the spatial alignment of the two objects; in this case, the larger the disparity cues available, the more the task might be improved. Some support for these ideas comes from the fact that going from no disparity cues (zero) to the 20 arc min limit improved positional performance by 78% but the same magnitude disparity increase only resulted in a 17% improvement in rotational performance. Clearly, the rotational task benefited differentially from S3D cues, and may have simply had more “room for improvement.”

## **Individual Performance Results**

Unlike Experiment One, here in Experiment Two there was not much in the way of individual inconsistencies in performance. Eleven of the twelve participants received a clear, large-magnitude benefit to performance from having stereo disparity cues (double-digit reductions in error on both performance measures), while one participant received only a small benefit consisting of single digit reductions from S3D (see summary in Table 14, and Figures 7-3 and 7-5). This single participant [Participant 8] received only a 7% improvement in positional performance from S3D, and only a 5% improvement in rotational performance, relative to no stereopsis cues. In comparison, the rest of the group's performance in S3D conditions was an average 81% reduction in positional errors and an average 30% reduction in rotational errors.



**Figure 7-3.** The effect of disparity limit on positional error, per participant. Error bars represent +/- 1 SEM.

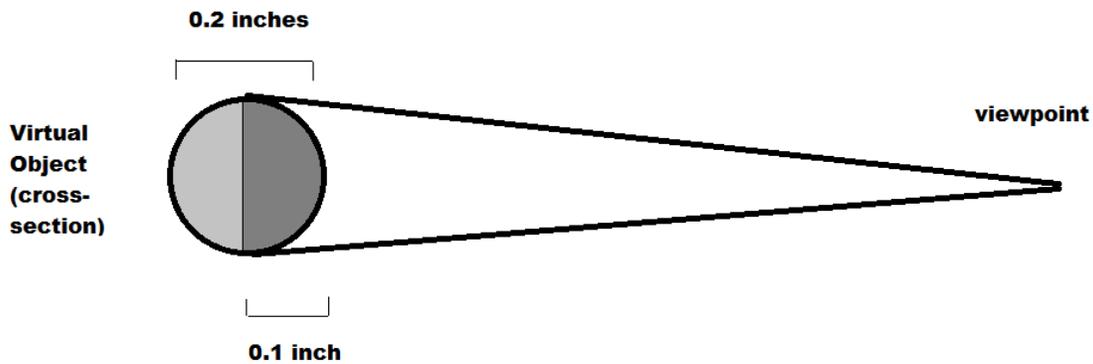
While Participant 8's performance patterns might suggest unique problems in viewing stereoscopic stimuli, as we had discovered for two participants in Experiment One, such explanations do not seem appropriate in this case. In fact, this participant had the best recorded stereoacuity threshold of 4.3 arc seconds (versus an average of 21.8 arc seconds for the rest of the group) and had extensive previous experience in working with and interacting with virtual 3D CAD models on non-stereo 2D display systems. In addition, this participant reported

practicing a “special strategy” which no other participants had adopted, and which involved manually positioning the virtual control object so that it was precisely superimposed over/within the target object, and then slowly “backing out” the control object’s position while retaining the precise orientation of the target. The consequences of this strategy were apparently lengthy trials (average RTs of 32 to 50 seconds/trial, versus an average of 13 seconds/trial for the group) and extremely precise performance even when no stereopsis cues were present (e.g., see Table 14). For instance, in the no-stereopsis session, Participant 8 achieved positional errors of only 0.14 virtual inches (versus an average of 1.13 for the rest of the group), and rotational errors of only 3.61 degrees (versus an average of 14.44 for the rest of the group). In summary, Participant 8 adopted a clever, unique, and unanticipated strategy and also had extensive 3D modeling experience on 2D displays which seems to have allowed for high performance using non-stereo cues, and so received only a small benefit to performance when S3D cues were provided.

**The benefit of S3D in the 20 arc min limit sessions.** Seven participants [3, 5, 6, 7, 9, 10, and 11] showed an interesting performance pattern in which they all received a large benefit from S3D cues in general, but seemed to have smaller relative benefits in the 20 arc min limit conditions than in the conditions with the larger limits (40 to 100 arc min). This might suggest that performance in the 20 arc min sessions was hampered by their stereoacuity limits, since this session presented disparities ranging from only -20 to +20 arc min. Indeed, these participants had relatively higher (poorer) stereoacuity thresholds, averaging 27.2 arc sec versus an average of 10.6 for the other five viewers. However, further analysis suggests that a stereoacuity threshold limitation *in regards to positional disparity of the virtual objects* was not responsible for these performance patterns, as 91% of the trials in the 20 arc min limited sessions presented positional disparities (relative to the display surface) of 4 arc min or larger, which were considerably above

these viewers' thresholds. Instead, additional analyses (provided below and in the section *Binocular Disparity and Performance: Positional Disparity versus Shape Disparity*) suggest that this smaller camera separation (at 20% of vIPD) resulted in extremely small binocular parallax cues to object shape, presumably harming fine relative positioning, and leading to suboptimal performances for these participants in this condition.

Evidence for this idea comes from the fact that the virtual objects' widths (the arrows' cylindrical bodies; see *Methods* section) spanned 0.2 inches in cross-sectional diameter. Any given viewpoint, then, allows at most 0.1 inch of an objects' shape in depth to be visible (see Figure 7-4 below). Assuming this virtual object is located in depth at the screen distance of 24 inches, with a camera separation of 20% of vIPD (0.52 inches), this provides a visible shape disparity (from the nearest point on the front object face, to its side edges) of 18.7 arc sec.



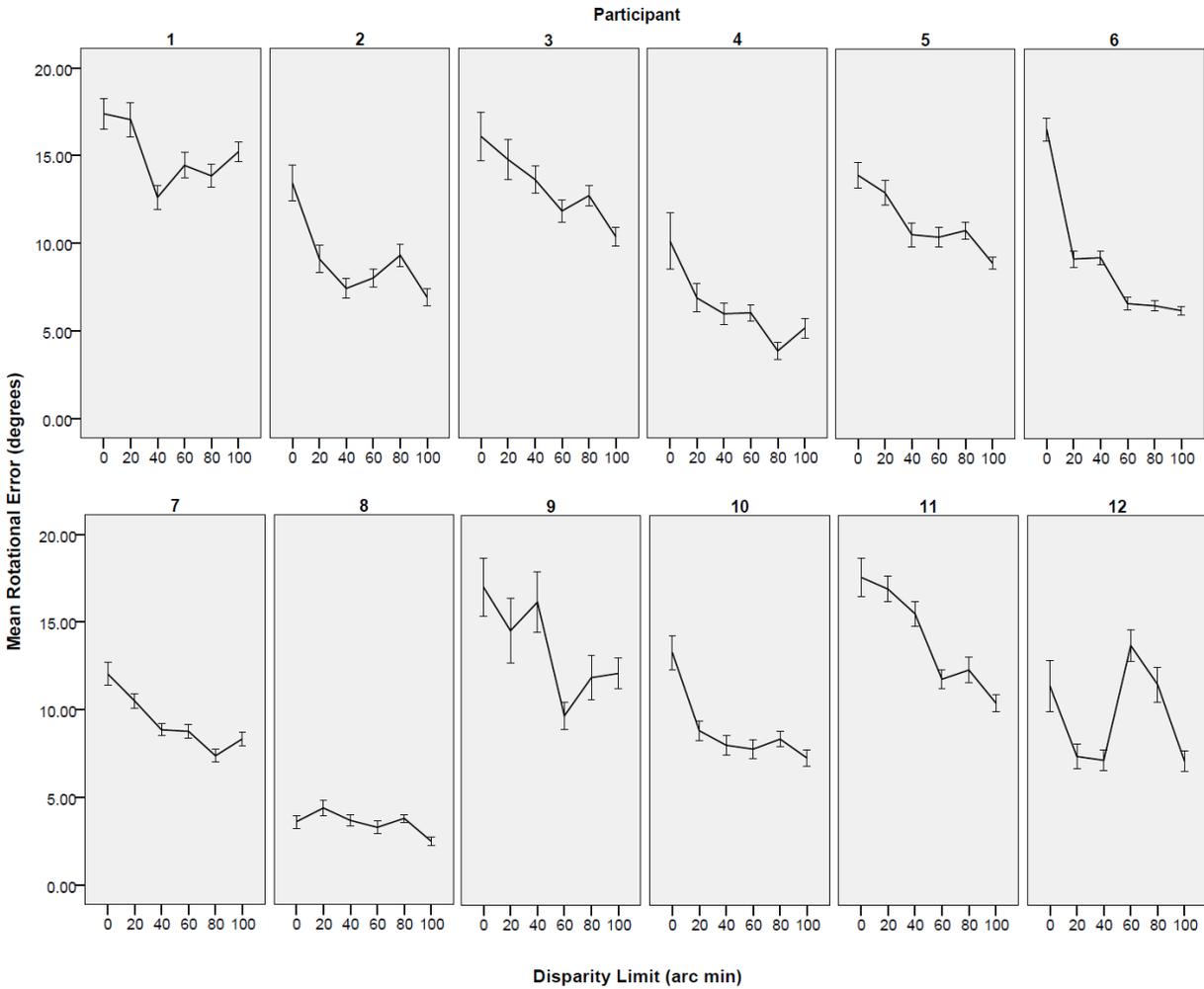
**Figure 7-4.** The virtual objects used in this study had a cross-sectional diameter of 0.2 inches, allowing any given viewpoint to see at most 0.1 inch of shape depth to be visible. The amount of shape disparity this results in, on a stereo 3D display, is also a function of the camera separation used to image this object.

Thus, it seems that these participants' atypical benefit from the S3D cues in the session with camera separations at 20% of virtual IPD (20 arc min limit) *was* likely due to their poorer stereoacuity thresholds being near or above the small relative surface disparities shown, calculated to be around 18.7 arc sec. In fact, six of the seven participants in this subgroup had stereoacuity thresholds of 18.0 arc sec or larger (again, this subgroup average was 27.2 arc sec). Although this possible explanation is elegant and fits very well with the experimental and optometric data, one participant's poorer S3D performance in the 20 arc min session does not seem explainable as a stereoacuity threshold limitation, as Participant 7 had excellent stereoacuity of 5.7 arc sec. Regardless of what the ultimate explanation may be, overall these data show that nearly all participants improved their positional performance when S3D cues were provided, and that optimal performances were generally achieved when disparity limits of 60 arc min or larger were utilized, equating to camera separations of 60 to 100% of virtual IPD.

**Individual Variability in Non-stereo versus S3D.** One additional instance of wide individual performance variability was in positional placement accuracy when no stereo cues were provided, versus when S3D cues were provided (first two data columns of Table 14). As discussed, when there were no stereo cues, Participant 8 demonstrated the smallest average errors of 0.14 virtual inches, but half the participants demonstrated average errors greater than 1 inch, with two participants even reaching average errors of 2 inches or larger. Clearly, when there were no stereo cues, the task was extremely difficult for some individuals, and resulted in very high variability across participants (and sometimes within participants, too). In contrast, when S3D cues of any magnitude were provided, there was little variability across or within participants: most everyone performed consistently very well in the higher disparity sessions

(participants' average positional errors ranged from a minimum of 0.09 inches to a maximum of 0.25 inches).

**Individual differences in rotational performance.** Individual differences in non-stereo viewing were considerably less pronounced in the rotational performance data (Figure 7-5). Some possibly strange patterns of individual results in the rotational data include Participants 1 and 12. However, these patterns are not alarming as they appear to be explainable from other data we collected. Participant 12's unexpected pattern seem explainable as order effects (training or practice), as this viewer experienced the higher disparity limit sessions as the first few sessions; in any case, this possible order effect only seemed obvious in the rotational but not positional performance data.



**Figure 7-5.** The effect of disparity limit on rotational error, per participant. Error bars represent +/- 1 SEM.

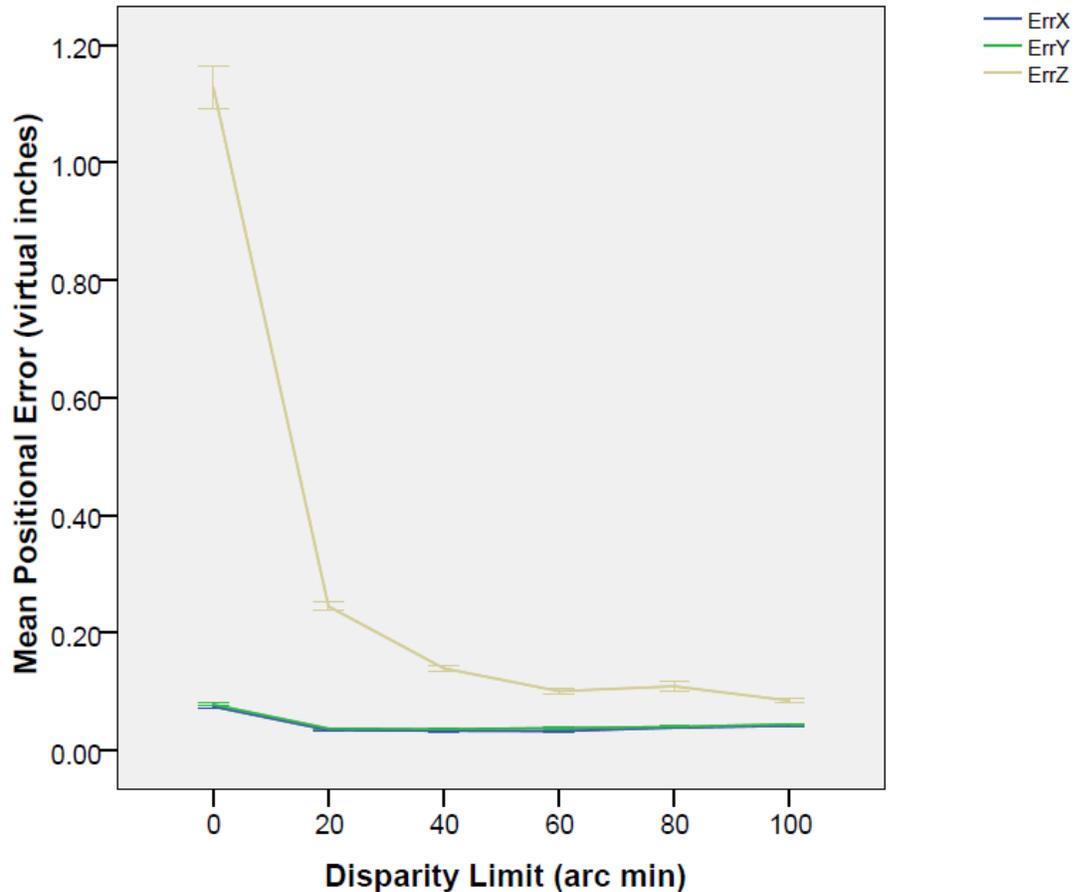
Participant 1's rotational *and* positional performance seems to reflect a preference for smaller magnitude S3D cues (as best performance occurred within the 20 to 40 arc min limited sessions). This speculation is strongly supported in Participant 1's optometric screening data, in which Participant 1 had the smallest clinically measured fusion range (at near) for the entire group: 24.5 prism diopters of total fusion range compared to an average of 34.0 prism diopters for the rest of the group. By further examining the clinical measurements that define total fusion

range, an outlier was evident: for fusion limits measured at a near distance, Participant 1 had a base-out (convergent) recovery point of only 2 prism diopters versus 16.2 prism diopters for the rest of the group. A measure of 2 prism diopters is equivalent to 1.04 degrees (or 62.4 min) of visual angle, suggesting this viewer had difficulty in achieving or maintaining binocular fusion for positional disparity magnitudes of 60 arc min or larger (for crossed disparities, which require convergent eye movements). These performance patterns and optometric screening patterns for Participant 1 imply that this viewer had a small effective binocular fusion range on S3D displays, particularly in the crossed direction, and might explain why this participant seemed to have the highest performance benefits from S3D within the 20 to 40 arc min disparity limits, and less of a benefit (but still a positive benefit) with disparity limits of 60 to 100 arc min.

### **Further Analysis of Performance Data**

Experiments One and Two were largely replication-and-extensions of Rosenberg (1993) who conducted a 2 DOF placement task almost identical to that used in Experiment One. As is common in this type of S3D research, Rosenberg had operationalized the positional performance measure as error only along the z-axis (in depth) and seems to have ignored (or not reported) the x or y-axis positional errors, while also ignoring orientation accuracy (which is not typically studied). Our primary positional error measure in both experiments was the absolute size of the Euclidean distance error in 3D space, but our raw data does allow us to break our positional errors into their respective x, y, and z error subcomponents for further investigation. Likewise, the combined rotational error measure can be subdivided via the raw data into rotational subcomponents around two axes (x and z).

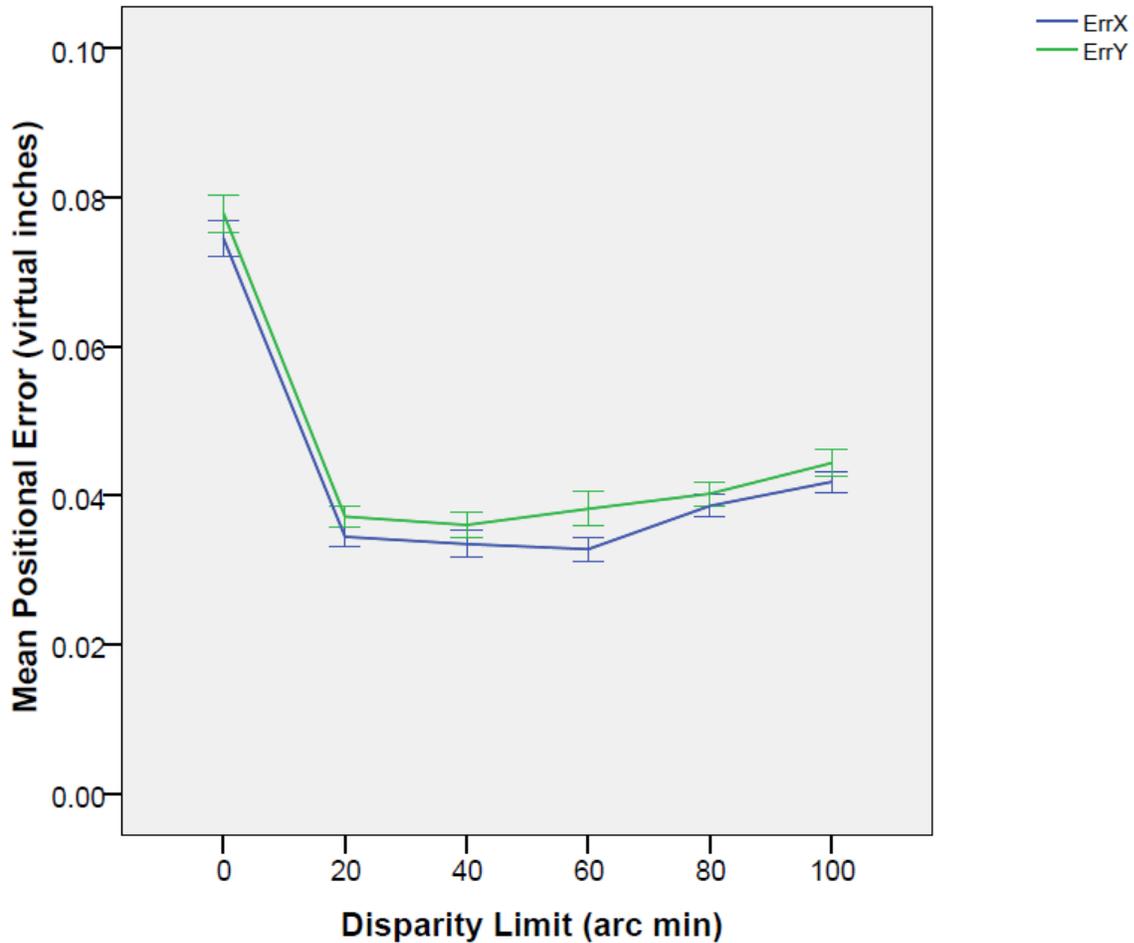
**Positional Error per Positional Dimension.** As expected, we found that our calculated Euclidean positional error was mostly a function of error in the z-dimension: the combined measure correlated almost perfectly with the positional error along the z-axis ( $r=.996$ ,  $p<.0001$ ) but much less so with the x or y-axis errors (correlations of .500 and .466, respectively). If we look at the magnitudes of these errors, plotted together, across the manipulation of disparity limits (Figure 7-6), a few observations are worth noting. First, with no stereo cues, positional errors in the x or y dimension were both very small (~0.10 virtual inches) while errors in the z dimension were an inch larger (~1.10 virtual inches). Second, providing stereo cues via disparity limits of 20 arc min or larger vastly improved positional performance along the z-axis, but provided relatively small benefits for x or y-axis errors, for which the two curves appeared nearly identical. This finding provides an expected confirmation that stereo 3D provides enhanced depth perception capabilities primarily in the z-dimension.



**Figure 7-6.** The effect of disparity limit on positional error, per spatial dimension (x, y, z). Error bars represent +/- 1 SEM.

It is, however, interesting to note that there was still a positive benefit of using S3D cues along the x and y-axis, as shown in Figure 7-7. Positional error magnitudes were basically halved when any level of S3D cues were provided (from ~0.08 inches to ~0.04 inches). This finding was surprising given that, at first blush, there is no reason to suspect that providing enhanced z-axis spatial information via stereo should *simultaneously* improve spatial performance in the x and y-dimensions by 50%. The reason for this finding, we believe, is the fact that orientation of the

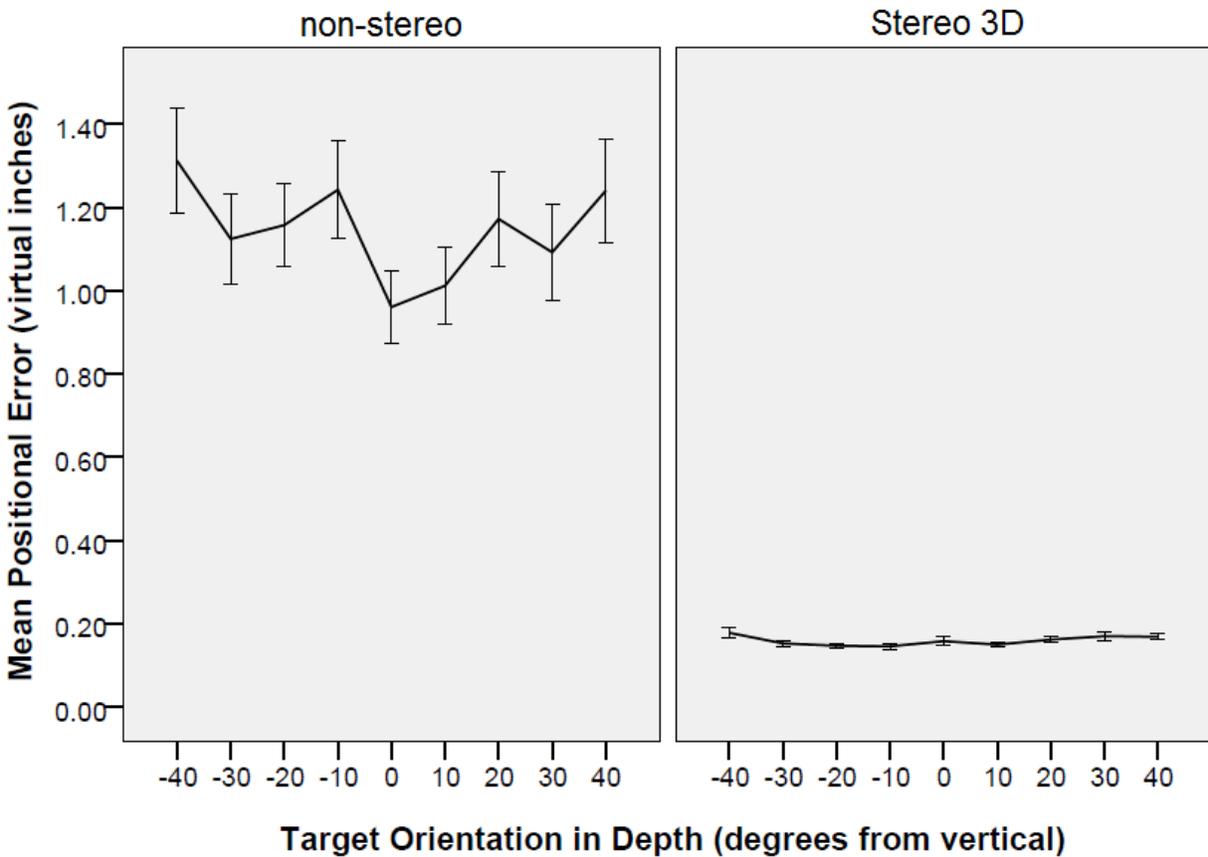
target object was manipulated across trials, in some cases degrading the quality of the spatial information visible in the x and y dimensions, as will be discussed next.



**Figure 7-7.** The effect of disparity limit on positional error, per spatial dimension (x, y). Error bars represent +/- 1 SEM.

**Positional Error by Target Orientation.** For particularly difficult orientations, in which the orientation of the target object was non-vertical and pointed toward or away from the viewer, slant-in-depth and/or perhaps occlusion issues could have hindered the ability for x-y information to be helpful about the relative positions of the object tips in space. Such situations would seemingly benefit from S3D depth cues regarding position. Evidence for this effect comes

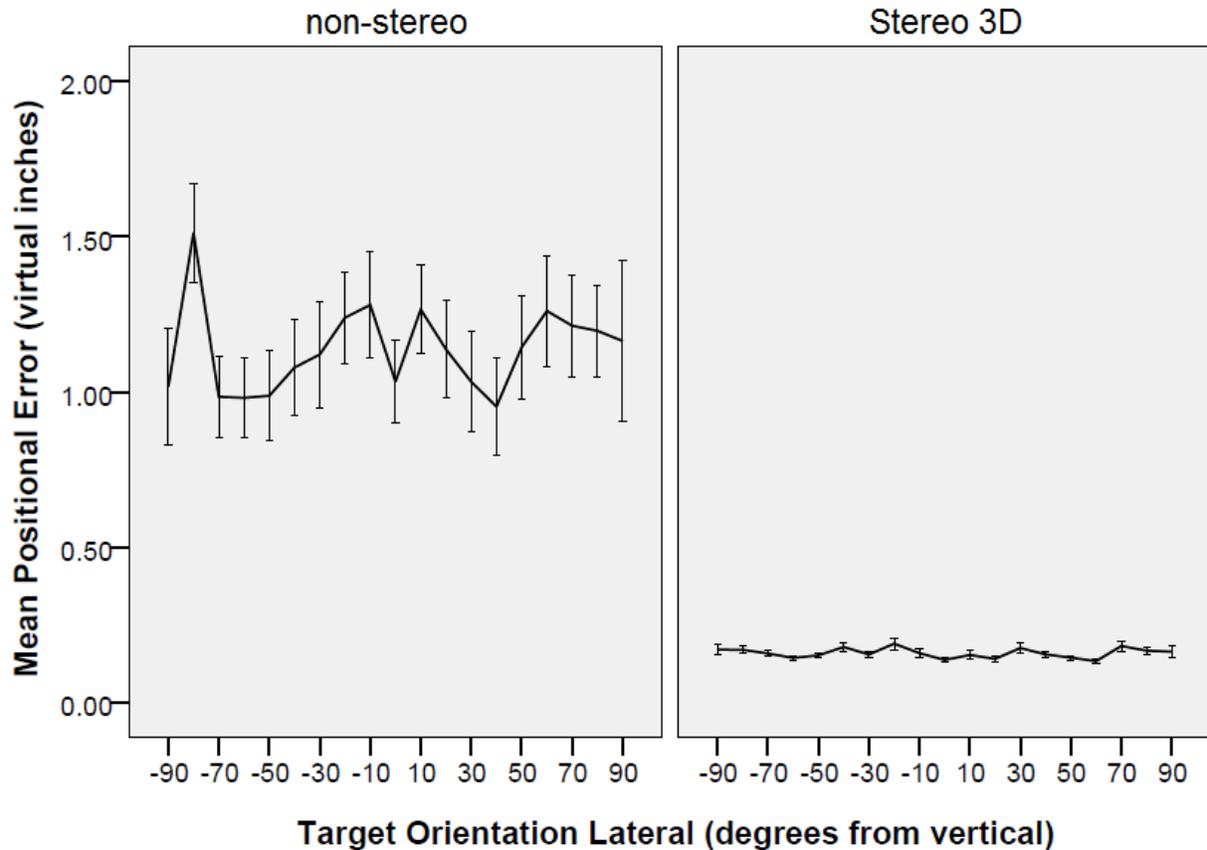
from the data as plotted according to the orientation of the target object (its angle in depth rotated about the x-axis); see Figure 7-8. Here we see that when no stereo cues are provided, errors in positioning tended to be highest when the target orientation was at the extremes of +/- 45 degrees rotation (forward/backward), and errors in positioning were smallest when the target possessed little or no orientation in depth. But the trend of target object orientation seems to disappear in the positional performance data once S3D cues of any magnitude were provided: positional error in x-y-z space was essentially unaffected by the orientation of the target object.



**Figure 7-8.** The effect of target orientation in depth on positional error, for trials without stereo cues (left panel) versus with S3D cues (right panel). Negative orientations indicate tilt of the target away from the viewer; positive orientations indicate tilt towards the viewer. Error bars represent +/- 1 SEM.

A somewhat related pattern emerges from target orientations that are left/right (lateral rotations from vertical ranging from -90 to +90 degrees): see Figure 7-9. When no stereo cues are given, positional performance *tended* to be best at the canonical orientations of 0, +/- 90 degrees, and also, surprisingly, at +/- 45 degrees (although the magnitude of these trends were very small). The only apparent large magnitude difference is that performance appeared substantially worse at around 80 degrees of leftward rotations from vertical than for most other orientations. This might also reflect difficulty due to awkward angle placement (far left-ward rotation simultaneously combined with potential rotations in depth) at non-canonical (but near canonical) orientations, perhaps indicating that hand orientation simply “snapped to” the assumption of the nearest canonical orientation when the target was approaching such orientations.

These orientation-specific performance trends are similar to the well-known *oblique effect* of stimulus orientation, observed throughout human and animal neurophysiological and performance data, in which a preference for canonically-oriented (horizontal or vertical) stimuli or objects is evident (e.g., Appelle, 1972; Westerheimer, 2003; Hermens & Gielen, 2003; Sasaki, Rajimehr, Kim, Ekstrom, Vanduffel, & Tootell, 2006; van Bergen, van Swieten, Williams, & Mon-Williams, 2007). Whatever the true reason might be, again, the trend of target orientation on positional accuracy disappeared once S3D cues of any magnitude were provided.

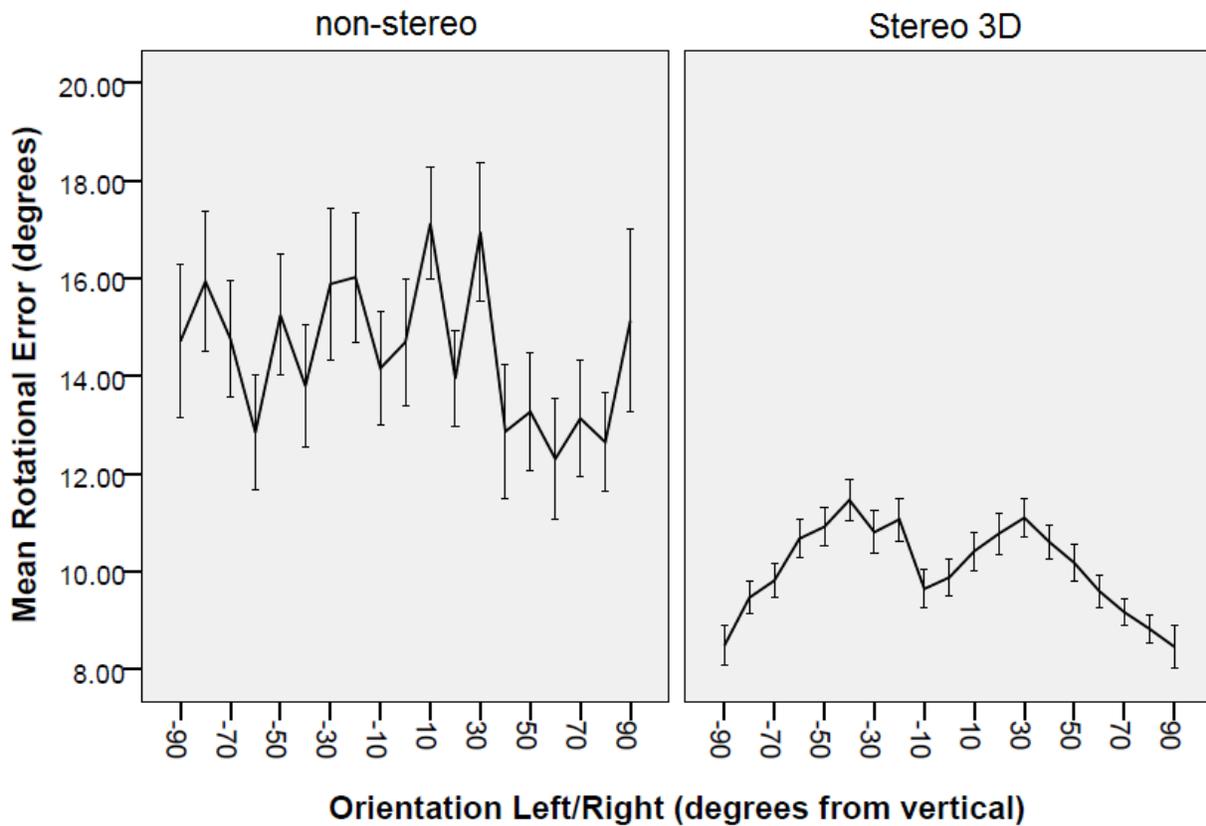


**Figure 7-9.** The effect of lateral target orientation on positional error, for trials without stereo cues (left panel) versus with S3D cues (right panel). Negative orientations indicate tilt of the target to the left (counter-clockwise); positive orientations indicate tilt towards the right (clockwise). Error bars represent +/- 1 SEM.

**Rotational Error per Rotational Dimension and Target Orientation.** Similar to positional error, it is important to verify our initial suspicion that S3D depth cues would help rotational performance more in the depth dimension. As expected, we found that the total combined rotational error was mostly a function of error in the depth dimension: the combined

measure correlated almost perfectly with rotational error in depth ( $r=.978, p<.0001$ ) but much less so with the lateral rotational error ( $r=.280, p<.01$ ).

When rotational accuracy is examined across different orientations of the target object, again some interesting trends seem to appear. When no S3D cues were provided, lateral orientations of the target had no clear, consistent trend in terms of performance. Once S3D cues were provided, however, rotational errors were smallest at the canonical orientations (-90, 0 and +90 degrees) and worst at intermediate orientations; see Figure 7-10.



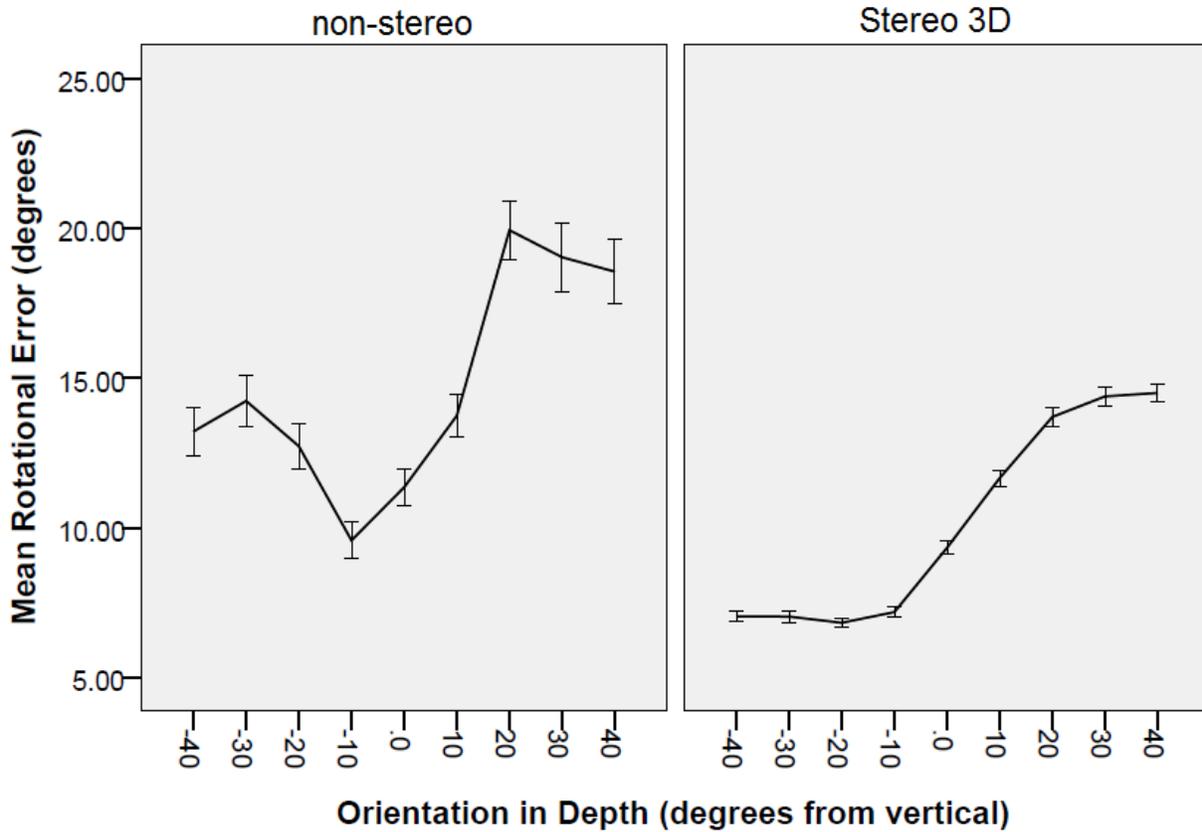
**Figure 7-10.** The effect of lateral target orientation on rotational error, for trials without stereo cues (left panel) versus with S3D cues (right panel). Negative orientations indicate tilt of the target to the left (counter-clockwise); positive

orientations indicate tilt towards the right (clockwise). Error bars represent +/- 1 SEM.

A more complex performance pattern emerges when target orientations in depth are considered (Figure 7-11). In the non-stereo case, we see that rotational performance was best when the target orientations were near zero; intermediate performance occurred when the target was oriented in depth away from the viewer; and poor performance occurred with orientations toward the viewer. In the S3D trials, orientation accuracy was best when the target was oriented at any angle less than zero up to -45 degrees (tilted away from the viewer), and this entire negative orientation range was flat in terms of performance. Accuracy systematically declined as the target object's orientation went in the opposite direction in depth, toward the viewer, from zero to +45 degrees. These patterns seem to suggest that in the 2D case, rotation in depth away from vertical *in either direction* harmed performance, particularly if oriented towards the viewer; but in the S3D trials, stereo provided for consistently good performance as long as orientations of the target did not tilt towards the viewer. This pattern is likely due to the fact that orientations of the target towards the viewer created situations in which the control object's nearer positioning occluded portions of the target object, apparently hampering rotational performance in both non-stereo and S3D trials.

The trends in these rotational data support the idea that *both positional and rotational* performance is best at (a) canonical orientations of the target object; also when (b) the target object is at orientations in which visual occlusion by the control object is not likely to be an interfering issue; and/or (c) the target orientations are not too "awkward" to require extensive

rotations, uncomfortable, or unnatural postures by the users' spatially manipulating the hand-held device.



**Figure 7-11.** The effect of target orientation in depth on rotational error, for trials without stereo cues (left panel) versus with S3D cues (right panel). Negative orientations indicate tilt of the target away from the viewer; positive orientations indicate tilt towards the viewer. Error bars represent +/- 1 SEM.

### **Binocular Disparity and Performance: Positional Disparity versus Shape Disparity.**

The location of a virtual object in 3D space, in conjunction with display viewing distance and viewer eye separation (or camera separation), are the factors that determine the ultimate binocular disparity that is presented to a viewer of an S3D system. It is this positional disparity

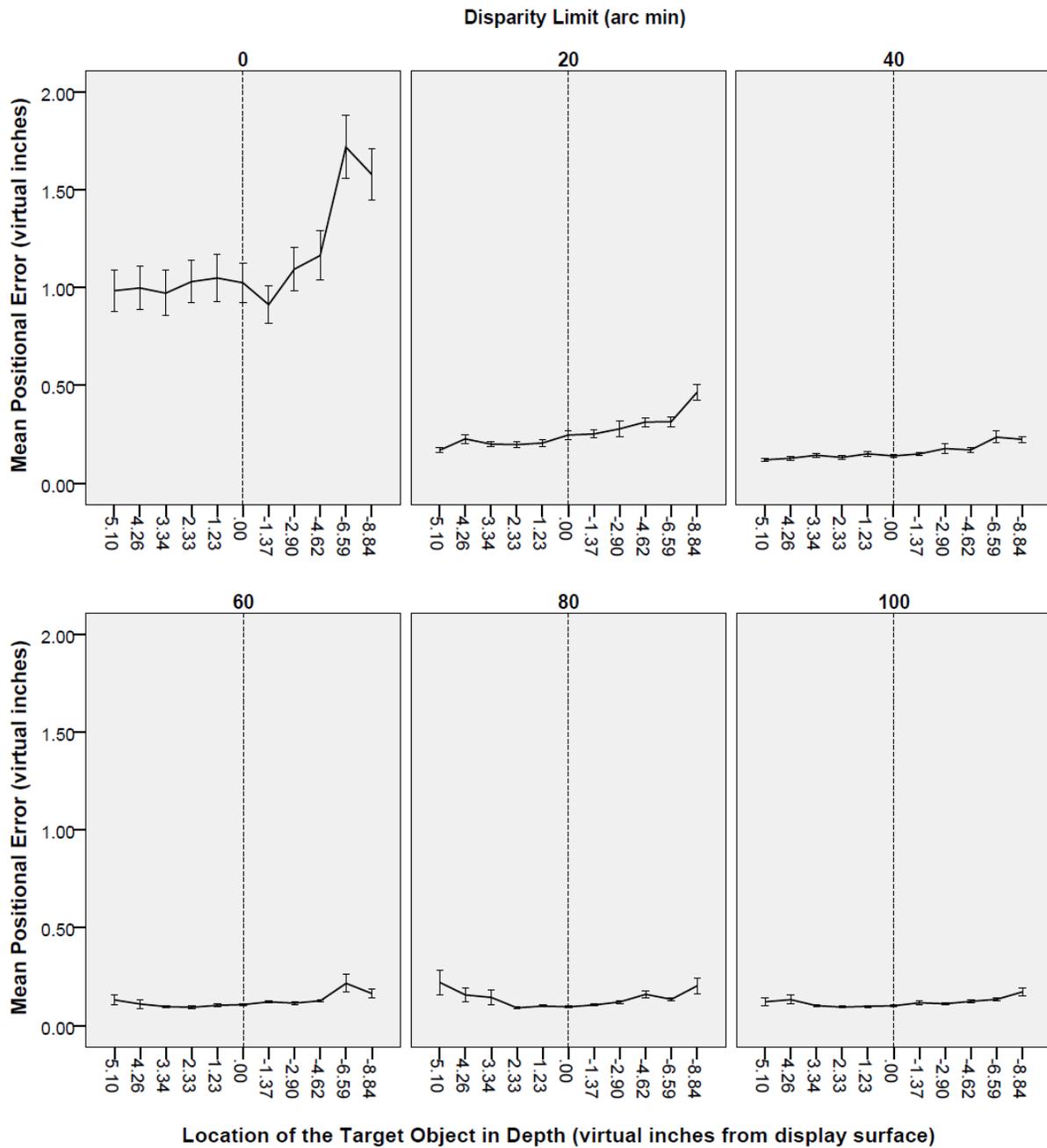
of a stimulus that is explicitly referred to by the One Degree Rule and the Zone of Comfort and to similar “depth budget” or “disparity limit” guidelines for ensuring viewer comfort. However, in terms of performance, it is not necessarily *only* the positional disparity of an object that allows for enhanced depth perception; the binocular parallax afforded by two slightly different views also provides enhanced shape perception of any given object. This is the concept of the “roundness factor” described by Kytö, Hakala, Oittinen, and Häkkinen (2012) and Yamanoue, Okui, and Okano (2006).

For instance, in a viewing condition with 100% camera separation (100% virtual IPD), an object whose center is located at the screen plane may technically have a positional measure of 0 degrees of binocular disparity relative to the screen. But this object could still give rise to a stereoscopic perception of shape and depth of the features of that object due to the different views afforded by the two camera views (via binocular parallax). In this study, differences in shapes of the virtual objects and the texture patterns visible to each camera’s viewpoint would permit the stereoscopic perception of objects’ shapes, in addition to those objects’ positions in space.

The question thus arises as to whether it is the absolute positional disparity (how far into or out of the screen plane) that provides a performance benefit for stereoscopic viewing, or whether it is the gross magnitude of the camera separation (the amount of binocular parallax) which improves performance? Our data clearly suggests that it is the latter, rather than the former: camera separation magnitudes primarily determined depth-task performance (see Figure 7-12). The actual location of the target object in the viewing volume has little effect on performance, although two trends involving location may be worth pointing out. First, in the zero disparity condition (no camera separation), performance was especially poor when the target

object was far into the screen, away from the viewer, at the extremes of the tested distances (~6 to 8 inches back in the virtual volume). This is not necessarily surprising given that the size of the visual cues for performing the task scale inversely with distance. We see a similar though more subdued trend involving distance in the condition using 20 arc min limits (this is equivalent to a camera separation of 20% of virtual IPD), and perhaps also the 40 arc min limits as well. However, at the larger camera separations resulting in 60 to 100 arc min limits, this effect of distance has largely disappeared.

The second trend worth noting is that in these larger camera separation conditions (limits of 60 to 100 arc min), there is a very slight U-shape to the curves, probably indicative of the fact that positional disparities *do matter, but only at extreme depth positions*, in which the fusion limits of the binocular visual system are likely taxed and under stress: i.e., the nearest and farthest locations in the conditions with the largest camera separations, as expected. Previous literature shows that depth-discrimination thresholds (disparity pedestals) start rising with pedestals as small as +/- 20 arc min and may continue increasing up to +/-1 or 2 degrees (Howard & Rogers, 2002). This is similar in concept to the present research in which there is large magnitude stereoscopic disparities requiring fine relative depth judgments, off the plane of the S3D display surface.



**Figure 7-12.** The effect of target location in depth on positional error, for each disparity limit condition. Target locations ranged from +5.10 virtual inches off the screen (near to the viewer) to -8.84 inches behind the screen (away from the viewer). The dashed lines indicate the depth of the screen plane at zero. Error bars represent +/- 1 SEM.

**Summary of Additional Analyses.** In summary, these further analyses of performance provided confirmation of intuitive predictions as well as providing additional insights into the data. First, we confirmed that stereo 3D cues help performance primarily in the z-axis depth dimension, as would be expected. This was true for both positional and rotational performance. Second, we showed that manipulations of the target's orientation seemed to result in worse performance for non-canonical orientations, similar to the commonly-observed *oblique effect*, although such effects largely disappeared once S3D cues were provided. Third, we also showed that target orientations resulting in possible visual occlusion from the control object, as well as possibly extreme orientations requiring awkward hand movements/positions, also seemed to negatively impact performance in both non-stereo and S3D conditions. Fourth, our data strongly suggests that the primary contributing factor in terms of performance improvements from S3D is the binocular parallax cues provided by stereo camera separations, and not necessarily the positional disparity afforded by an object's position in depth relative to the screen surface.

This last observation is particularly important, since most of the rules-of-thumbs and guidelines involving stereoscopic imaging and viewing (e.g., One Degree Rule; Zone of Comfort, etc.) refer to the positional disparities of objects. This observation also suggests where the nature of the comfort versus performance trade-off may lie: viewers need sufficient binocular disparity cues to see the *shapes* of objects in depth, and also to enhance the perception of the *positions* of objects in depth, but the binocular disparity cues must not be so large that tolerances in positional disparity are exceeded (thus resulting in discomfort and/or loss of fusion). This concern seems likely to be task-dependent, i.e., depending on the nature of the spatial task

viewers are required to perform, and the extent to which shape and/or position information is necessary for high performance.

### **Disparity Limits and Sickness, Discomfort, Eye Strain, and Balance**

**Simulator Sickness Questionnaire.** Again, similar to Experiment One, we failed to find a significant effect of the disparity limit on the Simulator Sickness Questionnaire Total Score pre/post change, using a repeated measures ANOVA [ $F(5,55)=1.105, p=.369$ ], nor on the SSQ's Oculomotor subscale [ $F(5,55)=0.828, p=.535$ ], as one might have expected. Both of these results fail to support the Cue Conflict Theory of simulator sickness as applied to S3D displays as predicted. Further, unlike Experiment One, there were no significant partial correlations between the disparity limits and the pre/post changes in SSQ scores, indicating that increases in disparity limits did not consistently result in higher discomfort symptoms (see Table 11).

Also, a paired sample one-tailed *t*-test was conducted between the average SSQ changes in the S3D conditions (sessions with non-zero disparity limits) versus the ratings in the 2D condition (sessions with the zero disparity limit). The results show no difference between 2D and S3D sessions in terms of discomfort (Table 12). These findings fail to support the utility of Cue Conflict Theory of simulator-sickness-type symptoms as applied to S3D displays. Similar to the reasons discussed in Experiment One, our failure to find strong support for Cue Conflict theory in this case may be due to one or more of the following reasons.

Range restrictions may again be an issue; although in this sample we included participants up to age 51 instead of restricting to age 40 or below, we again limited disparities across all conditions to a maximum of 100 arc min, and we limited S3D exposure time to 30 minutes. There were large differences observed across participants, but seven of the 12

participants indicated little or no discomfort by giving near-zero average ratings of discomfort across sessions (although a thirteenth volunteer dropped out due to visual discomfort). Of the 72 total pre/post observations, 51% were changes in score of zero (or lower), and 64% were changes of less than two.

Kennedy et al. (1993) warned that 40% to 75% of the SSQ rating scale items were likely to be zeros, and suggested that the more interesting values to experimenters are probably the non-zero items. In looking at the non-zero rated items, we found that all pre/post changes in SSQ scores of positive four or larger (17% of observations) only occurred in the S3D sessions. It may be worth adding, too, that one volunteer had to be withdrawn from the study halfway through data collection due to the extreme eyestrain experienced in the higher disparity sessions. While a few of these observations are interesting and suggestive, overall the results do not suggest a relationship between disparity limits and comfort, as the S3D display seemed fairly comfortable for most (but not all!) viewers, even when using disparity limits larger than the recommended One Degree Rule.

**Objective Measures of Eyestrain.** Only three of the possible 24 partial correlations between the objective and subjective measures were statistically significant (see Table 13). The three significant results were all from the same single objective measure: the *fusion far point* which correlated with the SSQ Nausea subscale, the SSQ Oculomotor subscale, and the SSQ Total score. Interestingly, none of the other objective eyestrain measurements correlated with subjective discomfort ratings, as might be expected. Also, only one of the objective eyestrain measures significantly correlated with the disparity limit manipulation: the pre/post change in the *vertical phoria* at far distance, a finding which seems difficult to explain given its lack of relationship to subjective discomfort. These data suggest that although the S3D system did not in

general cause much subjective eyestrain or viewing fatigue, the discomfort that did occur seemed to correlate with only one objective indicator, the *fusion far point* as measured on the S3D device.

This lone finding seems consistent with Experiment One and with previous research which suggests a possible relationship between viewer fatigue/eyestrain and fusion-related optometric measurements (e.g., see McIntire, Wright, Harrington, Havig, Watamaniuk, & Heft, 2014; also see Appendix 2). In totality, though, Experiment Two suggests that objective measures of eyestrain do not correlate strongly with subjective reports of discomfort, which in this case may be partly due to the low levels of discomfort apparently induced by the disparity limit manipulation (only up to 100 arc min) and the brief viewing times (limited to 30 min per session).

**Postural Instability.** Changes in postural stability before and after each session were measured as a possible alternative objective indicator of simulator sickness, disorientation, fatigue, and/or discomfort. The prediction was that virtual environment-related sickness or discomfort issues would be related to changes in postural stability. But changes in the average center-of-pressure velocity did not significantly correlate with self-reported discomfort as measured by the SSQ, or with the disparity limit manipulation on the S3D display (see Table 13). These results confirm those of Experiment One and suggest that the Postural Instability Theory of simulator sickness may not apply well to understanding discomfort on S3D displays, since pre-to-post changes in balance do not appear to be related to subjective reports of discomfort/sickness, nor to the disparity limit manipulation.

### **Predicting Performance with Optometric Measures and Pre-Session Measures**

Just as in Experiment One, an optometric screening process was conducted by a USAFSAM professional optometrist to investigate possible relationships with performance on S3D displays. We collected a variety of optometric screening measures (and further, computed several measures derived from these) for each participant, and also repeated some measures before each experimental session. Additionally, given our results and analysis from Experiment One, we also measured each participant's stereoacuity threshold (as described in the section *Experiment One – Follow-up Testing*). Then these measurements were correlated with the averages of S3D performance for each individual, which were the mean positional and rotational performance of all trials in which any magnitude of S3D cues were present (with correlations run separately for positional and rotational errors).

For the clinical screening measures and their derived measures, there were many significant correlations with either positional performance, rotational, or both (see Table 15). For the *fusion range* measures taken at distance, there were several significant correlations with performance. The base-out break point was related to both positional performance ( $r=-.72$ ,  $p=.004$ ) and rotational performance ( $r=-.68$ ,  $p=.007$ ). The base-out recovery point was related to rotational performance ( $r=-.81$ ,  $p<.001$ ). And the total fusion range was related to both positional ( $r=-.68$ ,  $p=.007$ ) and rotational performance ( $r=-.83$ ,  $p<.001$ ). Phorias measured at distance were not significantly related to performance in the suspected direction, nor were refractive errors.

Similar to the distance measures, the *fusion range* measures taken at near revealed some significant correlations with performance. Specifically, total fusion range was significantly correlated with both positional ( $r=-.52$ ,  $p=.042$ ) and rotational ( $r=-.71$ ,  $p=.005$ ) performance. Again, phorias measured at near were not significantly related to performance. The *stereoacuity thresholds* of the participants correlated with performance, but only rotational performance

( $r=.54$ ,  $p=.035$ ), not positional performance. This last observation may provide further support for the idea that rotational performance benefits more from incrementally larger magnitudes of S3D than positional performance, which was noted earlier in a comparison between Figures 7-1 and 7-2.

For the pre-session optometric measures, four were significantly correlated with performance in the suspected direction (see Table 16). All four significant measures were related to fusion ranges. *Fusion near point* (near point of convergence) correlated with positional performance ( $r=-.55$ ,  $p=.032$ ). *Fusion far point* (far point of convergence) correlated with rotational performance ( $r=.61$ ,  $p=.018$ ). And the combined measure of total *fusion range* on the S3D display correlated with both measures of performance: positional ( $r=-.57$ ,  $p=.026$ ) and rotational ( $r=-.50$ ,  $p=.049$ ). Here again, phorias failed to predict performance. These pre-session findings were consistent with the pre-experimental screening results.

Given the large number of statistical tests conducted for these analyses, the concern of Type I error rates (false positives) could become problematic. If we were to utilize a more stringent, conservative alpha level of .01 instead of .05, we would have found that only the clinical optometric measures relating to fusion range were significantly related to performance (not stereoacuity): see Table 15. The pre-session measurements in Table 16 would fail to achieve significance under these more stringent criteria. So our optometric predictor results should be interpreted with a bit of caution, and may suggest larger sample sizes could be appropriate for future studies if similar effect sizes are suspected. In our case, the consistency of results, in both experiments, and across pre-session versus clinical measurements (specifically regarding fusion ranges and stereoacuity) allows for confidence in our findings despite the large number of statistical tests utilized.

In summary, our results suggest that optometric measurements related to fusion ranges and stereoacuity thresholds were generally predictive of performance. Specifically, viewers with larger fusion ranges (convergent, divergent, and/or total ranges) and lower (better) stereoacuity thresholds generally performed better on the S3D display. Again, and in support of the results from Experiment One, these relationships suggest that some participants were better able to view the larger disparities on S3D displays before losing fusion, and/or were better able to utilize the available stereoscopic 3D cues, and might help explain why some viewers gain such a large relative performance advantage from S3D cues when compared to others.

### **Predicting Discomfort with Optometric Measures and Pre-Session Measures**

In an attempt to potentially predict which viewers might find S3D displays particularly uncomfortable, we correlated the SSQ self-reported changes in discomfort in the S3D display conditions with the pre-session optometric measurements and the USAFSAM clinical measurements (including refractive errors, horizontal and vertical phorias, fusion near and far limits, fusion ranges, and stereoacuity thresholds). We found no statistically significant correlations between optometric measures and reported discomfort as induced by the stereo display. We should note that for the one participant who dropped out halfway through data collection, due to eyestrain and discomfort from the larger disparity conditions, this participant had the smallest total fusion ranges (both near and at distance) of the 28 potential volunteers who received optometric pre-screenings during the course of this research. For distant fusion, this participant's base-in break/recovery was 6/2 prism diopters, base-out was 10/6; and for near fusion, base-in was 18/12 while base-out was 10/3. These gave a distance break-to-break range of 16 prism diopters (versus an average of 31.9 for all observers in Experiment One and Two) and a near break-to-break range of 28 prism diopters (versus an overall average of 42.1).

We might suppose that viewers who demonstrated more initial discomfort, even before testing in the experimental sessions, might have been more sensitive to discomfort experienced during the task. But initial discomfort as measured by the SSQ (see Appendix 4) did not seem to predict who would experience discomfort during the task: the pre-test Total SSQ scores did not significantly correlate with the pre-to-post changes in Total SSQ scores in the suspected direction ( $r=-.219, p>.05$ ). This lack of effect was also true for all three subscale SSQ scores and their correlations with the pre-to-post changes.

Postural Instability Theory specifically predicts that users with more initial postural instability should result in higher levels of discomfort from simulator/virtual environment exposure, but we also tested this correlation and found it to be non-significant ( $r=-.35, p=.735$ ). Further research on these topics using larger sample sizes and other optometric measurements may be warranted, given their importance and possible utility for operator selection/screening and for comfortable application of S3D technology.

## VIII. EXPERIMENTS ONE AND TWO – COMBINED ANALYSIS

The similarity in experimental design, structure, and resulting data from Experiments One and Two allows us to do a limited but combined analysis of some of the data. For this section, all of the optometric, balance, personal history/demographics, average SSQ discomfort ratings, and average S3D positional (but not rotational) performance can be conglomerated into a single data set, containing each of these measures for the 24 total participants. This will conceivably strengthen our ability to draw conclusions about possible correlations between the predictor variables and S3D performance and comfort. However, it should be kept in mind that the two Experiments, while extremely similar, were not strictly identical, and so any conclusions drawn from this analysis should be considered carefully.

In terms of predicting performance, the optometric measure of *total fusion range* ( $r=-.394, p=.028$ ) was significantly correlated with positional accuracy on the S3D system. This was the pre-session measure of fusion range as conducted on the S3D display system, not as measured clinically. We also found that the pre-session measure of *fusion near point* was just marginally non-significant ( $r=-.337, p=.054$ ). These results are generally consistent with the results from the analyses of the individual experiments, in which fusion range measures (especially total fusion ranges) were consistently predictive of S3D performance. One additional demographic measure did have a relationship with S3D performance: *gender* ( $r=-.375, p=.035$ ). Specifically, males performed slightly better in terms of positioning the virtual objects. Such a result is not unexpected, as it is common to observe small but statistically significant gender differences in spatial cognition and spatial motor tasks (e.g., Bosco, Longoni, & Vecchi, 2004; Parsons, Larson, Kratz, Thieboux, Bluestein, Buckwalter, & Rizzo, 2004).

In terms of predicting discomfort, the previous analyses in Experiment One and Two failed to find any significant correlations with the optometric measures. Again we found the

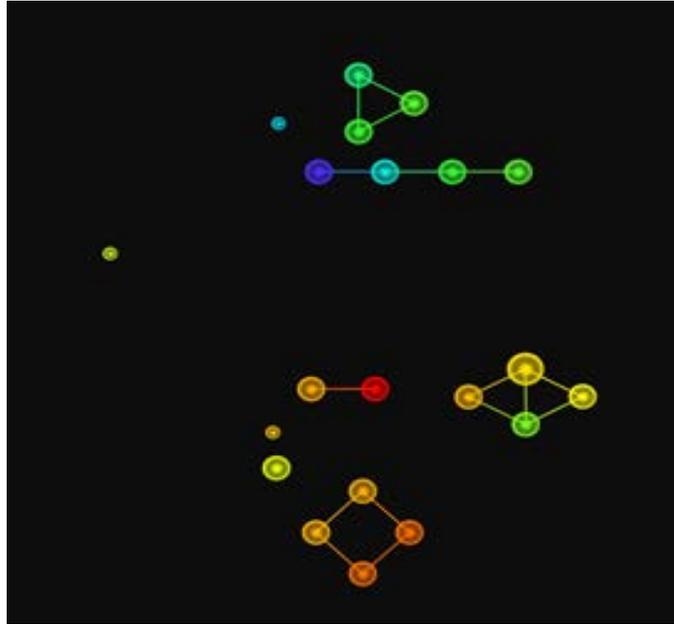
same lack of results in the combined dataset when using the same predictor variables. However, one demographic variable correlated significantly with reported discomfort on the S3D displays: *history of motion sickness* ( $r=.403$ ,  $p=.026$ ). Viewers who self-reported a history of experiencing motion sickness tended to report more discomfort from S3D display viewing. The most common type of motion sickness history reported was carsickness, followed by seasickness. Frequency or magnitude of previous motion sickness history was not assessed.

Such a finding was in fact predicted by the Cue Conflict Theory of motion and simulator sickness. The underlying concept is that some individuals are prone, for whatever reason, to feel discomfort and illness when experiencing sensory or perceptual conflict that is common within modes of transportation including cars, boats, airplanes, etc. This finding may be the first reported instance that these same people are also apparently more likely to feel discomfort/illness when viewing S3D displays, presumably due to the same issue of cue conflicts. There is at least one previous report of a possible relationship between a history of carsickness and 2D/S3D viewing discomfort (Solimini, 2013), though this was an observational study of movie-goers and had a small age range of participants of 18-30 years old. Our finding that a history of motion sickness may predict S3D viewing discomfort on an operationally-relevant type of task provides some evidential support for the utility of Cue Conflict Theory as applied to S3D displays.

### **Exploratory Topological Analysis of the Optometric Data**

An exploratory topological data analysis, often useful for studying multivariate high-dimensional datasets, was utilized to explore for any relationships between the many optometric assessment measures and the few dependent measures of interest. The techniques used herein are roughly analogous to some of the methods used in the topological analysis of biological data for cancer research by Nicolau, Levine, & Carlsson (2011). Simplex clustering was derived for the

data from two experiments using a normalized correlation metric combined with principle and secondary metric singular value decomposition. The resulting clusters showed two main groupings based on a combination of different phoria and fusion range-related measures.

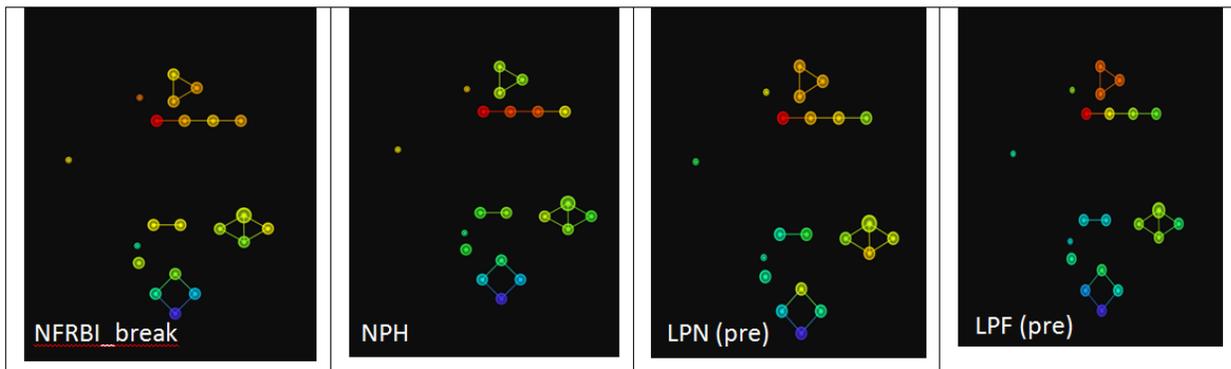


**Figure 8-1.** Topological space showing relationships between clusters of data points; clusterings represent similarity of measurements across individuals. This representation is color-coded by FFP (pre-test for fusion far point).

The image in Figure 8-1 illustrates the major groupings of data based on similarity between data points; Table 17 shows the strongest distinguishing optometric measures differentiating the groups. Note that the data points in the figures do not represent a single variable nor a single participant, but instead represent a cluster of measures; lines connecting these data points represent strong relationships between connected clusters. In all images, the red is the high end of the range of values on a selected dimension, while indigo is the lower end. The signed KS-score is the Kolmogorov-Smirnov two-sample test statistic, and that a larger value indicates a stronger difference in the distributions between the upper and lower groups. The

image in Figure 8-1 is color coded by FFP (the fusion far point pre-session test), capturing that the upper grouping exhibited lower FFP values, indicating larger divergent fusion ranges on the S3D display than those in the bottom clusters.

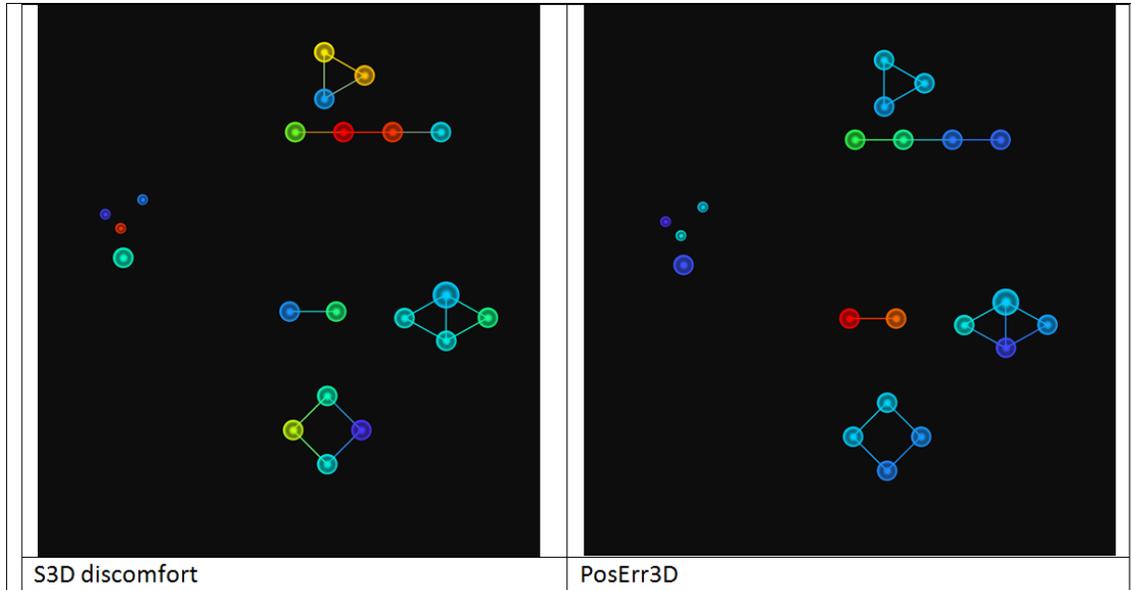
The six next most differentiating characteristics are listed in order in the table. Of these six optometric measures, three relate to *horizontal phorias* and three relate to *fusion ranges*. These data show that horizontal phorias and fusion range measures seem to cluster together, suggesting that individuals with larger, more divergent phorias (tending towards *exophoria*) also seem to demonstrate larger fusion range measurements in the divergent (base-in or far point) direction. Conversely, those with more convergent phorias (tending towards *esophoria*) seem to also demonstrate smaller divergent fusion range capabilities. The clusterings in Figure 8-2 suggest that individuals group together according to similar results on combinations of these optometric measures, and not just due to one measure alone (note the consistency of more reddish colors in the top group, versus more blue and indigo in the bottom group).



**Figure 8-2.** The topological clusterings color coded near fusion range base-in break point, near phoria in horizontal, lateral phoria at near (pre-session measure), and lateral phoria at far (pre-session measure), from left to right. This demonstrates consistency in the clusterings according to fusion ranges and lateral phorias.

These findings fit well with the optometric concepts encapsulated in *Percival's criterion* and *Sheard's criterion* (Stidwell & Fletcher, 2011) which relates to optometrically “balancing” the binocular system so that heterophoria is reduced and sufficient fusional reserves exist for both convergent and divergent eye movements. In terms of S3D viewing, our findings suggest more exophoric individuals will tend to demonstrate larger S3D fusion ranges for uncrossed disparities, but may result in discomfort, eyestrain, or difficulty for crossed stimuli if their relative total fusion range is insufficient. Conversely, esophoric individuals would conceivably demonstrate the opposite pattern, showing a preference for crossed stimuli and perhaps showing difficulty with uncrossed stimuli if their divergent fusion range abilities are limited. Some experimental support for these ideas was previously provided by Shibata et al. (2011b), who showed that measured phoria demand lines in relation to measured fusion limits (i.e., the Zone of Clear Single Binocular Vision) are moderately predictive of discomfort and eyestrain ratings, with overall correlations ranging from .208 to .275. If similar effect sizes were present in our dataset, they would be too small to detect as significant using traditional correlations given our sample size of 24, even if a direction of effect were specifiable.

This relationship between phoria and fusion ranges, and their relationship to discomfort with S3D, is also evidenced in the discomfort data (Figure 8-3, left panel), where the top group also shows higher levels of discomfort from desktop S3D viewing. The situation with positional error performance in S3D, however, is less clear (Figure 8-3, right panel): this plot suggests that the top group also seemed to show lower errors (better performance). It is unclear why this would be, and is opposite of expectation (we would expect some trade-off between comfort and performance, if both are highly dependent on the magnitude of S3D cues).



**Figure 8-3.** Topological clusterings color-coded by discomfort (left) and performance (right). Higher discomfort measures (reds, oranges, and yellows) tend to appear in the upper group. Poor performance appeared in the lower group.

By classifying the 24 participants into groups based on their distance phoria, we found that six participants demonstrated exophoria (positive horizontal phoria), nine demonstrated esophoria (negative horizontal phoria), and nine demonstrated orthophoria (zero). The averages on several optometric measures of interest are presented in Table 18. We see that both the esophoria and exophoria groups had smaller fusion near points than the ortho group. As hypothesized, we see that the exophoria group had a larger fusion far point average than the ortho group, which was larger than the eso group, just as expected. Interestingly, both the eso and exophoric groups demonstrated smaller total fusion ranges than the orthophoric group. And just as expected, viewers with any heterophoria (eso or exo) at distance tended to report more discomfort on S3D displays than the orthophoric viewers.

These results were very interesting given the analyses from Experiments One and Two, where we found that the optometric measures, when considered individually, suggest only fusion range measures to be useful predictors of performance, and no measures were predictive of discomfort. The present analysis, in contrast, suggests that more complex interactions of optometric variables might yield more insight into S3D discomfort and perhaps performance; specifically, measures of horizontal phoria and its possible interaction with fusion range-related measures. Further research on this topic is recommended.

## **IX. GENERAL RESULTS AND DISCUSSION**

In this work, we investigated performance and viewing discomfort in two experiments, using 2 DOF and 5 DOF manual spatial alignment tasks. We manipulated binocular disparity limits across experimental sessions, and correlated outcomes with several before-and-after measures (both objective and subjective) and with optometric screening data. Here we discuss the general results of both studies and the implications they may have in relation to previous research and in relation to several major theories currently guiding research in this field. Each section below tries to answer a specific hypothesis or major question of interest, which had initially guided the formulation of the two experiments.

### **Did the manipulation of disparity limits affect performance?**

Yes. In both experiments, spatial task performance generally improved whenever any level of S3D cues were provided. In Experiment One, in which only positional accuracy was required, performance improved by an average of 53% with S3D. In Experiment Two, in which both positional and rotational accuracy were required, positional performance improved by an average of 86% and rotational performance improved by an average of 29% with S3D cues.

### **Did the manipulation of disparity limits affect comfort?**

There was little to no effect. In both experiments, only moderate to low levels of discomfort were generally reported, although some occasional high ratings of discomfort were evident only in the S3D conditions. It is also worth observing that one participant had to be discontinued halfway through data collection due to the extreme eyestrain/discomfort experienced in the higher disparity conditions. In Experiment One, a few statistical tests indicated small but significant effects of increasing disparity limits on discomfort. In Experiment Two, no statistical tests in this regard reached significance. The overall lack of effect on comfort is probably due to the fact that very large magnitudes of binocular disparity were not displayed

(limited to at most 100 arc min), and also we used relatively brief viewing sessions (30 min per session). In real-world S3D viewing situations, larger disparities and lengthier viewing sessions are conceivable and probably common, for instance in S3D movies, gaming, surgical procedures, etc. so these lack of results should not be interpreted to suggest viewing comfort is not an issue for S3D display systems.

### **Is there a stereo “sweet spot” where both performance and comfort are high?**

Yes. Previous research had raised the intriguing possibility that low levels of binocular disparity (in the range of 10 to perhaps 50 arc min) might elicit improved performance while minimizing viewer discomfort. In both experiments, we found that providing even small magnitudes of disparity (limited to 20 or 40 arc min) greatly improved performance over the non-stereo conditions. However, the absolute best performance was always achieved in conditions with larger disparities than this lower range. In Experiment One, the highest accuracy was achieved in the range of 80-100 arc min disparity limits. In Experiment Two, for positional accuracy, performance plateaued in the range from 60-100 arc min disparity limits, and for rotational accuracy, performance continued to improve up to the 100 arc min disparity limit. Most magnitudes of disparity were generally comfortable and usable.

### **What are the implications for Microstereopsis versus Orthostereopsis?**

These results paint a somewhat complicated picture for the concept of microstereopsis. To the question “Does microstereopsis improve performance over no stereopsis?” the answer is clearly and unequivocally: Yes. Even small magnitudes of disparity (simulating camera separations of only 20 or 40% of viewer eye separations) provided substantial gains in performance across both experiments, with little or no effect on viewer discomfort. But the results also suggest that the best performance is achieved with orthostereoscopic or near-

orthostereoscopic levels of binocular disparity (simulating camera separations of 60 to 100% of viewer eye separations), again with little discernible effect on viewer discomfort. For situations in which fine spatial performance accuracy is not critical, the use of microstereopsis can be recommended based on our results. But in other limited situations, in which spatial accuracy may be absolutely critical, including surgery, robotic bomb disposal, and perhaps aerial refueling, the use of camera separations in the orthostereoscopic range should be seriously considered. The question arises as to what effect *even larger* separations (i.e., hyperstereopsis) might have on performance or comfort; the present research remains silent on this question, due to the fact that we tested levels only from no stereopsis, through microstereopsis, up to orthostereoscopic levels of camera separation. This remains an intriguing, open research question.

#### **What are the implications for the One Degree Rule regarding disparity limits?**

The One Degree Rule is based on extensive human factors research supporting the existence of a general “Zone of Comfort” of +/- 1 degree of binocular disparity, beyond which S3D stimuli are likely to result in viewer discomfort (see Introduction). This guideline is meant to be used as a practical limit for the imaging and display of stereoscopic media. In practice, this might mean choosing near and far points of interest in a scene, and then choosing an appropriate camera separation before imaging stereo content (although this technique must make assumptions about the resultant viewer/display geometry). Another technique might be to modify the on-screen half-image stereo-pair separations so that a One Degree limit will not be violated for viewers of a particular display system, regardless of the camera separations used to capture the imagery, although not all imagery may completely allow for this (e.g., if the near-to-far disparity range in the imagery is larger than two degrees).

Experiments One and Two chose the former method for testing, in that the virtual camera separation was manipulated across sessions so that the near/far limits of the virtual viewing volume corresponded to maximum disparity levels of +/- 20, 40, 60, 80, or 100 arc min limits. The results from both experiments suggest that using disparity limits beyond the One Degree Rule (the 80 and 100 arc min disparity limits) were actually quite tolerable, inducing little subjective discomfort, and even resulted in some performance gains relative to the lower disparity limits. In summary, we found little support for the usefulness of the One Degree Rule in our experimental data.

The reason for this finding, and its apparent conflict with extensive previous literature, may lie in the work of Shibata et al. (2011b), which was previously discussed in the Introduction (see Figure 1-8). Shibata et al. plot the shape of the Zone of Comfort in units of binocular disparity and across viewing distances, making note of common distances for desktop, television, and cinema viewing. We can see from Figure 1-8 that while instituting a One Degree Rule typically ensures comfortable viewing across a wide range of viewing distances, the actual width of the zone tends to increase in size for nearer distances. Their work thus suggests, in apparent conflict with the One Degree Rule, that disparities of up to about +/- 2 degrees or more may be comfortably tolerated at typical desktop viewing distances. Our experiments, conducted on a desktop S3D system, and testing disparity limits up to +/- 100 arc min (1.67 degrees), provide further support for this contention.

### **Did the objective measures of eyestrain/fatigue relate to subjective reports of discomfort?**

Mixed results. In Experiment One, two of the seven pre-to-post session objective measures significantly correlated with subjective reports of discomfort: *fusion near point* and *fusion range*, both of which were measured on the S3D display. In Experiment Two, only the

objective measure of *fusion far point* significantly correlated with subjective discomfort. A few key points can be made about these observations. The first is that overall, across the 14 total measures (7 per experiment), only 3 significant results were found. This might suggest generally that objective measures of eyestrain are poor substitutes for the ‘gold standard’ of subjective reports. A reasonable objection might be that since subjective discomfort was low overall (across participants and experiments), there is little we can conclude about these lack of significant correlations. It remains possible, especially given previous related research and a few of our findings herein, that at least some objective measures of eyestrain/discomfort are indeed generally useful but may require larger magnitudes of subjective discomfort to be induced before such effects become manifest in objective measures. Some support for this idea comes from our finding that in both Experiments One and Two, objective measures relating specifically to fusion ranges (near point, far point, and/or total range) were significantly correlated with subjective scores, at the very least suggesting some consistency for fusion-related objective measures. Additional support comes from the observation that the one participant in Experiment Two who dropped out due to eyestrain had the smallest binocular fusion ranges we measured.

**Did any of the optometric data help to predict performance and/or discomfort?**

Yes. Using individualized data from the pre-experimental optometric screening, and from the pre-session measures of phorias, fusion ranges, and balance, we were able to find significant predictors of subsequent individual performance on the S3D tasks, though no predictors of viewer discomfort were found. This was true for both individual experiments, though a combined analysis suggested that a personal history of motion sickness predicted discomfort from S3D display viewing. In Experiment One, the significant predictors of performance were: *fusion range*, *fusion near point*, and *stereoacuity*. Specifically, viewers with larger fusion ranges,

closer fusion near points, and lower (better) values of stereoacuity thresholds performed better on the S3D task (these are reported in McIntire, Wright, et al., 2014). In Experiment Two, there was further support for these measures: total *fusion ranges*, *fusion near points*, *fusion far points*, and *stereoacuity* all predicted better S3D task performance on one or both of the spatial performance measures.

### **What are the implications for Cue Conflict Theory versus Postural Instability Theory?**

Cue Conflict Theory and Postural Instability Theory are both general, overarching theories about the causes (and possible solutions to) virtual environment and simulator-induced sickness and discomfort. In a few instances, the two theories diverge regarding empirical predictions, which we can test given our dataset. However, given that only small magnitudes of discomfort were apparently induced in our two experiments, any conclusions favoring one theory over another should be considered tentative.

Cue Conflict Theory predicts that minimizing sensory/perceptual conflicts in a simulator or virtual environment system should result in lower levels of discomfort. In fact, our manipulation of virtual camera separations across experimental sessions can be considered a direct manipulation of the Accommodation-Convergence conflict (see *Introduction* and *General Method*). So, if Cue Conflict Theory is accurate, and can be usefully applied to S3D display systems, we would expect to see a relationship between disparity limits (i.e., camera separation levels) and reported discomfort. We found only partial support for this hypothesis. A significant correlation between disparity limits and discomfort was found in Experiment One but not Experiment Two; and repeated measures ANOVAs failed to find a significant effect in both experiments. A combined analysis found that a personal history of motion sickness (particularly

carsickness and seasickness) predicted S3D viewing discomfort, in support of Cue Conflict Theory.

Postural Instability Theory predicts that sickness/discomfort results from postural instability, and that instability is a primary cause of simulator sickness-type symptoms. Thus, if Postural Instability Theory is true and useful in regards to the study of S3D display systems, then two hypotheses can be generated from these ideas: (1) pre-to-post session changes in postural instability should correlate with pre-to-post changes in reported discomfort; and (2) posturally-unstable individuals (pre-test measure) should experience more discomfort in the experiments.

In regards to hypothesis (1), we failed to find any significant correlations involving pre-to-post changes in postural instability versus SSQ total or sub-score changes, in both experiments. In regards to hypothesis (2), initial postural instability of individuals did not predict reported discomfort from the virtual environment experience; this was true for both experiments as well.

In totality, our data provides little to no support for Postural Instability Theory as a useful means to understand, study, or combat simulator sickness/discomfort as applied specifically to stereoscopic 3D viewing systems. Our data provides some but limited support for the idea of Cue Conflict Theory as applied to S3D, and further research along these lines may be appropriate to solidify these tentative conclusions. It should be kept in mind that both of these theories were developed out of research involving motion sickness and also simulator sickness involving real or apparent (visually-induced) motion. Thus, these theories were originally applied to viewing situations in which the primary sensory/perceptual conflicts involved visual motion versus vestibular motion cues. Our work suggests these theories may have only limited utility in understanding or combating sickness involving other sensory, perceptual, or cognitive conflicts

specifically as experienced in S3D display systems. But this statement should not be interpreted as implying that Cue Conflict Theory or Postural Instability Theory are not good or useful theories for simulators or virtual environments; only that they may have limited utility for discomfort or sickness situations involving other conflicts in which (visual) motion is not a primary factor.

## **X. CONCLUSIONS AND FUTURE WORK**

The results of both experiments suggest that utilizing disparity limits of 60 to 100 arc min provide for optimal spatial task performance, although lower disparity limits still allow for vastly improved performance relative to the no-stereopsis conditions. Only low to moderate levels of viewer discomfort were induced across the range of disparity limits, and only for some viewers; this was also true for both experiments. For the concept of microstereopsis, these results suggest that using camera separations smaller than human IPD can improve performance while keeping discomfort low, but near-orthostereoscopic or orthostereoscopic camera separations are recommended when accurate spatial task performance is vital, and such separations still seem to be relatively comfortable. For the concept of the One Degree Rule for S3D viewing comfort, our results suggest that grossly violating this rule for near-viewed desktop S3D systems does not generally result in large magnitudes of viewer discomfort as might be expected, and that this rule might be safely relaxed up to at least +/- 1.67 degrees of disparity. We found some limited support for the idea of using Cue Conflict Theory to understand viewer discomfort with S3D systems, but no support for Postural Instability Theory.

### **Ideas for Future Research.**

To confirm and extend our results, we recommend future research using other spatial tasks, such as those involved with imagery analysis, route planning, navigation, complex robotic teleoperations, etc. This research focused on S3D stimuli ranging from microstereopsis to orthostereopsis, but the utility and human factors implications of hyperstereopsis has been somewhat ignored in the literature and seems a promising avenue of future study. For instance, research into hyperstereo (i.e., telestereo) has reported interesting perceptual phenomena like the “cardboard effect” and the “puppet theater effect” (e.g., Yamanoue, Okui, & Yuyama, 2000) which may bear some relationship to conflicts with stimuli that already possess familiar size cues

to a viewer. These effects might even be due to the relative magnitudes of an objects' shape disparity versus its relative positional disparity (also suggested by the results of Yamanoue et al., 2000), which is supported by our results from Experiment Two: perhaps an objects' shape disparity may be below a viewer's stereo threshold, while its positional disparity is well above threshold, with the resulting percept of flat cardboard-like surfaces ("puppets") placed at different relative depths in a scene. Such concepts, and their relationships to camera separation, micro- or hyperstereopsis, and the magnitude of binocular disparity, seem worthy of further experimental work. For the foreseeable future, and given the explosive growth and popularity in S3D technology as of late, the manipulation of binocular disparity (via camera separation) will likely remain as both a theoretically and practically interesting variable for studying issues of performance and comfort for users of such systems.

The optometric measurements predictive of performance were stereoacuity and fusion-range-related measures. Essentially, fusion range measurements fix an accommodative stimulus while varying vergence cues to test the limits of the binocular system (i.e., fusion break and recovery points). This is roughly how a stereoscopic stimulus functions as well; accommodation is fixed while vergence cues to depth vary relative to the screen depth. So it is not necessarily a surprise that fusion-range-related optometric measurements seem to predict depth task performance on S3D displays. It is also not theoretically surprising that stereoacuity threshold measurements predict performance on S3D displays. But our work seems to be the first to experimentally confirm these speculations regarding fusion ranges and stereoacuity; and the magnitudes of some of these correlations were also surprisingly high, suggesting such measures could have broad utility for applied settings and/or for operator selection purposes.

Further research into optometric measures will likely be valuable, including possibly measuring the strength of accommodation (via accommodative facility perhaps, or accommodative-convergence/accommodation ratio, or convergent-accommodation/convergence ratio; known as the AC/A or CA/C ratios), since the present work focused primarily on vergence-related binocular fusion measurements while holding accommodation constant. Additional optometric measures that might be interesting for S3D display evaluation include vergence facility, and perhaps fusion ranges defined by blur points (instead of break & recovery points). And eye tracking systems with high spatial and temporal resolution would be particularly helpful in determining how viewers' binocular visual systems are behaving during S3D display viewing sessions.

### **Lessons Learned.**

In this section I explore some of the lessons I learned during this experience, lessons that might be applied more generally in scientific or vision or human factors research; or other lessons that seem to apply more specifically to binocular vision or stereoscopic displays.

One important lesson learned is that each person's visual system is unique, complex, and idiosyncratic, despite seeing much general agreement in the performance data. This may be intuitively obvious to optometrists and others in related fields where individual oddities in people's visual systems are studied and explored in detail for research, selection, or treatment purposes. But this point was made explicit (to me) in the two experiments contained herein, in which some volunteers had to be excluded due to poor acuity in one or both eyes, unexplainable difficulty in viewing S3D stimuli, deficient binocular vision, and/or discomfort from S3D. Other volunteers who passed the screening nonetheless demonstrated sometimes very noisy performance data, or adopted unexpected or unpredictable task strategies that complicated

analysis and interpretation of their data. The wide range of performance observed in conditions with no stereo cues is also suggestive of large differences in inherent visual and manual skills, and perhaps reflects individual differences in how the various monocular visual cues to depth are weighted and combined into a singular percept of object depths in a scene.

Another important related lesson learned is that “normal” stereo vision is a vague, messy concept. For instance, official or professionally standardized norms for stereoacuity thresholds or fusion ranges for the most part do not exist, although there is ample data regarding their central tendencies and distributions in the general population and even within sub-populations of interest (i.e., military operators, males versus females, children versus adults, etc.). This issue of “normal” classification complicates binocular vision and S3D display researchers attempting to study only viewers with normal stereo vision, or who wish to compare groups of normal versus non-normal viewers. What’s more, some viewers with seemingly normal stereovision still might not utilize S3D cues, as would be expected. This is the concept of *pseudo-stereoanomaly* discussed in the results and follow-up testing in Experiment One, in which viewers who can otherwise see disparity cues fail to utilize such cues to improve their task performance on an S3D system. A few experiments have already suggested that some pseudo-stereoanomalous viewers can learn to use S3D cues via extensive training, but further research on this topic is warranted.

If I could repeat this study, what variables might I add, eliminate, or alter? If I was primarily interested in studying viewer discomfort specifically, I would have added a larger range of disparity limits, perhaps extending out to  $\pm 2$  or more degrees of binocular disparity. The limits we tested in this work simply did not induce much viewer discomfort at this viewing distance, and did not adversely affect a large portion of viewers. For investigating possible optometric and related predictors of viewer discomfort and/or performance, a larger sample size

of participants would have been even more valuable in solidifying our findings. If instead I was primarily interested in studying viewer performance specifically, I would have used a larger set of disparity limit (camera separation) conditions but that tested a smaller range (perhaps up to 40 arc min), to determine what seems to be the absolute minimum amount of disparity cues we could provide a viewer that would be beneficial over no stereo cues. Conceivably, the necessary magnitudes of such cues would be heavily dependent upon viewers' stereoacuity thresholds, and perhaps their own internalized depth-cue combination weighting functions.

### **Final Thoughts.**

In a review of visual depth perception in remote (tele-) operations, Reinhardt-Rutland (1996) had recommended avoiding stereoscopic viewing altogether, because of the depth cue conflicts and sensory system conflicts likely to be induced by such systems. Instead, it was recommended that an acceptable alternative strategy for selecting operators might be to rely on individuals who *lack* binocular function, since such individuals seem to have already compensated for everyday life without binocular vision (just as would be present in monoscopic teleoperative viewing). Our current research, and that of others, suggests such a strategy to be drastically over-conservative. While cue conflicts on stereo displays are certainly problematic for some viewers and likely contribute to eyestrain and fatigue issues, the performance benefits provided by stereo 3D viewing are too great to ignore, especially for remote operation of critical import such as the handling of dangerous materials or for surgical procedures. For instance, our data shows that providing S3D cues improved virtual object positional accuracy by as much as 53% in Experiment One, and as high as 86% in Experiment Two.

Stereoscopic 3D displays commonly suffer from the human factors-related problems of viewer fatigue, discomfort, and eyestrain. Such complaints may also affect user performance.

Given that S3D display technology is growing in availability and being adopted in widespread applications, the present research was meant to shed light on the factors that induce fatigue and that may also affect performance, and what might be done about this trade-off. Our data provide a somewhat complicated conclusion regarding the utility of such rules-of-thumb as *The One Degree Rule* and the concept of *microstereopsis*, and suggest that a more nuanced and application-specific approach may be more valuable. This view seems particularly true for spatial control applications in which the outcomes may be critical, perhaps including (but not limited to) S3D displays in surgery, bomb disposal, imagery analysis, air traffic control, teleoperative robotic control and navigation, and aerial refueling.

**Table 1. Different environmental situations and the implications of Cue Conflict Theory on motion and/or simulator sickness.**

Situation	Visual System	Vestibular System	Result
Walking, running, etc.	Motion is perceived.	Motion is perceived.	No perceptual conflicts are induced, and no sickness results.
Below deck on a boat during choppy seas.	No motion is perceived.	Motion is perceived.	Perceptual conflict is induced and results in “sea sickness.”
Passenger in a car, reading or with head down.	No motion is perceived.	Motion is perceived.	Perceptual conflict is induced and results in “car sickness.”
Virtually moving within a motion-based simulator.	Motion is perceived.	Motion cues are inadequate or inaccurate.	Perceptual conflict is induced and results in “simulator” or “cyber” sickness.
Virtually moving within a fixed-based simulator.	Motion is perceived.	No motion is perceived.	Perceptual conflict is induced and results in “simulator” or “cyber” sickness.

**Table 2. Summary of experimental studies that have manipulated binocular disparity to study the effects on performance and/or visual comfort/sickness.**

Authors	What they studied	What they found
Wöpking (1995)	Comfort/Sickness	With high resolution background imagery, comfort peaked at 35 arc min of disparity and dropped steadily thereafter, with considerable discomfort beyond 60 arc min. With low resolution background, comfort stayed reasonably high despite large disparities.
Rosenberg (1993)	Performance	Object placement accuracy improved with increasing disparity and peaked at about 20 arc min. Further increases in disparity beyond 20 arc min did not noticeably improve performance.
Ellis, Fishman, Hasser, & Stern (2005)	Performance	Object placement completion times were fastest with disparities at about 10 arc min and beyond.
Yano, Emoto, & Mitsuhashi (2004)	Comfort/Sickness	Evaluated objective and subjective measures of visual fatigue while manipulating disparity to fall within or outside the viewer's depth of focus ( $\pm 50$ arc min). They found a significant drop in comfort beyond 60 arc min of far disparity.
IJsselsteijn, de Ridder, & Vliegen (2000a/b)	Comfort/Sickness	Manipulated camera separation (disparity) and other camera setup parameters. They found subjective image quality peaked midway through the separation range, while eye strain increased steadily with increasing camera separations. Unfortunately, we were unable to derive the corresponding binocular disparities.
van Beurden, IJsselsteijn, & de Kort (2011)	Performance, Comfort/Sickness, and Workload	Manipulated disparity to study accuracy, response times, workload, and visual comfort on a visual path-tracing 3D task. Difficulty of the task was also manipulated. Performance seemed to be best at 25 arc min of disparity and higher. Strangely, discomfort and workload were lowest at 10 arc min in the easier condition, and were highest at 5 arc min the difficult condition. Statistical results were somewhat unclear, especially regarding the subjective measures.
Poyade, Reyes-Lecuona, & Viciano-Abad (2009)	Performance	Object placement accuracy was generally better with larger disparities, maxing out at 3.0 to 6.3 cm virtual camera separations. Corresponding binocular disparities cannot be derived from the given information.
Camposeco, Avilés, Careaga, Spindola, & Velázquez (2011)	Performance	Response times were assessed on a speeded precision placement task, and accuracy was assessed on a non-speeded precision positioning task. Both measurements suggested that performance plateaued at 63 mm and beyond of camera separation, roughly corresponding to 100% and greater of average IPD.

**Table 3. Binocular Disparities, Camera Separations, Virtual IPDs, and Viewing Distances to Virtual Object Locations.**

The viewing volume for the virtual object placements used in Experiments One and Two. The distances in depth to the virtual target object positions and their relationships to the Binocular Disparity Limits, Virtual Camera Separation, and percentage of virtual IPD are indicated. The binocular disparity calculations assume an average IPD of 2.6 inches (66 mm). The disparity calculations used the recommended formula in Appendix 1.

Note the white highlighted row which indicates the real distance of the display from the viewer (24 inches), which corresponds to zero disparity. The light grey lines indicate crossed disparity values (negative disparity), while the darker grey lines correspond to uncrossed disparity values (positive disparity).

Table of Binocular Disparity Values (in arc min)		Virtual Camera Separation, in mm (and alternatively, in % of virtual IPD)					
Distance to Target (virtual inches, Z depth)	Distance between Target and Display	0.0 mm vIPD = 0%	13.2 mm vIPD = 20%	26.4 mm vIPD = 40%	39.6 mm vIPD = 60%	52.8 mm vIPD = 80%	66.0 mm vIPD = 100%
18.90	- 5.10	0	- 20	- 40	- 60	- 80	- 100
19.74	- 4.26	0	- 16	- 32	- 48	- 64	- 80
20.66	- 3.34	0	- 12	- 24	- 36	- 48	- 60
21.67	- 2.33	0	- 8	- 16	- 24	- 32	- 40
22.77	- 1.23	0	- 4	- 8	- 12	- 16	- 20
24.00	0.00	0	0	0	0	0	0
25.37	1.37	0	4	8	12	16	20
26.90	2.90	0	8	16	24	32	40
28.62	4.62	0	12	24	36	48	60
30.59	6.59	0	16	32	48	64	80
32.84	8.84	0	20	40	60	80	100

**Table 4. Participants' Performance with 2D versus S3D Display in Experiment One**

In this table, mean placement error (in virtual inches) is reported for each participant in both the 2D and S3D display conditions, and the percentage improvement when changing from 2D to S3D. Each reported value represents the mean of all trials in which either S3D cues were present in any magnitude (S3D Display), or there were none (2D Display). The greyed cells indicate the participants classified as 'atypical'.

<b>Participant</b>	<b>2D Display Placement Error</b>	<b>S3D Display Placement Error</b>	<b>% Reduction in Errors with S3D Cues</b>
1	1.50	0.30	80
2	1.19	0.30	74
3	0.97	0.28	71
4	1.49	0.30	80
5	0.94	0.66	30
6	1.32	0.65	51
7	0.96	1.16	- 22
8	0.75	0.34	55
9	1.34	0.22	83
10	0.95	0.94	1
11	2.04	0.19	91
12	0.84	0.48	42
<b>Average</b>	<b>1.19</b>	<b>0.49</b>	<b>53</b>

**Table 5. Partial Correlations between Disparity Limits and the Simulator Sickness Questionnaire in Experiment One**

In this table, partial correlations controlling for participant and session (practice) effects were conducted. Significance tests were performed with a sample size of 72 (12 participants, 6 sessions each). Pre/post session changes in the SSQ, including the Nausea, Oculomotor, and Disorientation subscales, and the Total Score, were correlated with the disparity limit utilized in each session. Correlations greater than +/- .195 in the suspected direction were significant at the .05 level, as highlighted in grey. If a more conservative alpha value of .01 were utilized to account for multiple tests, none would have achieved significance.

<b>Sim Sickness Questionnaire</b>	<b>Partial correlation ( <i>r</i> ) with Binocular Disparity Limits</b>	<b>One-tailed Significance Level ( <i>p</i> )</b>
Nausea (subscale)	.13	.138
Oculomotor (subscale)	.26	.014
Disorientation (subscale)	.21	.038
Total SSQ Score	.26	.014

**Table 6. Simulator Sickness Questionnaire Results across Disparity Limits in Experiment One**

In this table, the raw data concerning SSQ Total scores (pre/post sessions changes) across disparity limits are reported, by participant and as an average across participants. The bottom two rows consist of paired sample one-tailed *t*-tests and the associated *p*-values comparing the non-zero disparity limit sessions (S3D) to the zero disparity limit sessions (2D): all were significant. Pre/post changes of 5 units or larger only occurred in the S3D sessions (light grey cells).

SSQ Total score ratings pre/post change	Disparity Limit (arc min)						Average in the 3D conditions (nonzero)
	0	20	40	60	80	100	
<b>Participant</b>	<b>0</b>	<b>20</b>	<b>40</b>	<b>60</b>	<b>80</b>	<b>100</b>	
<b>1</b>	0	1	7	6	13	4	<b>6.2</b>
<b>2</b>	-1	1	0	5	7	11	<b>4.8</b>
<b>3</b>	3	12	5	6	3	14	<b>8.0</b>
<b>4</b>	4	1	6	5	1	0	<b>2.6</b>
<b>5</b>	2	4	4	3	5	3	<b>3.8</b>
<b>6</b>	4	7	3	3	5	4	<b>4.4</b>
<b>7</b>	0	0	0	1	2	0	<b>0.6</b>
<b>8</b>	1	4	7	2	5	3	<b>4.2</b>
<b>9</b>	-3	0	0	0	1	0	<b>0.2</b>
<b>10</b>	1	0	1	1	1	1	<b>0.8</b>
<b>11</b>	1	2	1	2	0	2	<b>1.4</b>
<b>12</b>	0	0	0	0	0	0	<b>0.0</b>
<b>Average</b>	<b>1.0</b>	<b>2.7</b>	<b>2.8</b>	<b>2.8</b>	<b>3.6</b>	<b>3.5</b>	.
<i>t-value</i>		1.968	2.561	2.823	2.053	1.866	2.833
<i>p-value</i>		.038	.013	.008	.032	.044	.008

**Table 7. Partial Correlations between the Objective Measures of Eyestrain, Postural Instability, the Simulator Sickness Questionnaire, and Disparity Limits in Experiment One**

In this table, partial correlations controlling for participant and session (practice) effects were conducted. One-tailed significance tests were performed with a sample size of 72 (12 participants, 6 sessions each). Pre/post session changes in the SSQ, including the Nausea, Oculomotor, and Disorientation subscales, the Total Score, and Disparity Limits were correlated with the pre/post session changes in the objective eyestrain measurements. Correlations greater than +/- .195 in the suspected direction were significant at the .05 level, as highlighted in grey. Correlations significant at a more conservative .01 level, to account for multiple tests, are indicated with stars (\*).

<b>Sim Sickness Questionnaire</b>	<b>Lateral Phoria Near</b>	<b>Lateral Phoria Far</b>	<b>Vertical Phoria Far</b>	<b>Fusion Near Point</b>	<b>Fusion Far Point</b>	<b>Fusion Range</b>	<b>Postural Instability</b>
Nausea (subscale)	-.02	-.13	-.03	-.32*	.09	-.28*	.03
Oculomotor (subscale)	.03	-.04	.00	-.03	.02	-.02	-.19
Disorientation (subscale)	-.13	-.13	-.05	-.20	-.15	-.27*	.08
Total SSQ Score	-.04	-.12	-.03	-.20	-.02	-.21	-.06
Disparity Limit manipulation	.21	.15	-.01	-.08	-.02	-.08	.03

**Table 8. Correlations between USAFSAM Pre-screening Optometric Data and S3D Performance in Experiment One.**

In this table, correlations between the USAFSAM optometric screening data and performance on the S3D display are reported. Performance for each subject was summarized into one measure: an average of placement error for all trials in which S3D cues were present. One-tailed significance tests were performed with a sample size of 12. Correlations of +/- .50 or larger in the suspected direction were significant at the .05 level, as highlighted in grey. None were significant if using a more conservative alpha value of .01 to account for multiple tests.

<b>USAFSAM Optometric Measurements</b>	<b>Correlation (<i>r</i>-value)</b>	<b>Significance (<i>p</i>-value)</b>
Refractive Error (right eye)	.31	.163
Refractive Error (left eye)	.13	.344
Horizontal Phoria (distance)	-.29	.180
Vertical Phoria (distance)	-.24	.226
Fusion Range (distance) – Base-In Break	-.16	.310
Fusion Range (distance) – Base-In Recovery	-.21	.256
Fusion Range (distance) – Base-Out Break	-.31	.163
Fusion Range (distance) – Base-Out Recovery	-.10	.379
Fusion Range (distance)	-.38	.112
Horizontal Phoria (near)	-.19	.277
Vertical Phoria (near)	.09	.390
Fusion Range (near) – Base-In Break	-.14	.332
Fusion Range (near) – Base-In Recovery	-.09	.390
Fusion Range (near) – Base-Out Break	-.09	.390
Fusion Range (near) – Base-Out Recovery	-.40	.099
Fusion Range (near)	-.51	.045

**Table 9. Correlations between the Pre-session Optometric Data and S3D Performance in Experiment One.**

In this table, correlations between the averaged pre-session optometric data and performance on the S3D display are reported. Performance for each subject was summarized into one measure: an average of placement error for all trials in which S3D cues were present. One-tailed significance tests were performed with a sample size of 12. Correlations of +/- .50 or greater in the suspected direction were significant at the .05 level, as highlighted in grey. Given the number of statistical tests conducted, if a more conservative critical value for alpha of .01 were utilized (instead of .05), none of these tests would have achieved significance.

<b>Pre-session Optometric Measurements</b>	<b>Correlation (<i>r</i>-value)</b>	<b>Significance (<i>p</i>-value)</b>
Lateral Phoria (near)	-.45	.071
Lateral Phoria (far)	-.50	.049
Vertical Phoria (far)	-.45	.071
Fusion Near Point (Near Point of Convergence)	-.50	.049
Fusion Far Point (Far Point of Convergence)	.34	.140
Fusion Range	-.60	.020

**Table 10. Stereoacuity Threshold Measurements for Experiment One.**

In this table, participants' stereoacuity threshold estimates are provided individually, per measurement technique (in arc seconds). Performance in terms of average error magnitude of placement precision is also reported. Rows are ordered from best performance to worst performance descending. The two atypical participants are highlighted in grey.

<b>Participant</b>	<b>USAFSAM software</b>	<b>OVT and/or Randot estimate</b>	<b>S3D Mean Error (virtual inches)</b>
11	7.8	.	0.19
9	5.9	.	0.22
3	14.6	.	0.28
1	8.2	.	0.30
2	21.8	.	0.30
4	7.3	.	0.30
8	9.2	.	0.34
12	8.5	.	0.48
6	11.0	.	0.65
5	difficulty	20 - 30	0.66
10	25.5	.	0.94
7	difficulty	25 - 30	1.16

**Table 11. Partial Correlations between Disparity Limits and the Simulator Sickness Questionnaire in Experiment Two**

In this table, partial correlations controlling for participant and session (practice) effects were conducted. Significance tests were performed with a sample size of 72 (12 participants, 6 sessions each). Pre/post session changes in the SSQ, including the Nausea, Oculomotor, and Disorientation subscales, and the Total Score, were correlated with the disparity limit utilized in each session. Correlations greater than +/- .195 in the suspected direction were significant at the .05 level, though none were found to be significant in this data.

<b>Sim Sickness Questionnaire</b>	<b>Partial correlation ( <i>r</i> ) with Binocular Disparity Limits</b>	<b>One-tailed Significance Level ( <i>p</i> )</b>
Nausea (subscale)	.073	.271
Oculomotor (subscale)	.005	.483
Disorientation (subscale)	-.046	.649
Total SSQ Score	.005	.483

**Table 12. Simulator Sickness Questionnaire Results across Disparity Limits in Experiment Two**

In this table, the raw data concerning SSQ Total scores (pre/post sessions changes) across disparity limits are reported, by participant and as an average across participants. The bottom two rows consist of paired sample one-tailed *t*-tests and the associated *p*-values comparing the non-zero disparity limit sessions (S3D) to the zero disparity limit sessions (2D): none were significant, although pre/post changes of 4 units or larger only occurred in the S3D sessions (light grey cells).

SSQ Total score ratings pre/post change	Disparity Limit (arc min)						Average in the 3D conditions (nonzero)
	0	20	40	60	80	100	
<b>Participant</b>	<b>0</b>	<b>20</b>	<b>40</b>	<b>60</b>	<b>80</b>	<b>100</b>	
<b>1</b>	2	0	3	0	0	0	<b>0.6</b>
<b>2</b>	0	9	0	5	5	-2	<b>3.4</b>
<b>3</b>	3	1	1	0	1	0	<b>0.6</b>
<b>4</b>	3	-2	-2	1	1	3	<b>0.2</b>
<b>5</b>	1	7	3	0	3	1	<b>2.8</b>
<b>6</b>	2	8	16	-2	13	13	<b>9.6</b>
<b>7</b>	0	0	0	0	0	0	<b>0.0</b>
<b>8</b>	1	3	5	2	2	4	<b>3.2</b>
<b>9</b>	-1	0	0	0	0	0	<b>0.0</b>
<b>10</b>	0	0	0	0	0	0	<b>0.0</b>
<b>11</b>	0	0	0	0	1	0	<b>0.2</b>
<b>12</b>	3	5	3	2	5	0	<b>3.0</b>
<b>Average</b>	<b>1.2</b>	<b>2.6</b>	<b>2.4</b>	<b>0.7</b>	<b>2.6</b>	<b>1.6</b>	.
<i>t-value</i>		1.69	1.43	0.31	1.79	0.97	1.46
<i>p-value</i>		.12	.18	.76	.10	.35	.17

**Table 13. Partial Correlations between the Objective Measures of Eyestrain, Postural Instability, the Simulator Sickness Questionnaire, and Disparity Limits in Experiment Two**

In this table, partial correlations controlling for participant and session (practice) effects were conducted. One-tailed significance tests were performed with a sample size of 72 (12 participants, 6 sessions each). Pre/post session changes in the SSQ, including the Nausea, Oculomotor, and Disorientation subscales, the Total Score, and Disparity Limits were correlated with the pre/post session changes in the objective eyestrain measurements. Correlations greater than +/- .195 in the suspected direction were significant at the .05 level, as highlighted in grey. Given the number of statistical tests conducted, if a more conservative critical value for alpha of .01 were utilized (instead of .05), none of these tests would have achieved significance.

<b>Sim Sickness Questionnaire</b>	<b>Lateral Phoria Near</b>	<b>Lateral Phoria Far</b>	<b>Vertical Phoria Far</b>	<b>Fusion Near Point</b>	<b>Fusion Far Point</b>	<b>Fusion Range</b>	<b>Postural Instability</b>
Nausea (subscale)	-.11	-.14	.00	-.07	.21	-.12	.00
Oculomotor (subscale)	-.03	-.13	-.03	-.05	.24	-.11	-.19
Disorientation (subscale)	-.06	-.02	-.02	-.07	.15	-.11	.01
Total SSQ Score	-.06	-.12	-.02	-.07	.23	-.12	-.08
Disparity Limit manipulation	.02	-.16	.23	.14	-.05	.15	.08

**Table 14. Participants' Performance with 2D versus S3D Display in Experiment Two**

In this table, mean positional error (in virtual inches) and rotational error (in degrees) are reported for each participant in both the 2D and S3D display conditions, and the percentage improvement when changing from 2D to S3D. Each reported value represents the mean of all trials in which either S3D cues were present in any magnitude (S3D Display), or there were none (2D Display).

<b>Participant</b>	<b>2D Display Positional Errors</b>	<b>S3D Display Positional Errors</b>	<b>% Reduction in Positional Errors with S3D Cues</b>	<b>2D Display Rotational Errors</b>	<b>S3D Display Rotational Errors</b>	<b>% Reduction in Rotational Errors with S3D Cues</b>
1	2.31	0.25	89	17.38	14.52	16
2	0.33	0.11	67	13.43	8.09	40
3	1.51	0.13	91	16.09	12.43	23
4	0.48	0.09	81	10.12	5.50	46
5	0.28	0.13	54	13.93	10.61	24
6	1.26	0.18	86	16.54	7.52	55
7	1.36	0.15	89	12.04	8.91	26
8	0.14	0.13	7	3.61	3.43	5
9	2.13	0.15	93	17.04	12.33	28
10	0.54	0.09	83	13.31	8.07	39
11	1.87	0.21	89	17.60	13.69	22
12	0.39	0.12	69	11.40	9.97	13
<b>Average</b>	<b>1.05</b>	<b>0.15</b>	<b>86</b>	<b>13.54</b>	<b>9.59</b>	<b>29</b>

**Table 15. Correlations between USAFSAM Pre-screening Optometric Data and S3D Performance in Experiment Two.**

In this table, correlations between the USAFSAM optometric screening data and performance on the S3D display are reported. Performance for each subject was summarized into two measures (positional errors and rotational errors in S3D trials), which are presented as separate columns. One-tailed significance tests were performed with a sample size of 12. Correlations of +/- .50 or larger in the suspected direction were significant at the .05 level, as highlighted in grey. Stars (\*) indicate values that are also significant at the .01 level to account for multiple tests.

<b>USAFSAM Optometric Measurements</b>	<b>Positional Error</b>	<b>Rotational Error</b>
Refractive Error (right eye)	.29	.19
Refractive Error (left eye)	.28	.42
Horizontal Phoria (distance)	.11	-.04
Vertical Phoria (distance)	-.32	-.55
Fusion Range (distance) – Base-In Break	-.08	-.04
Fusion Range (distance) – Base-In Recovery	-.14	-.12
Fusion Range (distance) – Base-Out Break	-.72*	-.68*
Fusion Range (distance) – Base-Out Recovery	-.47	-.81*
Fusion Range (distance)	-.68*	-.83*
Horizontal Phoria (near)	.07	-.39
Vertical Phoria (near)	-.26	-.61
Fusion Range (near) – Base-In Break	-.25	-.26
Fusion Range (near) – Base-In Recovery	-.09	-.11
Fusion Range (near) – Base-Out Break	-.31	-.37
Fusion Range (near) – Base-Out Recovery	-.19	-.42
Fusion Range (near)	-.52	-.71*
Stereoacuity threshold	.12	.54

**Table 16. Correlations between the Pre-session Optometric Data and S3D Performance in Experiment Two.**

In this table, correlations between the averaged pre-session optometric data and performance on the S3D display are reported. Performance for each subject was summarized into two measures (positional errors and rotational errors in S3D trials), which are presented as separate columns. One-tailed significance tests were performed with a sample size of 12. Correlations of +/- .50 or greater in the suspected direction were significant at the .05 level, as highlighted in grey. Given the number of statistical tests conducted, if a more conservative critical value for alpha of .01 were utilized (instead of .05), none of these tests would have achieved significance.

<b>Pre-session Optometric Measurements</b>	<b>Positional Error Correlation</b>	<b>Rotational Error Correlation</b>
Lateral Phoria (near)	-.16	-.46
Lateral Phoria (far)	-.22	-.38
Vertical Phoria (far)	-.19	-.09
Fusion Near Point (Near Point of Convergence)	-.55	-.40
Fusion Far Point (Far Point of Convergence)	.35	.61
Fusion Range	-.57	-.50

**Table 17. Distinguishing Optometric Measurements of the Groupings in the Topological Exploratory Data Analysis.**

In this table, the top seven distinguishing optometric measurements are provided in rank order, according to the signed Kolmogorov-Smirnov two-sample test. Four of the seven measures relate to fusion ranges, the three others relate to horizontal phorias. Whether the variable was recorded by the experimenter in pre-session testing or by the USAFSAM optometrists in the clinic is indicated.

<b>Type of Measure</b>	<b>Variable</b>	<b>Signed KS score</b>
Fusion Range	Fusion Far Point (pre-session)	-0.889
Fusion Range	Near Fusion Range Base-In Breakpoint (USAFSAM)	0.786
Phoria (horizontal)	Near Phoria in Horizontal (USAFSAM)	0.746
Phoria (horizontal)	Lateral Phoria Near (pre-session)	0.706
Phoria (horizontal)	Lateral Phoria Far (pre-session)	0.675
Fusion Range	Distance Fusion Range Base-In Recovery point (USAFSAM)	0.675
Fusion Range	Near Fusion Range Base-In Recovery point (USAFSAM)	0.675

**Table 18. Comparing Exophoria, Esophoria, and Orthophoria Groups on the Optometric Measurements and their Relation to S3D Discomfort**

In this table, the reported values represent mean values of the subgroups. Phorias were measured clinically while fusion range measures were measured on the S3D display.

<b>Measure</b>	<b>Esophoria (n=9)</b>	<b>Ortho (n=9)</b>	<b>Exophoria (n=6)</b>
Horizontal Phoria at Distance (PD)	-1.94	0	+3.00
Horizontal Phoria at Near (PD)	+1.39	+2.61	+6.08
Fusion Near Point on S3D	1.32	2.80	1.59
Fusion Far Point on S3D	0.93	1.24	1.39
Fusion Range Total on S3D	2.25	4.04	2.98
S3D Discomfort Ratings (SSQ)	2.80	2.20	2.60

## Attachment 1. The Simulator Sickness Questionnaire (SSQ)

# -----SSQ----- SIMULATOR SICKNESS QUESTIONNAIRE

OBSERVER NUMBER: \_\_\_\_\_

DISPLAY: \_\_\_\_\_ CAMERA SEPARATION: \_\_\_\_\_

PRE/POST-TEST: \_\_\_\_\_

Please answer the following 16 questions by circling the number corresponding to your answer:

	<u>None</u>	<u>Low</u>	<u>Moderate</u>	<u>High</u>
General discomfort	0	1	2	3
Fatigue	0	1	2	3
Headache	0	1	2	3
Eyestrain	0	1	2	3
Difficulty focusing	0	1	2	3
Increased salivation	0	1	2	3
Sweating	0	1	2	3
Nausea	0	1	2	3
Difficulty concentrating	0	1	2	3
Fullness of head	0	1	2	3
Blurred vision	0	1	2	3
Dizzy (eyes open)	0	1	2	3
Dizzy (eyes closed)	0	1	2	3
Vertigo	0	1	2	3
Stomach awareness	0	1	2	3
Burping	0	1	2	3

## Attachment 2. Demographic Questionnaire

MCINTIRE  
AFRL / 711 HPW / RHCV

Observer number \_\_\_\_\_

Age \_\_\_\_\_ Gender \_\_\_\_\_ Dominant Eye \_\_\_\_\_ Dominant Hand \_\_\_\_\_

IPD \_\_\_\_\_ Visual Acuity \_\_\_\_\_ 3D Vision Test \_\_\_\_\_

Circle one: 20/20 Vision Corrected to 20/20  
Circle one: Glasses Contacts Neither

Hours per week playing video games:

If applicable, what type of video games do you play?

If applicable, what type of video game system do you use? Circle one:

Computer TV Portable

Hours per week watching TV:

Hours per week on a computer:

Is computer use for (circle):

Work Leisure Gaming

Have you ever viewed a 3D display before (watched a 3D movie, show, or played a 3D game)?

If so, what type of 3D (if known)?

Red/Blue Flickering Polarized Other

If applicable, how many 3D movies/shows/or games have you viewed/played?

Very few (1-2) Some (3-5) Many (6 or more)

If so, what was your impression of the technology?

Terrible Not great No opinion I Liked it I Loved it!

Why? \_\_\_\_\_

Do you get motion sickness? Yes / No

If so, from what (circle all that apply):

Planes Trains Automobiles  
Boats Rollercoasters

Do you get migraines regularly? Yes / No If so, how regularly? \_\_\_\_\_

About how many times in your life have you gotten food poisoning? \_\_\_\_\_

### **Attachment 3. Recruitment E-mail (“Invitation to Participate”)**

**To: AFRL DL-WRS Personnel**

**Subject:** Study Participation Opportunity

Good Morning,

The researchers in RHCV are interested in testing user performance on one of our 3D stereoscopic displays. I am looking for volunteers to take part in a study to collect subjective and performance data in relation to the amount of virtual camera separation that is displayed. The time requirement for each volunteer subject is anticipated to be a total of 6 sessions lasting approximately 40 minutes each, with a maximum of 2 sessions per day. No monetary compensation is being offered for participation. I have slots for 12 subjects available right now, and will take volunteers on a first come, first serve basis. Let me know if you are interested and I can send you more information. Thank you for your time.

V/R,

John McIntire

---

John McIntire  
Engineering Research Psychologist (DR-II)  
Battlespace Visualization Branch  
Warfighter Interface Division, 711 HPW/RHCV  
2255 H Street  
Wright-Patterson AFB, OH 45433-7022  
john.mcintire@wpafb.af.mil  
DSN: 785-0589  
(937) 255-0589

## Attachment 4. Summary of Procedures and Data Collection

### 1. Pre-experimental Screening and Tests

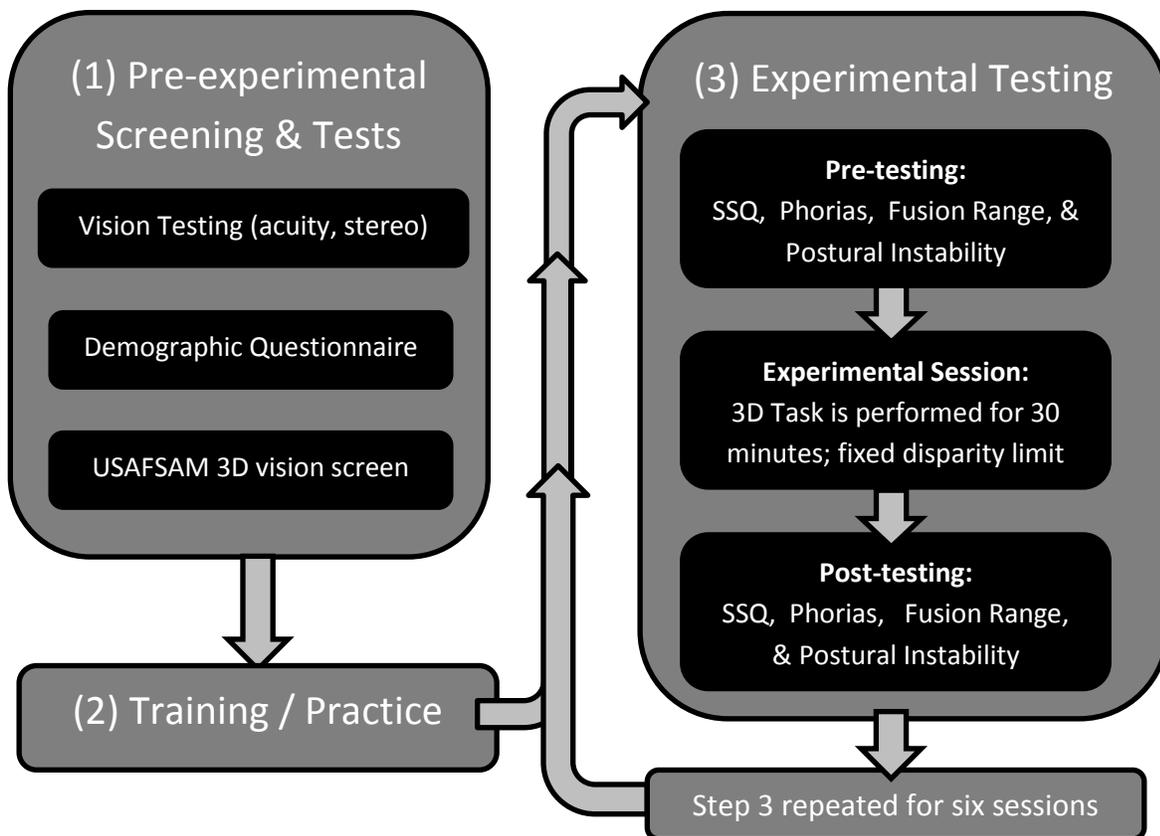
- Demographic Questionnaire (Attachment 2)
- Normal distance acuity testing (Snellen eye chart)
- Normal stereo vision testing (Keystone View Telebinocular, Stereo Fly)
- USAFSAM optometrists' collaborative binocular vision screening protocol (including refraction measurement, fusional ranges, phorias, and fixation disparities)

### 2. Training

- Several minutes of warm-up (practice trials), verification of stereo depth perception

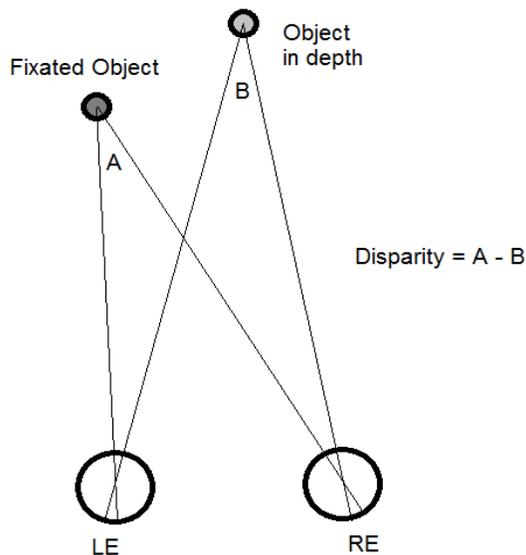
### 3. Experimental Testing (per session, each test occurs both before & after the experimental session; the presentation order of the pre and post tests were randomized across sessions for each participant))

- Simulator Sickness Questionnaire (SSQ; Attachment 1)
- Near and Far Lateral Phoria (measured via Keystone Telebinocular visual testing apparatus)
- Fusion Range (measured via 3D computer display)
- Postural Instability Test (via Wii Balance Board)



## Appendix 1. Calculations of Disparity

The term “disparity” in the context of 3D display systems can refer to two slightly different but related concepts. In the **vision science community**, *binocular disparity* or *interocular disparity* or *retinal disparity* or *geometric disparity* typically refers to the angular difference projected on the retinas between two different points in space, one point being the fixation point where the eyes meet, and the other usually being an object of interest at some depth other than that of the fixation point (e.g., Cho & Kang, 2012; De Silva, Fernando, Worrall, Arachchi, & Kondo, 2011; Leroy, Fuchs, Moreau, 2012; Wang, Barkowsky, Ricordel, & Le Callet, 2011; Hsu, Pizlo, et al., 1996). In the **display community**, the terms *horizontal disparity*, *on-screen disparity*, *display disparity*, *lateral disparity*, or *half-image separation* typically refer to the physical image separation of stereo-pair images as viewed on a stereoscopic display (e.g., Howard & Rogers, 2002; Poyade, Reyes-Lecuona, & Viciano-Abad, 2009). This measure is typically provided in units of mm or pixels, and is also the same concept as image parallax.



It may not be immediately obvious where this relationship  $\text{Disparity} = A - B$  comes from, or why the difference between Angle A and Angle B is an important or useful metric. Part of the answer comes from the fact that vision researchers are interested in the angular extent of stimuli that are projected on the retinas because these can be easily related to photoreceptor spacings, to ganglion receptive field sizes, to optics measurements, and to more complex psychophysical relationships. In the figures above, the left diagram shows the definition of angular disparity and how it is most commonly calculated (as the difference between angles A and B).

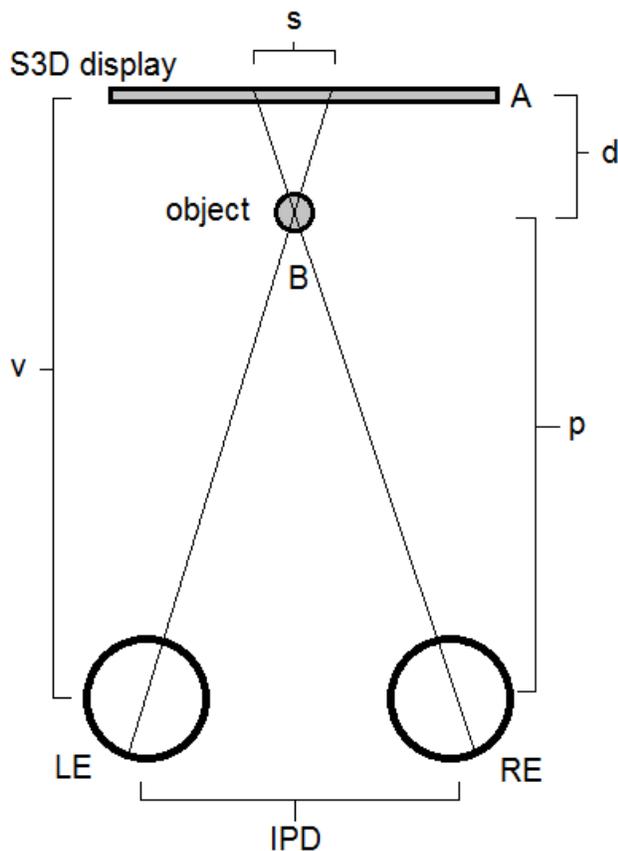
Notice that if the non-fixated object B is closer to the viewer than the fixated object A, angle B will be larger than angle A, and binocular disparity will be a negative value, indicating ‘crossed disparity’ or ‘negative parallax.’ Alternatively, if the non-fixated object B is farther from the viewer than object A, binocular disparity will be a positive value, indicating ‘uncrossed disparity’ or ‘positive parallax.’

We can move from considering disparities between two objects or points in space, to the situation in which we are considering the disparity of a virtual object B “in depth” relative to the display surface (A), as is the case for stereoscopic display system viewing. On-screen disparities (in units of length like mm or pixels) are ultimately related to binocular disparities (in units of visual angle) by viewing distance, viewer eye separation (IPD), and virtual depths of object(s). These calculations generally assume the object lies along the median plane of the viewer and display, when the viewer’s line-of-sight is perpendicular to the display surface, etc.

To estimate the **Binocular Disparity (D)** for a virtual point/object (B) relative to a stereo display surface (A):

$$D = A - B$$

$$D = 2 * \arctan \frac{IPD}{2*v} - 2 * \arctan \frac{IPD}{2*p}$$



You need to first measure/estimate:

**IPD = inter-pupillary distance** of the viewer (or assume an average, ~60-65 mm is common).

**v = viewing distance;** the distance between the viewer and display.

**p = point or object distance;** the distance between the viewer and virtual point/object.

**s = on-screen half-image separation;** by convention, positive values indicate far depth, negative values indicate near.

**d = v - p; object depth relative to display;** by convention, positive values indicate near depth, negative values indicate far.

\* If arctan function is in radians (as in Excel or SPSS), the conversion is 1 radian = 57.2957795 degrees. Disparity angles are typically reported in degrees, min, or sec of arc.

If the distance of a virtual object (**p**) is not easily determinable, but on-screen disparities (**s**) are measurable, then binocular disparity of a virtual object relative to the display surface can be easily calculated by plugging one of the following terms into the above equation for D, due to congruent triangles: [s, d] and [IPD, p] (Holliman, 2004).

$$p = \frac{-d * IPD}{s} = \frac{(p-v) * IPD}{s} = \frac{v * IPD}{IPD - s}$$

This gives a modified equation for D that utilizes only real-world-measurable values of **IPD**, **s**, and **v**:

$$D = 2 * \arctan \frac{IPD}{2*v} - 2 * \arctan \frac{IPD - s}{2*v}$$

As suggested by Hsu, Pizlo, et al (1996), binocular disparity can also be **estimated** by measuring on-screen disparities and then using the following definition of visual angle (VA), but be sure to use the same units for on-screen disparity (s) and viewing distance to display (v). Keep in mind that this alternative estimate (D') departs from D at extremes in viewing distances or objects depths, and  $D = A - B$  remains the preferred calculation, but for most general purposes this estimate works very well.

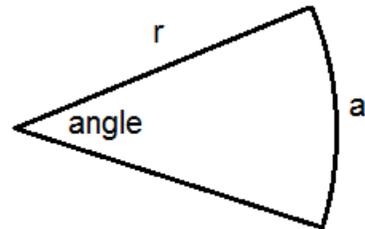
$$D' \approx VA = 2 * \arctan \frac{s}{2 * v}$$

Putting **s** (on-screen image separations) alone on one side of either equations involving D or D' can allow an experimenter to **compare predictions of on-screen disparity with actual measurements** using the appropriate assumed/measured values. Doing so can be useful for verifying that manipulations of object depths in a real or virtual camera/scene space correspond as desired to the actual viewer/display space.

$$s = \frac{IPD*(-d)}{p} = \frac{IPD * (p-v)}{p} \quad \text{(derived from equations above)}$$

$$s' = 2 * v * \tan \frac{D}{2} \quad \text{(estimate; derived from VA calculation)}$$

Other commonly seen formulas for binocular disparity (D), typically for estimating observers' stereoacuity, utilize circles (radians) instead of right triangle geometry (e.g., Stidwell & Fletcher, 2011). The advantage of using radians is the formulas are a bit simpler, but with the disadvantage that these estimates may give slightly different disparities from D and D' above, and may deviate more noticeably with extreme values. We start with the definition of a radian as the angle subtended by an arc of length (**a**), divided by the length of the radius of the circle (**r**):



$$\text{angle (in radians)} = \frac{a}{r}$$

Now Angle A can be estimated by letting the eye separation (**IPD**) be an estimate for the arc length (**a**), and letting the viewing distance to display (**v**) be an estimate for the radius (**r**). Likewise, Angle B can be estimated using **IPD** and the distance to the point or object of interest (**p**):

$$D = A - B \quad A' = \frac{IPD}{v} \quad B' = \frac{IPD}{p} = \frac{IPD}{v-d}$$

$$D'' = \frac{IPD}{v} - \frac{IPD}{v-d} = \frac{IPD(v-d)}{v(v-d)} - \frac{IPD*v}{v(v-d)} = \frac{-IPD*d}{v(v-d)} \dots$$

$$D'' = -\frac{IPD * (v-p)}{v * p} \quad \text{OR} \quad -\frac{IPD * d}{v^2 - v*d}$$

When the value of relative depth (**d**) is very small (such as when estimating fine stereoacuity thresholds), the term (**v\*d**) in the denominator approaches zero and is often dropped to reach an even simpler formula, for a somewhat poorer estimate. This is essentially the same formula commonly seen in many optometry textbooks for measuring an observer's stereoacuity in radians (to convert to arc sec, multiply by 206,265):

$$D''' = -\frac{IPD * d}{v^2} \quad \text{OR} \quad -\frac{IPD * (v-p)}{v^2}$$

More complex geometrical analysis may be needed for alternative setups and/or if higher precision is appropriate. Please see the references cited in this appendix, or see Williams & Parrish (1990), Ware, Gobrecht, & Paton (1998), Jones, Lee, Holliman, & Ezra (2001), Kim, Choi, & Sohn (2011), or Grinberg, Podar, & Siegel (1994).

## Appendix 2. Measures and Predictors of Eyestrain/Fatigue on Stereoscopic 3D Displays

The following measures were gleaned from the 3D display, human factors, and vision science literature. This is not a comprehensive list. These represent measures that were actually used or only were suggested as potentially useful for characterizing and/or predicting eyestrain and fatigue on stereoscopic 3D displays.

### Optometric

- acuity (Häkkinen et al., 2006; Mon-Williams, Wann, & Rushton, 1993; Tourancheau et al., 2012)
- stereoscopic or binocular acuity (Kooi & Toet, 2004; Barkowsky, Cousseau, & Le Callet, 2011; Chao et al., 2012; Fortuin et al., 2010; Häkkinen et al., 2006; Hoffman et al., 2008; Neveu et al., 2010; Rushton, Mon-Williams, & Wann, 1994; Tourancheau et al., 2012; Stone, Watts, Rosenquist, 2012)
- phorias (Kooi & Toet, 2004; Häkkinen et al., 2006; Fortuin et al., 2011; Howarth, 1999; Rushton et al., 1994; Lambooi et al., 2011; Lambooi et al., 2012; Neveu et al., 2010; Judge & Miles, 1985; Fisher & Ciuffreda, 1990; Bobier & MacRae, 1996; Mon-Williams, Wann, & Rushton, 1993; Vienne, Blondé, & Doyen, 2012)
- fixation disparity (Fortuin et al., 2010; Lambooi et al., 2011; Vienne, Blondé, & Doyen, 2012)
- amount and velocity of the step ocular convergence (Sharples et al., 2008)
- accommodation response (Yano et al., 2004; Emoto et al., 2005; Fortuin et al., 2011; Inoue & Ohzu, 1997; Lambooi et al., 2011; Yang & Sheedy, 2011; Yano et al., 2004)
- esotropia (Tsukuda & Murai, 1988)
- contrast sensitivity (Chao et al., 2012)
- critical flicker frequency (Mitsubishi, 1996; Chao et al., 2012)
- AC/A ratio (Emoto et al., 2004; Wann & Mon-Williams, 2002; Neveu et al., 2010; Judge & Miles, 1985; Miles et al., 1987; Bobier & MacRae, 1996)
- CA/C ratio (Wann & Mon-Williams, 2002; Miles et al., 1987)
- fusional amplitude/range/limits (Emoto et al., 2004; Lambooi et al., 2011; Nojiri et al., 2004; Fortuin et al., 2011; Lambooi et al., 2011; Kim, Choi, & Sohn, 2011; Neveu et al., 2010)
- fusional reserves (Fortuin et al., 2011; Lambooi et al., 2012; Neveu et al., 2010)
- near point of accommodation (Häkkinen et al., 2006; Wee & Moon, 2014)
- vergence/prism facility (Fortuin et al., 2011; Lambooi et al., 2011)
- foveal suppression (Lambooi et al., 2011)
- accommodation facility (Lambooi et al., 2011)
- accommodative amplitude/load (Omori et al., 2005; Rushton, Mon-Williams, & Wann, 1994; Wee & Moon, 2014)
- near point of convergence (Mon-Williams, Wann, & Rushton, 1993)
- positive accommodation velocity (Suzuki, Onda, Katada, Ino, & Ifukube, 2004)
- vergence response (Yang & Sheedy, 2011)
- fusion response time (Hoffman et al., 2008; Kim, Choi, & Sohn, 2011; Kim, Choi, Park, & Sohn, 2011)
- inter pupillary distance (Rushton, Mon-Williams, & Wann, 1994)

- pupil size/pupillary reflex (Chao et al., 2012; Oyamada et al. 2007; Inoue & Ohzu, 1997; Rushton, Mon-Williams, & Wann, 1994)
- eye blink rates (Cho & Kang, 2012; Kim, Choi, Park, & Sohn, 2011; Barkowsky & Le Callet, 2010; Vienne, Blondé, & Doyen, 2012; Wee & Moon, 2014)
- binocular intraocular pressure (Chao et al., 2012)
- binocular high frequency components of accommodative micro fluctuations (Chao et al., 2012; Vienne, Blondé, & Doyen, 2012)
- ocular protection index (OPI; Wee & Moon, 2014)

#### Subjective

- Simulator Sickness Questionnaire (SSQ) (Häkkinen et al., 2006; Hale & Stanney, 2006; Vienne, Blondé, & Doyen, 2012)
- Convergence Insufficiency Symptom Survey (CISS) (Chao et al., 2012; Fortuin et al., 2011; Lambooij et al., 2011; Lambooij et al., 2012)
- Interpretation Based Quality (IBQ) methodology (Hirahara, Shiraishi, & Kawai, 2012)
- modified Single Stimulus Continuous Quality Evaluation (SSCQE) methodology (Sohn et al., 2012; Speranza et al., 2006; Noriji et al., 2004; Yano et al., 2004)
- modified Double Stimulus Continuous Quality Evaluation (DSCQE) methodology (Tam et al., 2011)
- non-standard comfort/quality ratings (Kooi & Toet, 2004; Speranza et al. 2006; Emoto et al., 2005; Hoffman et al., 2008; IJsselsteijn, de Ridder, & Vliegen, 2000; Lambooij et al., 2007; Neveu et al., 2010; Wöpking, 1991; Pastoor, 1993; Rushton, Mon-Williams, & Wann, 1994; Shibata et al., 2011a/b; Tam, Vazquez, & Speranza, 2007; Kuze & Ukai, 2008; Tourancheau et al., 2012; Yang & Sheedy, 2011; Yano et al., 2004; Wee & Moon; 2014)

#### Physiologic/Other

- cross-correlation from blood pressure to heart rate (Yoshizawa et al., 2001, 2004)
- visual evoked cortical potentials (Emoto et al., 2005)
- event-related potentials (Li, Seo, Kham, & Lee, 2008)
- heart rate (Hirahara, Shiraishi, & Kawai, 2012; Kim, Yoshitake, Morikawa, Kawai, Yamada, & Iguchi, 2011)
- cerebral oxygenation (Kim, Yoshitake, Morikawa, Kawai, Yamada, & Iguchi, 2011)
- Wilkins Rate of Reading test (WRRT) (Lambooij et al., 2011)
- Real-world depth perception tasks (timed precision placement; Stone, Watts, & Rosenquist, 2012)
- cardio-vascular reflex (Oyamada et al., 2007).

### Appendix 3. Follow-up Testing for Atypical Observers in Experiment One

Here we describe a series of brief follow-up tests that were conducted on the atypical participants in Experiment One, in an attempt to explain their results:

- **Stereoacuity threshold measure.** This test takes only minutes to administer and utilizes an adaptive staircase thresholding technique (QUEST) to arrive at an accurate estimate of a viewer's distance stereoacuity. The accuracy is on the order of seconds of arc (custom in-house developed USAFSAM software). The purpose of this test is to get a measure of how accurate each viewer's stereovision can be at the very limits of their perception.
- **Randot stereotest.** The Randot is a similar booklet to the Titmus that tests stereoscopic acuity from 400 down to 20 arc seconds, but which cannot be passed monocularly at any disparity level (Fricke & Siderov, 1997), unlike the Titmus test.
- **Static suppression test.** There is a static image binocular suppression test in the Randot test booklet, to determine whether stereo imagery is causing one eye's view to be inadvertently suppressed.
- **Forced binocularity task.** A brief experimental session (nearly identical to Experiment One) will be replicated using inconsistent size and/or texture cues (2D versus S3D), lasting about 5 minutes total. This test will determine whether the participants will utilize the stereoscopic 3D cues to depth if forced to, due to task demands. In this case, the monocular cues to depth, including size and texture, will be inconsistent across trials and will therefore help little or not at all for completing the task. Basically, the only useful cue would be stereopsis, and it should be abundantly obvious from the data whether or not they were able to utilize the S3D cues.
- **USAFSAM suppression testing.** Dr. Steven Wright, an optometrist from the USAFSAM OBVA Lab, will do some additional follow-up testing that may shed light on this issue, using the following tests. These tests might be utilized in conjunction with either prism or lenses to potentially de-link the accommodation and vergence systems.
  - **AO (American Optical) suppression test.** This test presents a line of letters on an eye chart with some letters seen only by the right eye, some letters seen only by the left eye, and some letters are seen by both eyes. While similar to the

suppression test stimuli on the Stereo Fly, the AO test can present much smaller and precise visual targets which are more prone to suppression.

- **Worth light.** A red/green flashlight which, again, has some lights visible only to the right/left eye and one common light, allowing suppression to be investigated.
- **Base out prism.** This is an objective test in which the clinician presents a base out prism over one eye and observes whether the subject demonstrates re-fixation of the opposite eye.

**Appendix 4. Simulator Sickness Questionnaire Pre-test Results across Disparity Limits in Experiment Two**

In this table, the raw data concerning SSQ Total scores (pre-session values) across disparity limits are reported, by participant and as an average across the sessions. The pre-to-post session changes are given in Table 12; this table is provided as a comparison.

SSQ Total score ratings pre-session scores	Disparity Limit (arc min)						Average across the conditions
	Participant	0	20	40	60	80	
<b>1</b>	6	3	12	2	2	4	<b>4.8</b>
<b>2</b>	6	3	1	0	0	6	<b>2.7</b>
<b>3</b>	0	0	0	0	0	0	<b>0.0</b>
<b>4</b>	4	5	5	0	0	0	<b>2.3</b>
<b>5</b>	1	1	3	7	0	0	<b>2.0</b>
<b>6</b>	0	0	0	8	0	0	<b>1.3</b>
<b>7</b>	0	0	0	0	0	0	<b>0.0</b>
<b>8</b>	2	2	1	2	1	0	<b>1.3</b>
<b>9</b>	1	0	0	0	0	0	<b>0.2</b>
<b>10</b>	0	0	0	0	0	0	<b>0.0</b>
<b>11</b>	0	0	0	0	0	0	<b>0.0</b>
<b>12</b>	0	0	0	1	0	0	<b>0.2</b>

## Appendix 5. Publications derived wholly or in part from this dissertation.

### Journal/proceedings publications:

McIntire, J. P., Havig, P. R., & Geiselman, E. (2012). "What is 3D good for? A review of the human performance benefits of stereoscopic 3-D displays." *Proceedings of SPIE*, 8383, 83830X.

McIntire, J. P., Havig, P. R., & Geiselman, E. (2014). "Stereoscopic 3D displays and human performance: A comprehensive review." *Displays*, 35(1), pp. 18-26.

McIntire, J. P., Wright, S. T., Harrington, L. K., Havig, P. R., Watamaniuk, S. N. J., & Heft, E. (2014). "Optometric measurements predict performance but not comfort on a virtual object placement task with a stereoscopic 3D display." *Optical Engineering*, Special Issue on Human Vision, 53(6), p. 061711-1.

### In-progress manuscripts:

McIntire, J. P. Havig, P. R., Harrington, L. K., Wright, S. T., Watamaniuk, S. N. J., & Heft, E. (unpublished manuscript). "The effect of microstereopsis on performance and comfort: Precision alignment tasks with a stereoscopic 3D display."

McIntire, J. P. Havig, P. R., Harrington, L. K., Wright, S. T., Watamaniuk, S. N. J., & Heft, E. (unpublished manuscript). "Clinically normal stereopsis does not ensure a performance benefit from stereoscopic 3D depth cues."

McIntire, J. P., Heft, E., Havig, P. R., & Harrington, L. K. (unpublished manuscript). "Calculations of binocular disparity and the implications for stereoscopic 3D displays."

## REFERENCES

- Aitsiselmi, Y. & Holliman, N. S. (2009). "Using mental rotation to evaluate the benefits of stereoscopic displays." *Proceedings of SPIE*, 7237, 72370Q.
- American Optometric Association; AOA.org (2010). "3-D TV and movies look to attract viewers but not everyone can 'see' what all the hype is about." [digital press release] on May 18, 2010. Available: <http://www.aoa.org/x15123.xml>
- Appelle, S. (1972). "Perception and discrimination as a function of stimulus orientation: The 'oblique effect' in man and animals." *Psychological Bulletin*, 78(4), pp. 266-278.
- Banks, M. S., Akeley, K., Hoffman, D. M., & Girshick, A. R. (2008). "Consequences of incorrect focus cues in stereo displays." *Journal for the Society of Information Display*, 24(7), p. 7-11.
- Barkowsky, M., Cousseau, R., & Le Callet, P. (2011). "Is visual fatigue changing the perceived depth accuracy on an autostereoscopic display?" *Proceedings of SPIE*, 7863, 78631V.
- Blake, R. & Wilson, H. (2011). "Binocular Vision." *Vision Research*, 51, pp. 754-770.
- Bobier, W. R., & McRae, M. (1996). "Gain changes in the accommodative convergence cross-link." *Ophthalmic & Physiological Optics*, 16, 318-325.
- Bosco, A., Longoni, A. M., & Vecchi, T. (2004). "Gender effects in spatial orientation: Cognitive profiles and mental strategies." *Applied Cognitive Psychology*, 18(5), pp. 519-532.
- Camposeco, F., Avilés, C., Careaga, B., Spindola, L., & Velázquez, R. (2011). "Constraints on human stereo vision for tele-operation." *Proceedings of the 2011 IEEE IX Latin American and IEEE Colombian Conference on Automatic Control and Industry Applications (LARC)*, October 1-4, 2011, pp. 1-6.
- Chao, W.-C., Chen, C.-W., Dong, R.-L., & Chen, Y.-S. (2012). "An ergonomic evaluation of stereoscopic and deadzone-free autostereoscopic three-dimensional displays." *Society for Information Display SID Symposium Digest of Technical Papers*, 43(1), June, pp. 887-890.
- Chen, J. Y. C., Oden, R. V. N., Drexler, J. M., & Merritt, J. O. (2010). "Evaluation of stereoscopic display for indirect-vision driving and robot teleoperation." *The WSTIAC Quarterly*, 10(2), pp. 1-9.
- Chen, J. Y. C., Oden, R. V. N., & Merritt, J. O. (2014). "Utility of stereoscopic displays for indirect-vision driving and robot teleoperation." *Ergonomics*, 57(1), pp. 12-22.
- Cho, S.-H., & Kang, H.-B. (2012). "The measurement of eyestrain caused from diverse binocular disparities, viewing time and display sizes in watching stereoscopic 3D

- content.” Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 16-21 June 2012, pp. 23-28.
- Clark, R., Bryant, A., Pua, Y., McCrory, P., Bennell, K., et al. (2010). “Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance.” *Gait and Posture*, 31(3), pp. 307-310.
- Cobb, S. V. G., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). “Virtual reality-induced symptoms and effects (VRISE).” *Presence: Teleoperators and Virtual Environments*, 8(2), pp. 169-186.
- De Silva, V., Fernando, A., Worrall, S., Arachchi, H. K., & Kondoz, A. (2011). “Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-D displays.” *IEEE Transactions on Multimedia*, 13(3), pp. 498-506.
- Ehrlich, J. A. (1997). “Simulator sickness and HMD configurations.” Proceedings of SPIE, 3206, pp. 170-178.
- Ellis, S. R., Fishman, J. M., Hasser, C. J., & Stern, J. D. (2005). “Effect of reduced stereoscopic camera separation on ring placement with a surgical telerobot.” Proceedings of SPIE, 5664, pp. 372-379.
- Fielder, A. R., & Moseley, M. J. (1996). “Does stereopsis matter in humans?” *Eye*, 10, pp. 233-238.
- Fisher, S. K., & Ciuffreda, K. J. (1990). “Adaptation to optically-increased interocular separation under naturalistic viewing conditions.” *Perception*, 19, pp. 171-180.
- Fortuin, M. F., Lambooi, M. T., IJsselsteijn, W. A., Heynderickx, I., Edgar, D. F., & Evans, B. J. W. (2011). “An exploration of the initial effects of stereoscopic displays on optometric parameters.” *Ophthalmic and Physiological Optics*, 31, pp. 33-44.
- Fricke, T. R., & Siderov, J. (1997). “Stereopsis, stereotests, and their relation to vision screening and clinical practice.” *Clinical and Experimental Optometry*, 80(5), pp. 165-172.
- Froner, B. (2011). “Stereoscopic 3D technologies for accurate depth tasks: A theoretical and empirical study.” Durham thesis, Durham University. Available online at: <http://etheses.dur.ac.uk/3324/>
- Fujisaki, H., Yamashita, H., Kihara, K., & Ohtsuka, S. (2012). “Individual differences in the use of binocular and monocular depth cues in 3D-graphics environments.” SID 2012 Digest, pp. 1190-1193.
- Grinberg, V. S., Podnar, G., & Siegel, M. W. (1994). “Geometry of binocular imaging,” Proceedings of the IST/SPIE Symposium on Electronic Imaging Stereoscopic Displays and Applications 2177.
- Häkkinen, J., Pölönen, M., Takatalo, J., & Nyman, G. (2006). “Simulator sickness in virtual display gaming: A comparison of stereoscopic and non-stereoscopic situations.”

- Proceedings of the 8<sup>th</sup> Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'06), ACM New York, pp. 227-229.
- Hale, K. S., & Stanney, K. M. (2006). "Effects of low stereo acuity on performance, presence and sickness within a virtual environment." *Applied Ergonomics*, 37, pp. 329-339.
- Hermens, F. & Gielen, S. (2003). "Visual and haptic matching of perceived orientations of lines." *Perception*, 32(2), pp. 235-248.
- Hettinger, L., & Riccio, G. (1992). "Visually induced motion sickness in virtual environments." *Presence*, 1(3), pp. 306-310.
- Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue." *Journal of Vision*, 8(3), pp. 1-30.
- Holliman, N. (2004). "Mapping perceived depth to regions of interest in stereoscopic images." Proceedings of SPIE, Stereoscopic Displays and Applications XV.
- Howard, I. P. (2002). "Seeing in Depth: Volume I, Basic Mechanisms." University of Toronto Press: I Porteous, Toronto.
- Howard, I. P., & Rogers, B. J. (2002). "Seeing in Depth: Volume II, Depth Perception." Chapter 30, Stereoscopic Techniques and Applications. University of Toronto Press: I Porteous, Toronto.
- Howarth, P. A. (1999). "Oculomotor changes within virtual environments." *Applied Ergonomics*, 30(1), pp. 59-67.
- Hsu, J., Pizlo, Z., Chelberg, D. M., Babbs, C. F., & Delp, E. J. (1996). "Issues in the design of studies to test the effectiveness of stereo imaging." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 26(6), pp. 810-819.
- IJsselsteijn, W. A., de Ridder, H., & Vliegen, J. (2000a). "Effects of stereoscopic filming parameters and display duration on the subjective assessment of eyestrain." Proceedings of SPIE, 3957, pp. 12-22.
- IJsselsteijn, W. A., de Ridder, H., & Vliegen, J. (2000b). "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration." *IEEE Transactions on Circuits and Systems for Video Technology*, 10(2), pp. 225-233.
- Inoue, T., & Ohzu, H. (1997). "Accommodative responses to stereoscopic three-dimensional display." *Applied Optics*, 36(19), pp. 4509-4515.
- Intuitive Surgical, Inc. (2013). "The da Vinci Surgical System." [webpage] Available: [http://www.intuitivesurgical.com/products/davinci\\_surgical\\_system/](http://www.intuitivesurgical.com/products/davinci_surgical_system/)
- Johnson, D. M. (2005). "Introduction to and Review of Simulator Sickness Research." U.S. Army Research Institute for the Behavioral and Social Sciences, Research Report 1832.

Available online: <http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-HFM-121-PART-II//TR-HFM-121-Part-II-05.pdf>

- Jones, R. (1977). "Anomalies of disparity detection in the human visual system." *Journal of Physiology*, 264, pp. 621-640.
- Jones, G., Lee, D., Holliman, N., Ezra, D. (2001). "Controlling perceived depth in stereoscopic images," *Proceedings of SPIE 4297*, 42-53.
- Judge, S. J., & Miles, F. A. (1985). "Changes in the coupling between accommodation and vergence eye movements induced in human subjects by altering the effective interocular separation." *Perception*, 14, pp. 617-629.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). "Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness." *International Journal of Aviation Psychology*, 3(3), pp. 203-220.
- Kihara, K., Fujisaki, H., Ohtsuka, S., Miyao, M., Shimamura, J., Arai, H., & Taniguchi, Y. (2013). "Age differences in the use of binocular disparity and pictorial depth cues in 3D-graphics environments." *SID 2013 Digest*, pp. 501-504.
- Kim, D., Choi, S., & Sohn, K. (2011). "Depth adjustment for stereoscopic images and subjective preference evaluation." *Journal of Electronic Imaging*, 20(3), pp. 033011-1.
- Kolasinski, E. M., Jones, S. A., Kennedy, R. S., & Gilson, R. D. (1994). "Postural stability and its relation to simulator sickness." *Proceedings of the 38<sup>th</sup> Annual Meeting of the Human Factors and Ergonomics Society*, 2, p. 980.
- Kolasinski, E. M. (1995). "Simulator sickness in virtual environments." U.S. Army Research Institute for the Behavioral and Social Sciences, Technical Report 1027. Available online: <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA295861>
- Kooi, F. L., & Toet, A. (2004). "Visual comfort of binocular and 3D displays." *Displays*, 25, pp. 99-108.
- Kuze, J., & Ukai, K. (2008). "Subjective evaluation of visual fatigue caused by motion images." *Displays*, 29(2), pp. 159-166.
- Kytö, M., Hakala, J., Oittinen, P., & Häkkinen, J. (2012). "Effect of camera separation on the viewing experience of stereoscopic photographs." *Journal of Electronic Imaging*, 21(1), p. 011011-1.
- Lambooi, M. T. M., Fortuin, M., IJsselsteijn, W. A., Heynderickx, I. (2010a). "Measuring visual discomfort associated with 3D displays." *Proceedings of SPIE 7237*, 72370K.
- Lambooi, M. T. M., Fortuin, M., IJsselsteijn, W. A., Heynderickx, I. (2010b). "Measuring visual fatigue and visual discomfort associated with 3-D displays." *Journal of the Society for Information Display*, 18(11), pp. 931-943.

- Lambooij, M. T. M., Fortuin, M., IJsselsteijn, W. A., Evans, B. J. W., & Heynderickx, I. (2011). "Susceptibility to visual discomfort of 3-D displays by visual performance measures." *IEEE Transactions on Circuits and Systems for Video Technology*, 21(12), pp. 1913-1923
- Lambooij, M. T. M., (2012). "Reading performance as screening tool for visual complaints from stereoscopic content." *Displays*, 33, pp. 84-90.
- Lambooij, M. T. M., IJsselsteijn, W. A., & Heynderickx, I. (2007). "Visual discomfort in stereoscopic displays: A review." Proceedings of SPIE, 6490, 64900I. Available online: [http://iss.bu.edu/jkonrad/Publications/refs/pdf/SDA-2007/Lambooij\\_etal\\_SDA\\_2007.pdf](http://iss.bu.edu/jkonrad/Publications/refs/pdf/SDA-2007/Lambooij_etal_SDA_2007.pdf)
- LaViola, J. J. (2000). "A discussion of cybersickness in virtual environments." Society for Information Graphics and Computer-Human Interaction (SIGCHI) Bulletin, 32(1), pp. 47-56.
- Lampton, D. R., Gildea, J. P., McDonald, D. P., & Kolasinski, E. M. (1996). "Effects of display type on performance in virtual environments." U.S. Army Research Institute, Technical Report 1049.
- Lawson, B. D., Graeber, D. A., Mead, A. M., & Muth, E. R. (2002). "Signs and symptoms of human syndromes associated with synthetic experiences." In HANDBOOK OF VIRTUAL ENVIRONMENTS: DESIGN, IMPLEMENTATION, AND APPLICATIONS by K. M. Stanney (Ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Leroy, L., Fuchs, P., & Moreau, G. (2012). "Visual fatigue reduction for immersive stereoscopic displays by disparity, content, and focus-point adapted blur." *IEEE Transactions on Industrial Electronics*, 59(10), pp. 3998-4004.
- McIntire, J. P., Havig, P. R., & Geiselman, E. (2012). "What is 3D good for? A review of the human performance benefits of stereoscopic 3-D displays." Proceedings of SPIE, 8383, 83830X.
- McIntire, J. P., Havig, P. R., & Geiselman, E. (2014). "Stereoscopic 3D displays and human performance: A comprehensive review." *Displays*, 35(1), pp. 18-26.
- McIntire, J. P., Wright, S. T., Harrington, L. K., Havig, P. R., Watamaniuk, S. N. J., & Heft, E. (2014). "Optometric measurements predict performance but not comfort on a virtual object placement task with a stereoscopic 3D display." *Optical Engineering*, Special Issue on Human Vision, 53(6), p. 061711-1.
- McKee, S. P., & Taylor, D. G. (2010). "The precision of binocular and monocular depth judgments in natural settings." *Journal of Vision*, 10(10):5, pp. 1-13.
- Miles, F. A., Judge, S. J., & Optican, L. M. (1987). "Optically induced changes in the couplings between vergence and accommodation." *Journal of Neuroscience*, 7, pp. 2576-2589.

- Mitsubishi, T. (1996). "Evaluation of stereoscopic picture quality with CFF." *Ergonomics*, 39(11), pp. 1344-1356.
- Mollenhauer, M. A. (2004). "Simulator adaptation syndrome literature review." Realtime Technologies, Inc., Technical Report. Available online: <http://www.simcreator.com/documents/papers/simadaptsyndrome.pdf>
- Mon-Williams, M., Wann, J. P., & Rushton, S. (1993). "Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display." *Ophthalmic and Physiological Optics*, 13(4), pp. 387-391.
- Neveu, P., Priot, A.-E., Plantier, J., & Roumes, C. (2010). "Short exposure to telestereoscope affects the oculomotor system." *Ophthalmic and Physiological Optics*, 30, pp. 806-815.
- Nicolau, M., Levine, A. J., & Carlsson, G. (2011). "Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival." *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, 108(17), pp. 7265-7270.
- Nojiri, Y., Yamanoue, H., Hanazato, A., Emoto, M., & Okano, F. (2004). "Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing." *Proceedings of SPIE*, 5291, Stereoscopic Displays and Virtual Reality Systems XI, pp. 303-313.
- NOVA (2012). "3D Spies of WWII." [television program] Aired January 18, 2012 on PBS. Available: <http://www.pbs.org/wgbh/nova/military/spies-3d.html>
- O'Connor, A. R., Birch, E. E., Anderson, S., Draper, H., and the FSOS Research Group (2010). "The functional significance of stereopsis." *Investigative Ophthalmology and Visual Science*, 51(4), pp. 2019-2023.
- Omori, M., Ishihara, S., Hasegawa, S., Ishigaki, H., Watanabe, T., Miyao, M., & Tahara, H. (2005). "Accommodative load for stereoscopic displays." *Proceedings of SPIE*, 5664, pp. 64-71.
- Oyamada, H., Iijima, A., Tanaka, A., Ukai, K., Toda, H., Sugita, N., Yoshizawa, M., & Bando, T. (2007). "A pilot study on pupillary and cardiovascular changes induced by stereoscopic video movies." *Journal of Neuroengineering and Rehabilitation*, 4, 37.
- Parsons, T. D., Larson, P., Kratz, K., Thiebaut, M., Bluestein, B., Buckwalter, J. G., & Rizzo, A. A. (2004). "Sex differences in mental rotation and spatial rotation in a virtual environment." *Neuropsychologia*, 42, pp. 555-562.
- Pastoor, S. (1993). "Human factors of 3D displays in advanced image communications." *Displays*, 14(3), pp. 150-157.
- Patterson, R. (2009). "Human factors of stereo displays: An update." *Journal of the Society for Information Display*, 17(12), pp. 987-996.

- Pausch, R., Crea, T., & Conway, M. (1992). "A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness." *Presence*, 1(3), pp. 344-363.
- Peinsipp-Byma, E., Rehfeld, N., & Eck, R. (2009). "Evaluation of stereoscopic 3D displays for image analysis tasks." *Proceedings of SPIE*, 7237, 72370L.
- Poyade, M., Reyes-Lecuona, A., & Viciano-Abad, R. (2009). "Influence of binocular disparity in depth perception mechanisms in virtual environments" (Chapter 2). In J.A. Macias et al. (eds.) *NEW TRENDS ON HUMAN-COMPUTER INTERACTION*, Springer-Verlag, London Limited, pp. 13-22.
- Reason, J. T. (1970). "Motion sickness: A special case of sensory rearrangement." *Advancement of Science*, 26, pp. 386-393.
- Regan, E. C. (1995). "Some evidence of adaptation to immersion in virtual reality." *Displays*, 16(3), pp. 135-139.
- Reinhardt-Rutland, A. H. (1996). "Remote operation: A selective review of research into visual depth perception." *Journal of General Psychology*, 123(3), pp. 237-248.
- Remote Vision System (2011). "Why 3D?" Rockwell Collins: Horizons, 16(4), pp. 10-11. Available: <http://www.rockwellcollins.com/~media/Files/Unsecure/Pages/Horizons/HorizonsVol16Issue4.ashx>
- Riccio, G. E., & Stoffgren, T. A. (1991). "An ecological theory of motion sickness and postural instability." *Ecological Psychology*, 3(3), pp. 195-240.
- Richards, W. (1971). "Anomalous stereoscopic depth perception." *Journal of the Optical Society of America*, 61(3), pp. 410-414.
- Rosenberg, L. (1993). "The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks." *Proceedings of the IEEE VRAIS '93*, pp. 27-31.
- Rowlands, J. J., & Killian, J. R. (1937). "Seeing solid: The third dimension at work and play." *Technology Review*, March, pp. 191-210.
- Rushton, S., Mon-Williams, M., & Wann, J. P. (1994). "Binocular vision in a bi-ocular world: New-generation head-mounted displays avoid causing visual deficit." *Displays*, 15(4), pp. 255-260.
- Russi, J. G. (2013). "Effects of stereoscopic 3D digital radar displays on air traffic controller performance." United States Air Force, Air University, Air Force Institute of Technology, Wright-Patterson AFB, OH [unpublished master's thesis]. AFIT-ENV-13-M-24.

- Sasaki, Y., Rajimehr, R., Kim, B. W., Ekstrom, L. B., Vanduffel, W., & Tootell, R. B. H. (2006). "The radial bias: A different slant on visual orientation sensitivity in human and nonhuman primates." *Neuron*, 51(5), pp. 661-670.
- Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). "Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems." *Displays*, 29(2), pp. 58-69.
- Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011a). "Visual discomfort with stereo displays: Effects of viewing distance and direction of vergence-accommodation conflict." *Proceedings of SPIE*, 7863, 78630P.
- Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011b). "The zone of comfort: Predicting visual discomfort with stereoscopic displays." *Journal of Vision*, 11(8), pp. 1-29.
- Siegel, M. (1999). "Just enough reality: A kinder gentler approach to stereo." *Proceedings of the SPIE Conference on Cockpit Displays VI: Displays for Defense Applications*, Orlando, FL, April 1999, vol. 3690, pp. 173-179.
- Siegel, M., & Nagata, S. (2000). "Just enough reality: Comfortable 3-D viewing via microstereopsis." *IEEE Transactions on Circuits and Systems for Video Technology*, 10(3), pp. 387-396.
- Siegel, M., Tobinaga, Y., & Akiya, T. (1999). "Kinder gentler stereo." *Proceedings of the SPIE Conference on Stereoscopic Displays and Applications X*, San Jose, CA, January 1999, vol. 3639, pp. 18-27.
- Singer, M. J., Ehrlich, J., & Cinq-Mars, S. (1995). "Task performance in virtual environments: Stereoscopic versus monoscopic displays and head-coupling." U.S. Army Research Institute for the Behavioral and Social Sciences, Technical Report 1034. Available online: <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA306720>
- Solimini, A. G. (2013). "Are there side effects to watching 3D movies? A prospective crossover observational study on visually induced motion sickness." *PLoS ONE*, 8(2), e56160.
- Speranza, F., Tam, W. J., Renaud, R., & Hur, N. (2006). "Effect of disparity and motion on visual comfort of stereoscopic images." *Proceedings of SPIE*, 6055, 60550B.
- Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997). "Cybersickness is not simulator sickness." *Proceedings of the Human Factors and Ergonomics Society 41<sup>st</sup> Annual Meeting*, pp. 1138-1142.
- Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). "Human factors issues in virtual environments: A review of the literature." *Presence*, 7(4), pp. 327-351.

- Stidwell, D., & Fletcher, R. (2011). "Depth Perception." Chapter 11 in *NORMAL BINOCULAR VISION: THEORY, INVESTIGATION, AND PRACTICAL ASPECTS*. Malden, MA: Blackwell Publishing Ltd.
- Stone, R. T., Watts, K. P., & Rosenquist, B. E. (2012). "Evaluation of 3D television: Impact on depth perception." *Proceedings of the Human Factors and Ergonomics Society 56<sup>th</sup> Annual Meeting*, pp. 733-737.
- Suzuki, Y., Onda, Y., Katada, S., Ino, S., & Ifukube, T. (2004). "Effects of an eyeglass-free 3-D display on the human visual system." *Japanese Journal of Ophthalmology*, 48(1), pp. 1-6.
- Tourancheau, S., Sjöström, M., Olsson, R., Persson, A., & Ericson, T. (2012). "Evaluation of quality of experience in interactive 3D visualization: methodology and results." *Proceedings of SPIE*, 8288, 82880O.
- Treisman, M. (1977). "Motion sickness: An evolutionary hypothesis." *Science*, 197, pp. 493-495.
- Tsukuda, S., & Murai, Y. (1988). "A case report of manifest esotropia after viewing anaglyph stereoscopic movie." *Japanese Orthoptic Journal*, 18, pp. 69-72.
- Ukai, K., Howarth, P. A. (2008). "Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations." *Displays*, 29, pp. 106-116.
- van Bergen, E., van Swieten, L. M., Williams, J. H. G., & Mon-Williams, M. (2007). "The effect of orientation on prehension movement time." *Experimental Brain Research*, 178, pp. 180-193.
- van Beurden, M. H. P. H., IJsselsteijn, W. A., & de Kort, Y. A. W. (2011). "Evaluating stereoscopic displays: Both efficiency measures and perceived workload sensitive to manipulations in binocular disparity." *Proceedings of SPIE-IS&T Electronic Imaging, Stereoscopic Displays and Applications XXII*, 7863, pp. 786316-1.
- Vienne, C., Blondé, L., & Doyen, D. (2012). "Visual fatigue versus eye-movements." *Proceedings of SPIE*, 8288, 828812-1.
- von Noorden, G. K., & Campos, E. C. (2002). "The Near Vision Complex." Chapter 5 in *BINOCULAR VISION AND OCULAR MOTILITY: THEORY AND MANAGEMENT OF STRABISMUS*. Mosby: Philadelphia, PA. Free electronic versions available at: [http://telemedicine.orbis.org/bins/content\\_page.asp?cid=1-2193](http://telemedicine.orbis.org/bins/content_page.asp?cid=1-2193)
- Waack, F. G. (1987). "A Little History." Chapter 2 in *STEREO PHOTOGRAPHY: AN INTRODUCTION TO STEREO PHOTO TECHNOLOGY AND PRACTICAL SUGGESTIONS FOR STEREO PHOTOGRAPHY*. Limited chapters available online from Stereoscopy.com at: <http://www.stereoscopy.com/library/waack-contents.html>
- Wade, N. J. (1987). "On the late invention of the stereoscope." *Perception*, 16, pp. 785-818.

- Wang, J., Barkowsky, M., Ricordel, V., & Le Callet, P. (2011). "Quantifying how the combination of blur and disparity affects the perceived depth." *Proceedings of SPIE*, 7865, 78650K.
- Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). "Natural problems for stereoscopic depth perception in virtual environments." *Vision Research*, 35(19), pp. 2731-2736.
- Ware, C., Gobrecht, C., & Paton, M. A. (1998). "Dynamic adjustment of stereo display parameters." *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 28(1), pp. 56-65.
- Watson, A. B., & Pelli, D. G. (1983). "QUEST: A Bayesian adaptive psychometric method." *Perception & Psychophysics*, 33(2), pp. 113-120.
- Wee, S. w., & Moon, N. J. (2014). "Clinical evaluation of accommodation and ocular surface stability relevant to visual asthenopia with 3D displays." *BMC Ophthalmology*, 14(29), doi: 10.1186/1471-2415-14-29.
- Westheimer, G. (2003). "Meridional anisotropy in visual processing: Implications for the neural site of the oblique effect." *Vision Research*, 43, pp. 2281-2289.
- Wheatstone, C. (1838). "On some remarkable, and hitherto unobserved, phenomena of binocular vision." *Philosophical Transactions of the Royal Society of London*, 128, pp. 371-394.
- Wilkinson, S. (2011). "3D Sickness Explained?" Home Theater [web article], 25 January 2011. Available: <http://www.hometheater.com/content/3d-sickness-explained>
- Williams, S. P., & Parrish, R. V. (1990). "New computational control techniques and increased understanding for stereo 3-D displays," *Proceedings of SPIE* 1256, pp. 73-82.
- Wöpking, M. (1995). "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus." *Journal of the Society for Information Display*, 3(3), pp. 101-103.
- Yamanoue, H., Okui, M., & Yuyama, I. (2000). "A study on the relationship between shooting conditions and cardboard effect of stereoscopic images." *IEEE Transactions on Circuits and Systems for Video Technology*, 10(3), pp. 411-416.
- Yamanoue, H., Okui, M., & Okano, F. (2006). "Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images." *IEEE Transactions on Circuits and Systems for Video Technology*, 16(6), pp. 744-752.
- Yang, S. N., & Sheedy, J. E. (2011). "Effects of vergence and accommodative responses on viewer's comfort in viewing 3D stimuli." *Proceedings of SPIE*, 7863, 78630Q.
- Yano, S., Emoto, M., & Mitsunashi, T. (2004). "Two factors in visual fatigue caused by stereoscopic HDTV images." *Displays*, 25, pp. 141-150.