THE TEST–RETEST RELIABILITY OF FOUR BINOCULAR VISION TESTS

EVALUATING THE TEST–RETEST RELIABILITY OF FOUR NEWLY DEVELOPED BINOCULAR VISION TESTS

By FERMIN H. RETNAVARATHAN, H.B.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University

McMaster University, MASTER OF SCIENCE (2025)

Hamilton, Ontario (Neuroscience)

TITLE: Evaluating the Test–Retest Reliability of a Battery of Newly Developed Binocular Vision Tests

AUTHOR: Fermin H. Retnavarathan, H.B.A. (McMaster University)

SUPERVISOR: Dr. David I. Shore

NUMBER OF PAGES: xii, 63

LAY ABSTRACT

Patients with amblyopia typically have poor vision in one of their eyes. Since one eye has worse vision, the brain tends to rely on input from the stronger eye. This has consequences as individuals with amblyopia do not develop the ability to integrate the images seen by each eye—*binocular vision.* Without binocular vision, patients struggle with various daily activities. Remarkably, some amblyopic patients were able to perceive the *Pulfrich effect*– a depth illusion that requires binocular vision (Maehara et al., 2019). To explore this *hidden binocularity* further, we developed a battery of tasks to measure binocular vision in this population. Before administering these tasks to patients, the present study evaluated the reliability of four of the tasks by having control participants complete the tasks twice, one week apart. Three of the four tasks demonstrated strong reliability supporting their use as reliable tools to measure binocular vision in this population.

ABSTRACT

Amblyopia will affect 200 million people around the world by 2030 (Fu et al., 2020). Characterized by poor vision, primarily in one eye, this condition arises when an individual does not receive concordant visual input early in life due to strabismus (misalignment of the eyes), cataracts, or high differential refractive error between the two eyes. Due to a lack of normal binocular input early in life, individuals with amblyopia do not develop binocular vision. The disruption of binocular vision prevents accurate depth perception, which causes challenges with everyday tasks such as driving and reading (Levi, Knill & Bavelier, 2015; Birch et al., 2018). Even after corrective surgery, deficits often persist throughout life. Remarkably, in a recent paper by Maehara et al. (2019), a subset of amblyopia patients, who failed all clinical tests of binocular vision, demonstrated a *Pulfrich effect*. The Pulfrich effect occurs when horizontally moving objects are presented to both eyes with a neutral density filter over one eye. The reduced contrast to the one eye delays visual processing, which the perceptual system perceives as spatial disparity, inducing depth perception. Evidently, binocular vision is necessary to perceive this effect implying these patients have residual *hidden binocularity.* To explore this phenomenon further, we developed a battery of binocular vision tests (most of which are motion-based). The present project evaluated the test–retest reliability of four tasks: Letter Dominance, Pulfrich, Plaid Motion, and Motion Parallax. Participants with typically developed vision completed these four tasks twice, one week apart. We observed a strong positive correlation between performance on week one and week two for the Letter Dominance, Pulfrich, and Plaid Motion tasks. This represents a foundational step in a research program which aims to obtain more sensitive measures of binocular vision in this population.

ACKNOWLEDGEMENTS

I am first extending my deepest gratitude to my supervisor David Shore. David and I met in an unconventional way—I was seeking a thesis supervisor amidst my last year in the PNB program. The semester had already started, and I was starting to lose hope until David reached out to me himself. I will always be thankful to David for contacting me and for his unwavering support and generosity these last few years. David has taught me lessons that I will carry with me both in and beyond the lab—among them, the value of collaboration, attention to detail (yes, even knowing the difference between en and em dashes), and how to navigate setbacks in research with resilience. Anytime I encounter a challenge in my research, I hear David's voice in my head saying "Well, that's science!" When we met, I thought he was the coolest professor in the department, and I still stand by that. I look forward to working together with David during my PhD and beyond.

Next, I am extending my deepest thanks to my committee members Daphne Maurer (O.C.) and Daniel Goldreich for supporting me during the completion of my Master's. My first introduction to Dr. Goldreich was when I took the course PNB 2XB3 as an undergraduate student. As a second year student, a neurophysiology & neuroanatomy course was daunting to say the least. But Dr. Goldreich has a remarkable way of making complex concepts engaging and accessible. Years later, although we did not meet often, the time we spent together discussing my research has been invaluable to me. Dr. Goldreich's insights and support as a member of my committee have made a lasting impact on my academic journey.

Without Daphne, my research simply would not exist. I see Daphne as the godmother of all amblyopia research. My introduction to Daphne's work was during my undergraduate thesis, when I explored the audiotactile simultaneity perception of adult congenital cataract patients tracked by Daphne since birth. It is truly an honour for me to not only have Daphne as a mentor, but to be able to carry forward some of the research she began decades ago. As a woman in science, Daphne is one of my greatest inspirations—and while I aspire to follow in her footsteps (which are certainly big shoes to fill!), I know I can do it with her by my side.

I am extending my thanks next to my mentor, Brendan Stanley. Brendan was my graduate supervisor while I was working on my undergraduate thesis. Working on Brendan's project was my first introduction to multisensory and psychophysical research. Brendan consistently went above and beyond in his mentorship. I'll never forget when he offered to have a late night Zoom call with me after driving in from Montreal just so that I could rehearse for my first poster presentation. Brendan instilled a sense of confidence in me that I still carry with me today. Without Brendan, I would not be where I am today.

I am next extending my thanks to all members of our multi-institutional CIHR team: Andrew Silva, Xiaoxin Chen, Ben Thompson, Xiaoqing Gao, Haotian Lin, Agnes Wong, Daphne Maurer, and Ewa Niechwiej-Szwedo. Feedback from our monthly meetings were incredibly informative and helped me grow into the vision researcher that I am today. I am extending a special thank you to Andy and Xiaoxin for making time to meet with me throughout the year when I had doubts regarding the binocular battery and for teaching me skills such as how to gamma-correct a monitor. I am very grateful to be a part of such an intellectually diverse and talented team.

This research would not have been possible without the undergraduate student researchers on the binocular vision team at McMaster. I am extending my thanks to Lynn Turkstra, Riley Morgan, Dasnoor Saini, Angelica Maniglia, Bahtinur Yilmaz, and Ashton Poopalasingham. I am extending a special thank you to Tara Nichols for working passionately with me on this project. I will never forget our unique side quests like trying to glue the first version of the response box together with Krazy glue. I am glad to call Tara a friend and I look forward to continuing to work together in the future.

I am next extending my deepest thanks to all my dear friends both in and outside of the department. Prior to joining graduate school, I didn't realize how important the connections I made along the way would be. There are too many names to list, but I know you all know who you are. You all have supported me in various ways, and I cannot thank you enough.

Next, I am extending my heartfelt thanks to my family for always having my back during all the highs and low of graduate school. I want to thank my brother Favian Retnavarathan, my aunt Finez Mehar, my uncle Rexon James, my uncle Fayaz Mehar, my grandmother Farina Mehar, my grandfather Fujrush Mehar, my cousins Jerusha and Jeruby Retnakanthan, and the rest of my dearest cousins. They say, "it takes a village to raise a child," and when I reflect on my journey, I see that village in my family. You have all carried me in ways I will never forget, and I am endlessly grateful.

I am next extending a special thank you to my dad Retnavarathan Thiyagarajah. Many people know that my dad came to this country seeking refuge from a civil war. Ever since I was a kid, I saw my dad tirelessly work multiple jobs to give my brother and I a better life. As a first-generation student, I am incredibly proud to say I am my father's daughter, and I hope I made all his sacrifices feel worthwhile.

I am next extending a special and deeply personal thank you to my mother Fathima Mehar. Graduate school felt like a rollercoaster with moments of unpredictable highs and lows. One of the biggest reasons I was able to make it this far was because of your support. Mom, I know you sacrificed a lot for me to be here today. Your strength, love, and unwavering belief in

me carried me through my toughest challenges. I truly could not have made it this far without your endless support. I hope this accomplishment reflects even a fraction of what you have given me, and I hope I made you proud.

Last, but certainly not least (in fact, it is quite the opposite), I am expressing my deepest gratitude to my step-dad Uthayasangary (Bob) Tharmasangary. Your guidance and unwavering belief in me are among the biggest reasons I have made it this far in my academic journey. You have been not only a steadfast supporter, but also a constant source of motivation and strength. You are the person most invested in my academic journey and during my hardest moments, you remind me that I am never alone in this. I am endlessly thankful for everything you have done for me. This degree does not belong to just me, it belongs to everyone who helped me reach this milestone, especially you.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

DECLARATION OF ACADEMIC ACHIEVEMENT

This thesis was written by Fermin H. Retnavarathan and edited by Dr. David I. Shore, Dr. Daphne Maurer (O.C.), and Dr. Daniel Goldreich. All binocular vision tests conducted in the battery were coded by Dr. Andrew Silva and Dr. Xiaoxin Chen. The response boxes used in the Pulfrich task were designed by Dr. Xiaoqing Gao. The first version of the enclosure was built by Dr. David Shore and Dr. Brendan Stanley. The first versions of the response boxes used in the Pulfrich task were built by Fermin H. Retnavarathan and Tara Nichols. The second version of the enclosure and the response boxes were built by Gary Fekete. The binocular battery was conceptualized and developed by a multi-institutional team consisting of Dr. Andrew Silva, Dr. Xiaoxin Chen, Dr. Ben Thompson, Dr. Xiaoqing Gao, Dr. Haotian Lin, Dr. Agnes Wong, Dr. Daphne Maurer, Dr. David Shore, and Dr. Ewa Niechwiej-Szwedo. Additional contributions to the methods were made by Fermin H. Retnavarathan and Tara Nichols. The experiment was conducted by Fermin H. Retnavarathan alongside undergraduate students Tara Nichols, Lynn Turkstra, Riley Morgan, Dasnoor Saini, Angelica Maniglia, and Bahtinur Yilmaz. Final preparations and submission for publication were completed by Fermin H. Retnavarathan.

Introduction

The term *amblyopia* originates from the Greek word meaning “dull sight” (Daw, 1998). True to its Greek origin, amblyopia is a neurodevelopmental condition characterized by reduced vision typically in one eye (Levi, Knill & Bavelier, 2015). This condition arises because the input from the two eyes does not match and thus these patients do not experience binocular vision. The disruption of binocular vision prevents accurate depth perception, which produces consequences for daily living (Levi, Knill & Bavelier, 2015). Patients with amblyopia past seven years of age are often informed that they can never recover their visual acuity or stereoacuity as their visual system is beyond the critical period to develop binocular vision (Levi, Knill & Bavelier, 2015). However, in 2019, Maehara et al. discovered that a small group of amblyopia patients, who failed all clinical tests of binocular vision, were able to demonstrate the *Pulfrich effect* (Maehara et al., 2019). The Pulfrich effect is a depth illusion that requires intact binocular abilities. Perceiving depth through this illusion implies some intact binocular vision, which we refer to as *hidden binocularity*. Hidden, because current standard clinical tests fail to capture these binocular capabilities. To explore hidden binocularity further, we developed a battery of eight tasks. The present study evaluates the test–retest reliability of four of these tasks: a baseline task of eye dominance (*letter dominance),* a version of the *pulfrich* to potentially replicate the findings of Maehara et al., a *plaid motion* task (Chow, A., et al, 2021), and a new task testing the monocular depth cue of *motion parallax* (Gibson et al., 1959), both with and without disparity added.

Binocular Dysfunction in Amblyopia

Individuals with amblyopia typically do not demonstrate binocular vision and stereovision. The three major types of amblyopia are strabismic (misalignment of the eyes), anisometropic (high difference in refractive error between the eyes), and deprivation amblyopia (vision is impacted by a physical obstruction in one or both eyes) (Hamm et al., 2017). Normal visual input early in life is critical for the development of the visual system (Hubel and Wiesel, 1964). Amblyopic patients were deprived of focused concordant visual input to both eyes early in childhood. As a result, neuronal connections were altered in areas as early as the LGN. These neuronal alterations produce a range of visual deficits such as loss of visual acuity, stereopsis, position acuity, and contrast sensitivity in the deprived eye (Levi, Knill & Bavelier, 2015). Since the fellow eye of amblyopia patients dominates viewing, cortical inputs from the amblyopic eye are suppressed or actively inhibited leading to the loss of binocular vision (also known as interocular suppression; Hamm et al., 2017). Binocular vision is necessary for stereovision—the brain's ability to perceive depth from the slight differences between the images seen by each eye. Reducing vision in one eye whether through blurring, filtering, or reduced contrast, results in impaired stereovision (Levi, Knill & Bavelier, 2015). Consistent with these findings, impaired stereoscopic depth perception is the most common deficit associated with amblyopia.

Functional Consequences

Individuals living with amblyopia experience lifelong consequences to all aspects of their lives, from fundamental reaching and grasping to complex behaviours such as aiming, catching, and reading, all of which negatively impacts their daily lives. Even when compared with control participants with an induced blur in one eye, patients with anisometropic amblyopia demonstrated significantly different kinematics when performing a reach-to-touch task (Niechwiej-Szwedo et al., 2012). Many additional studies (e.g. Birch, Kelly & Giaschi, 2020; Kelly et al., 2020) demonstrate that poor visual acuity alone cannot explain the idiosyncrasies and deficits in motor skills demonstrated by amblyopic patients. Given the deficits in these foundational actions, it is not surprising that individuals with amblyopia show deficits in reading and eye–hand coordination (Birch et al., 2018; Niechwiej-Szwedo et al., 2012). These deficits appear to produce lower social, scholastic, and athletic competence (Birch et al., 2018). Moreover, slower reading speed and worse performance in aiming and catching were associated with lower self-perception (Birch et al., 2018). Ultimately, amblyopia can limit future career options as some professions—such as surgery—require accurate stereoability (Levi, Knill & Bavelier, 2015). Finding routes to rehabilitation forms a fundamental goal for many research programs.

The Pulfrich Effect Reveals *Hidden Binocularity*

Remarkably, a subset of amblyopia patients reported seeing a Pulfrich effect demonstrating *hidden binocularity*. Two patients, with no stereopsis as measured by the TNO (Netherlands Organization for Applied Scientific Research) test, reported viewing a Pulfrich effect (Maehara et al., 2019). The Pulfrich effect, discovered by Carl Pulfrich (1922), is most easily demonstrated using a pendulum. To elicit the effect, a pendulum is presented swinging horizontally in front of the observer with a light-attenuating filter (i.e., a neutral density filter) placed over one eye. The combination of the motion of the pendulum and the reduced luminance to one eye causes an illusion of depth—the pendulum appears to move in an elliptical path. The decreased luminance in the filtered eye causes a temporal delay in retinal processing (Heng & Dutton, 2011). This delay in processing for one eye causes the position of the moving pendulum to lag behind the image seen by the other eye, which produces a spatial disparity between the two eyes. This spatial disparity then activates disparity-tuned neurons to signal a difference in depth for the left-to-right motion compared to the right-to-left motion, giving rise to the perceived elliptical path of the pendulum (Heng & Dutton, 2011). Since this effect relies on disparity-tuned neurons and their production of depth perception, the finding that amblyopic patients, with no measurable stereopsis, can perceive this effect raises the interesting possibility that these patients do have stereopsis, that cannot be detected by current clinical tests. Critically, the Pulfrich effect, and other related stimulation, may provide a more sensitive measure of stereopsis.

Preserved Binocular Abilities in Amblyopia

There is growing evidence that individuals with amblyopia possess preserved binocular abilities. In a study by Baker et al. (2007), strabismic amblyopia patients demonstrated normal levels of binocular contrast summation when detection thresholds were adjusted in the amblyopic eye to equal monocular detectability. Likewise, Hess, Mansouri & Thompson (2011) found that strabismic amblyopes demonstrated normal binocular vision in viewing conditions where stimuli of different contrasts were presented to each eye. When lower contrasts were presented to the fellow eye, participants were able to obtain balanced dichoptic performance. In another study by Hamm et al. (2017), the authors found that when interocular contrast was adjusted to favour the amblyopic eye, most children with deprivation amblyopia could overcome interocular suppression. Results from these studies along with many others provide evidence that the neural architecture necessary to support binocular vision is present but suppressed under normal viewing conditions in this population.

Pathway to Rehabilitation

The creation of artificial conditions under which binocular combination can occur can strengthen binocular vision in this population. Despite these methods originally being created to measure interocular suppression, many studies have shown that creating artificial conditions whereby the amblyopic subject can experience binocular combination can improve their binocular abilities. Hess, Mansouri & Thompson (2010) demonstrated that prolonged viewing of stimuli presented at different contrasts to each eye leads to improved binocular vision in amblyopic subjects with strabismus. Moreover, it led these subjects to combine binocular information under natural viewing conditions over time—a remarkable finding. Reduced suppression because of prolonged viewing was also associated with improved visual acuity in the amblyopic eye as well as the establishment of stereovision. Given that the restoration of binocular vision is critical to recovering stereovision in this population, uncovering potential *hidden binocularity* is an important first step towards rehabilitation.

Utility of Motion-in-Depth Stimuli

Motion-in-depth stimuli may provide a measure of binocular vision that static tests do not index. Dynamic tests contain changing patterns which provide the viewer with several independent samples to estimate depth and correspond the images seen by each eye (Tidbury et al., 2016). When presented with stereoscopic stimuli in a static disparity condition and a dynamic disparity condition, control participant thresholds for detecting depth were 50% lower in the dynamic disparity condition (Tidbury et al., 2016). Likewise, when strabismic amblyopia patients were presented with targets where the disparity changed dynamically, individuals with more severe strabismus could detect these changes in their peripheral vision (Verghese, 2023).

Newly Developed Binocular Battery

We developed a battery of eight binocular vision tests, most of which use dynamic stimuli, which we predict will provide additional measures of binocular vision in this population. To measure the reliability of the battery, we have chosen to include four tasks in the present test–retest study. Participants completed each of the four tasks from the battery twice, one week apart. Each task is described below including the general methods and the specific measure extracted.

Letter Dominance

Participants are presented two letters on a gray background—one above a central fixation and one below (see Figure 1a). The contrast of the letters to the two eyes are opposite—the top letter will be brighter than the background to one eye and darker to the other eye. The bottom letter has the opposite arrangement—the eye with the brighter top letter will have a darker bottom letter and the other eye will have a darker top letter and a brighter bottom letter. Finally, one eye will receive high contrast stimulation and the other low contrast stimulation. Once fused these rivalrous stimuli will produce the perception of two letters of a medium gray luminance, but one letter will appear relatively brighter depending on which eye is dominant. The participant indicates which letter appears brighter. Across trials, we vary the relative contrast to the two eyes, which provides the critical independent variable for this task.

The Letter Dominance task was designed to assess degree of suppression in patients with amblyopia. As coined by Wang et al. (2019), *binocular eye balance* can provide us with a measure of the relative contribution of each eye in viewing a binocular percept and thus, the degree of inhibition present in the amblyopic eye. Binocular eye balance can be quantitatively assessed when similar monocular patterns are presented to each eye and then, fused into a cyclopean percept (Wang et al., 2019). One paradigm which has been used to investigate binocular combination includes the use of contrast (Huang et al., 2010). To measure binocular eye balance, this paradigm involves adjusting the interocular contrast ratio until both eyes are contributing equally to binocular vision.

We have replicated this paradigm in the Letter Dominance task. Using a double-interleaved staircase procedure followed by a method of constant stimuli, we can measure the contrast ratio at which both eyes are contributing equally to binocular vision. When contrast presented to the left eye is plotted against proportion of left eye responses the participant made, the data form a psychometric function. The relative contrast value that produces 50% left and right eye responses provides us with a quantification of binocular eye balance.

Plaid Motion

Participants are presented with a circular aperture containing diagonally oriented gratings drifting perpendicular to their angle (see Figure 1b). Each eye receives orthogonal gratings: the left eye sees a grating oriented at 45°. In the right eye, participants would see a grating oriented at 135°. These rivalrous stimuli produce a fluctuating perception that can vary between a left-eye dominated perception of a grating moving up and to the left or a right-eye dominated perception of a grating moving up and to the right. If the two eyes are balanced, the stimulus from each eye will be fused and the participant will perceive a plaid design moving upwards. In some cases, the participant will perceive a mixed or sliding percept where part of the visual field is moving up and to the left and other parts are moving to the right, creating a dynamic interplay of motion directions that slide past each other in visual space—an imperfect stage of fusion that often precedes the viewing of a perfect plaid percept. The participant viewed this stimulus for 3 minutes and provided continuous verbal report of their perception using the numbers indicated on the aperture: 1 and 2 indicated leftward motion of varying degrees; 4 and 5 indicated rightward motion; 3 indicated upward motion; and 0 indicated the sliding percept. As described below, the critical measure of fusion was the proportion of time spent indicating 0 or 3. A measure of eye dominance can also be extracted by comparing the relative time reported 1 and 2 versus 4 and 5.

The Plaid Motion task was designed to provide a measure of individual fusion ability. In doing so, this task can also provide us with another measure of binocular eye balance. Another method to quantitatively assess binocular eye balance is by presenting different monocular patterns to each eye, producing perceptual rivalry (Wang et al., 2019). Despite commonly being used to measure the neural correlates of visual perception, binocular rivalry has been used to measure sensory eye dominance. In a study by Handa et al. (2004), authors investigated the relationship between the sighting dominant eye and sensory dominant eye using a binocular rivalry task where participants were presented with rightward titled gratings to the right eye and leftward titled gratings to the left eye. If two orthogonal gratings are drifting in opposite directions, when viewed independently, they appear to move in their independent directions as expected, referred to as component motion. When these two images are superimposed, the gratings fuse together forming a 'plaid' design, commonly described as pattern motion (Adelson & Movshon, 1982). When orthogonally oriented gratings are viewed dichoptically, the two percepts compete for dominance and the participant will view either one of the monocular gratings (component motion) or a fluid 'mosaic' which can be described as an imperfect combination of each monocular view (pattern motion) (Andrews & Blakemore, 2001).

Our team designed the Plaid Motion task based on similar stimuli such as that used in the aforementioned studies. Based on whether the participant viewed more leftward or rightward motion, we could obtain a measure of eye dominance. Based on the amount of time a participant spent viewing a fluid 'mosaic' or plaid design, we could measure time spent in binocular fusion. Both measures are critical in helping us validate sensory ocular dominance in amblyopia patients as well as residual binocular abilities, if present.

Motion Parallax

Participants are presented with a large square surface with four small squares, one in each quadrant (see Figure 1c). On each trial, the large square tilts on the horizontal plane very quickly.

Using cues of motion parallax, one of the small squares is given a different depth—it appears to float above the larger square. On some trials the target square also has disparity cues added to signal depth. Regardless of the source of the depth, the participant indicates the small square that was perceived to have depth.

Understanding of the two depth cues: (1) motion parallax and (2) disparity are fundamental to understanding the design of our Motion Parallax task. Motion Parallax is a monocular depth cue created by the translation of an observer's head—movement of the head to the left or right (Nawrot & Joyce, 2006). Disparity is a depth cue which relies on the minor difference in an image's position as seen by each eye. Due to the horizontal separation of the eyes in space, images land on slightly different points on the retina of each eye. If an object is closer to you than your point of fixation, the object will appear doubled due to crossed disparity. Whereas if the object is farther than your point of fixation, the object will appear doubled, but due to uncrossed disparity. The brain interprets these patterns to determine whether an object is in front of or behind the fixation point. Additionally, the magnitude of the disparity is proportional to the depth between the object and fixation point, allowing the brain to extract more precise depth information (Nawrot & Joyce, 2006).

Similarly, in motion parallax, objects moving in opposite directions on the retina indicate to the brain opposite depths relative to the fixation point, allowing the brain to extract depth information (Nawrot & Joyce, 2006). During translation of the observer's viewpoint, objects in the scene shift relative to one another. Objects which are farther than the fixation point, move in the same direction on your retina. Objects which are closer than the fixation point, move in the opposite direction on the retina. These objects appear to shift faster than objects in the distance. This is why when you are in a car on the highway, the divider appears to be moving by faster than the trees or buildings in the background. In order to maintain a stable view of one point in the scene, the eyes automatically move in the opposite direction of the translation. This ocular compensation in combination with head translation creates a stimulus of shifting patterns on the retina from which the brain can extract depth information (Nawrot & Joyce, 2006). Additionally, the speed of an object's image on the retina is proportional to the depth between the object and fixation point allowing the brain to extract additional depth information (Nawrot & Joyce, 2006).

In order for the observer to perceive depth from motion parallax, either the observer or the visual scene has to be in motion (Schiller et al., 2012). In a study by Schiller et al. (2012), authors assessed stereopsis and motion parallax through presentation of a dynamic display where random-dot stereograms were rocked forwards and backwards on a vertical axis. This portion of the study involved three conditions: (1) disparity cues only, (2) differential motion cues only, and (3) both cues presented together. Our team created the Motion Parallax task based on the design of this study. There are two conditions in our task: the target square would appear to have depth either through (1) the motion parallax cue only, or (2) the motion parallax cue plus an additional disparity cue. We designed this task with two conditions to assess whether amblyopic patients can perceive the added binocular disparity in the second. We can evaluate this by measuring whether performance improved in the added disparity condition compared to the parallax only condition. To do this, we administer two interleaved staircases to obtain threshold estimates for each condition. The difference between these thresholds serves as a measure of performance advantage and ultimately, *hidden binocularity*.

Our predictions for this task are two-fold. First, we predict individuals with normal binocular vision will demonstrate a performance advantage in the second condition over patients with amblyopia as they will benefit from the additional disparity cue. Second, we predict patients with amblyopia who have hidden binocular abilities will demonstrate a performance advantage on the second condition of this task over patients with amblyopia who do not have preserved binocularity as they will be able to detect the added disparity.

Pulfrich

Participants participated in two versions of the Pulfrich task: the Disparity task and the Sweeping Pulfrich task. In the Disparity task, participants are presented with stationary stars and moons moving leftwards across the screen (see Figure 1d). A neutral density filter is placed over one side of the monitor which, when viewed through a stereoscope, presents over one of the participant's eyes. The combined motion of the moons as well as the neutral density filter over one eye creates the illusion that the moving moons and stationary stars are not on the same depth plane—a replication of the Pulfrich effect (Pulfrich, 1922). At the bottom of the monitor, there is one stationary moon that the participant can adjust in depth using a disparity manipulation. That is, the static moon is presented at slightly different positions in each eye so that when fused they appear in depth. The participant's task is to move the stationary moon so that it is on the same depth plane as the moving moons. After each trial, the participant also reports the perceived depth using a physical response box. The response box contains stars hanging from the ceiling of the box as well as moons on a mobile slider. Using the slider, participants can slide the moons to a location relative to the stars to indicate the amount of depth they viewed on the monitor. We created this second measure of the Pulfrich effect as we anticipate some adults and children with amblyopia will have difficulty with the digital task since they are not sensitive to static disparity.

In the Sweeping Pulfrich task, the participant is presented with an identical display, however, there is no stationary moon at the bottom and the neutral density filter starts at 100% darkness over the participant's dominant eye. The participant adjusts the darkness of the neutral density filter from completely over the dominant eye to completely over the non-dominant eye; they report four points along this continuum: where they first perceive different depth for the moons and starts, where that depth disappears, then when the depth reappears with the filter over the non-dominant eye and then where it disappears.

For the disparity task, the participant is presented a different level of luminance on each trial and they indicate their perceived depth by adjusting the disparity of the static moon or the depth of the moons in the response box. Our dependent variable of interest is the slope of the line relating level of luminance with the response. For the sweeping pulfrich task, the dependent variable is the contrast level where the observer sees or does not see depth.

Hypotheses

The primary aim of this study is to evaluate the test–retest reliability of a battery of newly developed binocular vision tasks. Based on previous piloting, we predict all tasks will demonstrate strong reliability across the two weeks. Specifically, Pearson's correlation coefficients are expected to fall within the high range ($r \geq 0.70$) as outlined by Cohen (1988). Intraclass correlation coefficients are expected to fall within the good range ($0.60 \geq 0.70$) or the excellent range ($0.75 \geq 1.00$) (Cicchetti, 1994). Given all tasks were administered under standardized conditions, minimal variability between sessions is anticipated. Individual differences in binocular vision may also be revealed through the tasks. Overall, we predict performance on the tasks across the two runs will be highly correlated, supporting their use in the measurement of binocular vision in patients with amblyopia in a research setting.

Method

Participants

All participants were McMaster undergraduate students, recruited through SONA (an online research participation system). Experiment 1a recruited 25 participants (mean age = 18.5 years; age range = 16-25 years) while Experiment 1b recruited 12 participants (mean age = 19.8 years; age range = 18-22 years). The vast majority of participants had self-reported normal vision and passed vision screening with normal or corrected-to-normal vision: at least 20/20 (-2) visual acuity, a minimum of 40s of arc on the Randot test of stereoacuity, and normal fusional ability as demonstrated on Worth's Four Dot test (see Appendix A for individual performance). The few participants (n=3) who did not pass vision screening on both runs, are clearly identified in the analysis and figures. Participants attended the Multisensory Perception Laboratory at McMaster University twice, separated by one week; they completed the vision screening and a battery of four tasks during each visit. Upon completion of the second session, participants were compensated three credits towards a learning objective in their introductory psychology course. All participants provided written consent prior to participation. The study conformed to the Tri-Council Statement on Ethical Conduct of Research Involving Humans (TCPS2; Canada) and was approved by the McMaster Research Ethics Board (protocol #1733).

Apparatus and Stimuli

Participants sat in a chair at a fixed height of 43cm and rested their chin in a chin rest attached to the front edge of the table (see Appendix B). The height of the table was adjusted to the participant's preference while in the chin rest. A mirror stereoscope, consisting of two fixed mirrors and two adjustable mirrors (one per eye), was positioned directly in front of the chin rest. An LG UltraFine monitor (model no. 27UP600) was positioned 99.5cm in front of the chin rest.

At this distance, one pixel on the screen subtended 0.00892 degrees of visual angle or 32.11 arcsec. White paper with a square hole cut out was affixed to each mirror to conceal the metal mounts of the mirrors and any extraneous stimuli (see Appendix B).

The chinrest, stereoscope, and monitor were enclosed in material to avoid extraneous stimuli in the participant's field of view. For Experiment 1a, the enclosure was constructed from white corrugated plastic sheets supported by a wood square structure (see Appendix B). White Bristol board was placed over the corrugated plastic sheets to eliminate reflection from the monitor. Two LED light bars were placed face-down on the ceiling of the enclosure illuminating the inside of the enclosure with white light. For Experiment 1b, the enclosure was made with plywood that was painted black and held together with L brackets (see Appendix B). The second enclosure was sturdier than the first and the black paint reduced reflections and extraneous stimuli. Instead of LED light bars, a lamp was placed on top of the second enclosure to provide dim lighting in the testing room. Finally, a black cloth was draped over the back opening of the enclosure to eliminate any backlighting of the monitor.

Participants entered their responses on a number keypad placed under their right hand. In addition, for the Pulfrich task, we built a physical response box, which was placed on a small, custom-built table to the right of the participant. The physical response box is 17cm x 16cm x 18.2cm and contains a replication of the stimuli in the digital Pulfrich task (see Appendix B). The box contains a set of plastic moons attached to a metal rail as well as stationary stars which hang from the ceiling of the box. Using the handles on both sides of the box, the participant can slide the set of moons either in front of or behind the stars to indicate how much depth they viewed between the moons and the stars on the monitor. We created the physical response box as we anticipate the digital task may be difficult for children and adults with amblyopia. There were two versions of the response box used in this experiment. Major changes included the use of stronger adhesives, and the stars and moons were perfectly aligned at 0cm on the scale located on the side of the box. This was not the case in the previous version.

To support alignment of the two eyes, all stimuli were presented within a concordant black and white square frame identical in each eye. Before engaging in the battery of binocular vision tests, participants completed a fixation procedure to ensure proper eye alignment and binocular fusion. The horizontal segment of the fixation was presented to the left eye while the vertical segment was presented to the right eye. Participants indicated the degree of misalignment and the images were adjusted until aligned from the participant's point of view.

Letter Dominance Stimuli:

Stimuli presented to the left eye were two letters, one presented at the top of the screen and one at the bottom separated by a fixation cross in the centre (see Figure 1a). Each letter subtended 3 degrees of visual angle and was presented at 3.75 degrees eccentricity above or below the fixation cross. The same two letters were presented to the right eye separated by a fixation cross. The two letters presented change every trial. The contrast of the letters to the two eyes are opposite—the top letter will be brighter than the background to one eye and darker to the other eye. The bottom letter has the opposite arrangement—the eye with the brighter top letter will have a darker bottom letter and the other eye will have a darker top letter and a brighter bottom letter. On any given trial, the pair of letters presented to one eye will have high contrast and the pair in the alternate eye will have low contrast.

Plaid Motion Stimuli:

Stimuli presented in this task were two circular apertures containing diagonally oriented gratings drifting perpendicular to their angle (see Figure 1b). Each grating had a diameter of 4 degrees and a spatial frequency of 1 cycle per degree. Each eye receives orthogonal gratings: the left eye sees a grating oriented at 45°. In the right eye, participants would see a grating oriented at 135°. Above each circle are the numbers 1-5.

Motion Parallax Stimuli:

In this task, a large grey square with one small grey square in each quadrant was presented to each eye (see Figure 1c). In the centre of each large grey square, there is a fixation cross. Each corner of the black and white frame contained a number ranging from 1-4. Stimuli on each trial slightly differed based on the condition of the task being presented. The task was programmed so that on every trial, the large grey square tilts very quickly on the vertical plane. During this time, one of the small squares appears to have depth. In the parallax only condition, the amount of depth seen in the square that is popping out is proportional to the amount of parallax that would elicit that depth. In the parallax + disparity condition, there is additional disparity added to the square that is popping out. This is done by shifting the position of the pop-out square horizontally between the left and right eyes.

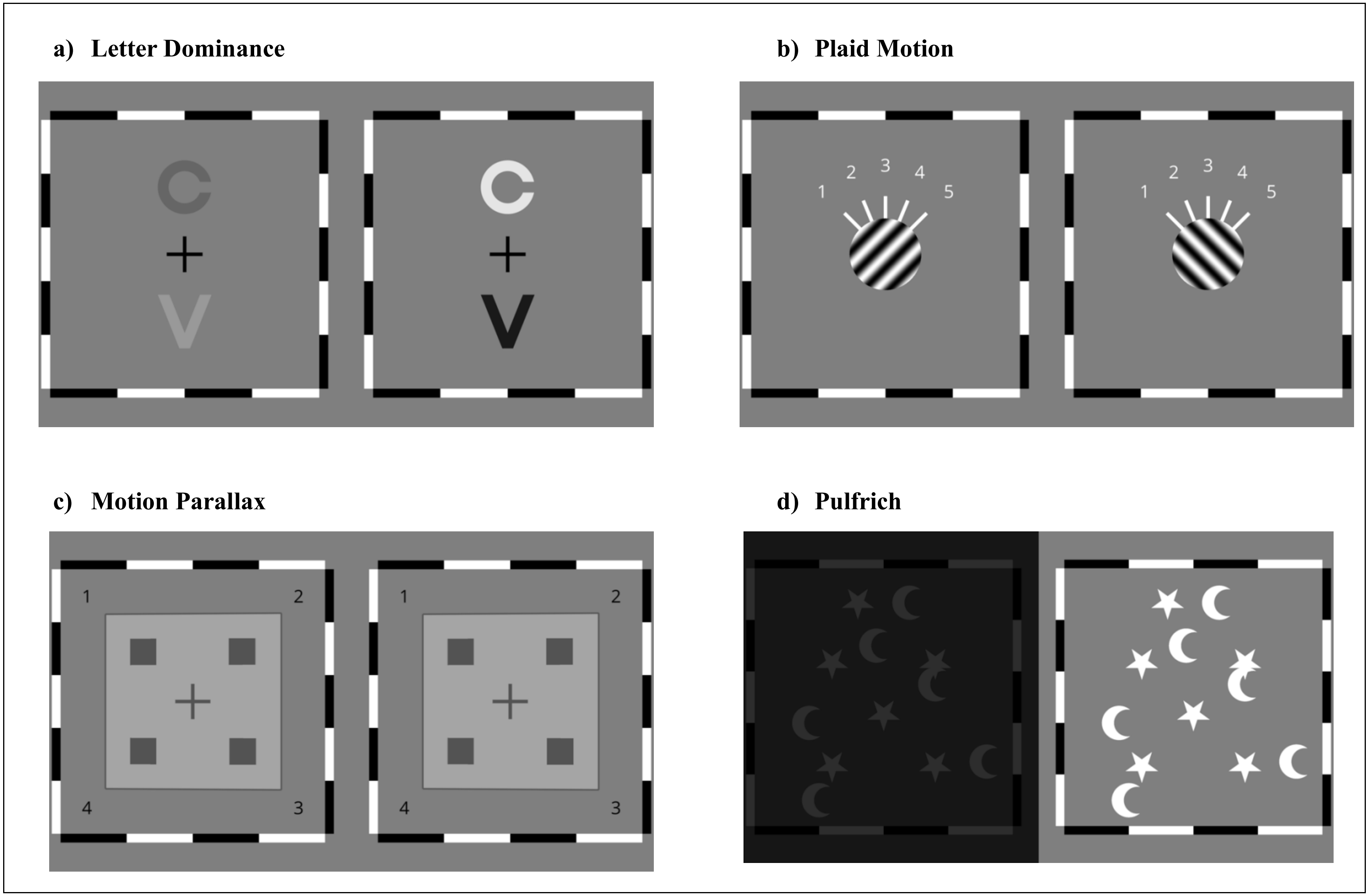
Pulfrich Stimuli:

Participants completed two versions of the digital Pulfrich task: the Disparity task and the Sweeping task. In the Disparity task, participants were presented with white stationary stars and moons moving leftwards across the screen (see Figure 1d). At the bottom of the screen, there was one stationary moon. A neutral density filter would be placed over one eye on the monitor to elicit the Pulfrich illusion so that the moving moons appeared either in front of or behind the stars.

In the Sweeping task, participants were presented with the same Pulfrich display. However, there was no stationary moon, and the neutral density filter was selected so that it would present on the same side as the participant's dominant eye to begin with. The filter is always presented at 100% darkness over the participant's dominant eye.

All four binocular vision tests were generated and controlled using the python libraries in PsychoPy® and were presented to the participant on the LG monitor in 4k display. Prior to testing, the LG monitor was linearized using a PR®-650 SpectraScan® Colorimeter to ensure the monitor displayed stimuli with accurate and consistent levels of luminance.

Figure 1
Binocular Vision Task Stimuli



Design and Procedure

Participants sat with their chin in the chin rest, facing the stereoscope. The table height was adjusted upwards or downwards until the participant reported being comfortably seated. Next, participants were presented with the initial fixation.

Binder clips on the outermost mirrors of the stereoscope were removed and participants were instructed to close their left eye and use their right eye to look through the right inner mirror while adjusting the sheet of the paper on the right outer mirror. Participants were instructed to adjust the sheet of paper until they saw the entire black and white frame on the monitor. This was repeated with the alternate eye to ensure participants could see the entire frame within which stimuli were presented.

Once participants were finished with the adjustments on the outermost mirrors, participants were asked to look through the stereoscope with both of their eyes open and report whether they viewed a perfect cross (+) in the center of the screen. If yes, participants would proceed to the first binocular vision task. If no, participants would be asked whether they wanted the vertical line to be moved to the left or right until the participant viewed a perfect, stable cross.

Letter Dominance:

Once participants fused the images seen by both eyes, the participant would view two letters on the screen, one letter brighter than the background and the other darker. Since the stimuli presented to each eye are rivalrous, it will produce the perception of two letters of a medium gray luminance, but one letter will appear relatively brighter depending on which eye is dominant. While maintaining fixation on the central plus sign, the participant indicates which letter appears brighter (see Appendix C for full participant instructions).

There were 28 trials in the practice task. Practice task data plots are printed immediately by the program and are identical to the data plots described earlier for actual trials. The x-axis represents contrast presented to the left eye, plotted against proportion of left eye responses on the y-axis. To achieve normal practice performance, the data should form a psychometric (s-shaped) curve with a PSE between the range of 0.4-0.6. If the experimenter determined that the participant did not understand the task based on the practice plot, the participant would be reinstructed and complete the practice task again. This was common for the Letter Dominance task. All participants completed 140 trials in the experimental task.

It is important to note that participants in experiment 1b were tested on a larger range of contrast values than those in experiment 1a. This task deploys a two interleaved staircase procedure followed by a method of constant stimuli to refine the threshold measurement. One staircase began with a contrast of 0.8 and one with a contrast of 0.2. The method of constant stimuli then tested seven contrasts centered around the participant's estimated balance point (balance point, ±0.03, ±0.06, and ±0.09).

Plaid Motion:

Once the participant fuses the stimuli presented to each eye, the participant would view one circle containing gratings moving in different directions based on the individual's fusion state. The possible directions of movement were listed above the circle and ranged from 1–5: 1 indicating the most leftward movement possible and 5 indicating the most rightward movement possible. There were two fused state responses, sliding and 3, with 3 indicating perfect fusion. The sliding response is to be used when the participant sees the gratings moving in various directions at once. The 3 response is to be used when the participant sees the gratings moving consistently upwards or downwards. There are 3 practice trials. All participants completed 3 experimental trials which were each one minute long.

Motion Parallax:

Participants were instructed to fixate on a central cross for the duration of the task. During each trial, the large square would tilt for a very brief period of time and one of the small squares would appear to pop out. Participants were instructed to indicate which of the small squares had depth by verbally stating the corresponding number for the square where they saw depth. All participants completed practice trials. Participants repeated the actual task three times.

It is important to note that due to a bug in the code, data from participants 1-21 in experiment 1a are not interpretable.

Pulfrich:

Participants first completed one practice trial. On this practice trial, participants would view the same stimuli presented in the Disparity Task. Participants were instructed to move the stationary moon using the keypad so that it matched the depth of the moving moons.

Next, participants would complete the Sweeping Task. The participant was instructed to use their keypad and adjust the strength of the neutral density filter. The filter would begin to reduce in luminance until it completely disappeared. Then, it would begin to darken over the participant's non-dominant eye. Participants were asked to verbally report four points—where they began to see depth between the moons and stars, where they stopped seeing depth, the second point where they began to see depth, and the final point where they stopped seeing depth.

Lastly, before engaging in the Disparity task, participants who used version 1 of the response box had to complete a response box alignment procedure. Since the moons and stars on the box were not aligned at 0cm, we conducted an alignment procedure to record where each participant perceived zero depth between the moons and stars. During this procedure, the experimenter moved the moons to four different distances (two in front of the stars, two behind). The participant was instructed to move the moons so that they aligned with the stars at each of the four distances. The distances at which participants moved the moons were recorded.

Identical to the practice trial, participants were instructed to move the stationary moon using the keypad so that it matched the depth of the moving moons. After each trial, participants were instructed to slide the set of moons in the response box so that it is the same distance away from the stars that they viewed on the monitor.

Participants in experiment 1a completed 3 trials at 4 different neutral density filter values for a total of 12 trials in the Disparity Task. Participants in experiment 1b completed 3 trials at 6 neutral density filter values for a total of 18 trials. It is important to note that subjects 21-25 in experiment 1a and subjects 1-12 in experiment 1b used the second version of the response box.

Participants returned exactly a week later, at the same time, and completed all four tasks again. The duration of each session was less than 90 minutes, and breaks were encouraged as needed.

Analysis

All analyses were conducted in R Studio running R version 4.2.2 (R Core Team, 2021).

Letter Dominance

To obtain the contrast threshold where a participant's eyes are balanced, the program deploys two interleaved staircases followed by a method of constant stimuli. When plotting contrast presented to the left eye against proportion of left eye responses, the result is a psychometric function. From this curve, we plotted a linear regression line and extracted each participant's point of subjective equality (PSE). A PSE of around 0.5 is typical of control participants and represents equal binocular balance between the two eyes. If the participant required more contrast to be presented to their left eye to achieve binocular balance (PSE > 0.5), we concluded that they are right eye dominant. If they needed less contrast presented to their left eye (PSE < 0.5), we concluded that they are left eye dominant. To assess test–retest reliability, a Pearson correlation analysis as well as an Intraclass correlation coefficient were conducted using week 1 and week 2 PSE values.

Plaid Motion

We grouped participant responses based on whether they were in the fused or unfused state. Responses 1 and 2 indicate left eye dominance and responses 4 and 5 indicate right eye dominance therefore these were grouped as unfused responses. Responses 3 and “sliding” indicate strong binocular fusion and were grouped as fused responses. By evaluating the frequency of responses, we were able to extract the proportion of time each participant spent in a state of binocular fusion. A Pearson correlation analysis and an Intraclass correlation coefficient were conducted using the proportion of time spent in binocular fusion on week 1 and week 2.

Motion Parallax

The Motion Parallax task uses a two-down, one-up staircase procedure. We calculated the threshold for each participant for each condition– parallax and parallax+disparity. The threshold calculation was done by taking the geometric mean of the last six reversals on each staircase. Since participants completed this task three times, we averaged the thresholds for each condition. We then calculated the difference between the two thresholds, in other words the performance advantage, on each week for each participant. A Pearson correlation analysis and an Intraclass correlation coefficient were conducted using this difference score on week 1 and week 2.

Pulfrich

To measure participant accuracy on the task, we plotted disparity responses made by the participant against the neutral density filter values tested and calculated a line of best fit. The same was done for data from the response boxes. The slopes of these lines (perceived depth/contrast) were extracted and compared between each week for each participant. A Pearson correlation analysis and an Intraclass correlation coefficient were conducted using the perceived depth/contrast values on week 1 and week 2.

Results

Exclusions and Fit Failures

Experiment 1a:

Of the initial 25 participants, 2 participants were excluded from the analysis as they did not attend the session on week 2. 1 participant was excluded from the Letter Dominance analysis due to an error in the code. From the remaining 22 participants, 9 participants were excluded from the Letter Dominance analysis. Our exclusion criteria for the Letter Dominance task are two-fold. First, it must be clear based on the psychometric curve that the participant did not understand the task and second, it must be very difficult to fit a line to the curve.

Experiment 1b:

Of the initial 12 participants, 2 participants were excluded from the analysis as they did not attend the session on week 2. 2 participants were excluded from the Letter Dominance analysis due to an error in the code. From the remaining 8 participants, 3 participants were excluded from the Letter Dominance analysis based on the criteria described above. 1 participant was excluded from the Pulfrich analysis as it was clear they misunderstood the task as they responded with nearly identical answers on every trial.

Pearson Correlation Analysis

There was a significant positive correlation between PSE values on week 1 and week 2 in the Letter Dominance task, $r(19)= .85$, $p=1.1e-06$ (see Figure 2a). There was a significant positive correlation between proportion of time spent in binocular fusion on week 1 and week 2 in the Plaid Motion task, $r(31)= .73$, $p=1.3e-06$ (see Figure 2b). There was a significant positive correlation between perceived depth/contrast values on week 1 and week 2 in the Pulfrich Disparity task, $r(30)= .74$, $p=1.5e-06$ (see Figure 2c). There was a significant positive correlation between perceived depth/contrast values on week 1 and week 2 in the Pulfrich Response Box task, $r(30)= .70$, $p=9.2e-06$ (see Appendix D). No significant correlation was found between performance advantage values on week 1 and week 2 in the Motion Parallax task $r(13)= .11$, $p=.71$ (see Figure 2d). To view individual data plots for each task including exclusions, see Appendix E.

Intraclass Correlation Analysis

Intraclass correlation coefficients (ICC) are often preferred over Pearson's correlation coefficients for test–retest reliability studies. The ICC evaluates both the consistency and agreement of scores whereas Pearson's correlations only measure whether two scores change together in a systematic way. ICC values can tell us not just whether two scores are linearly related but also if they are similar in value. We calculated an ICC (3,1) for each task—a two-way mixed-effects model for single measurements, assessing consistency (see Table 1). We obtained an ICC(3,1) of 0.85 for the Letter Dominance task indicating excellent reliability. An ICC(3,1) of 0.71 indicated good reliability for the Plaid Motion task. An ICC(3,1) of 0.70 indicated good reliability for the Pulfrich Disparity task. An ICC(3,1) of 0.69 indicated good reliability for the Pulfrich Response Box task. Lastly, an ICC(3,1) of 0.14 indicated poor reliability for the Motion Parallax task.

Table 1

Intraclass Correlation Coefficients

Table 1. Intraclass Correlation Coefficients

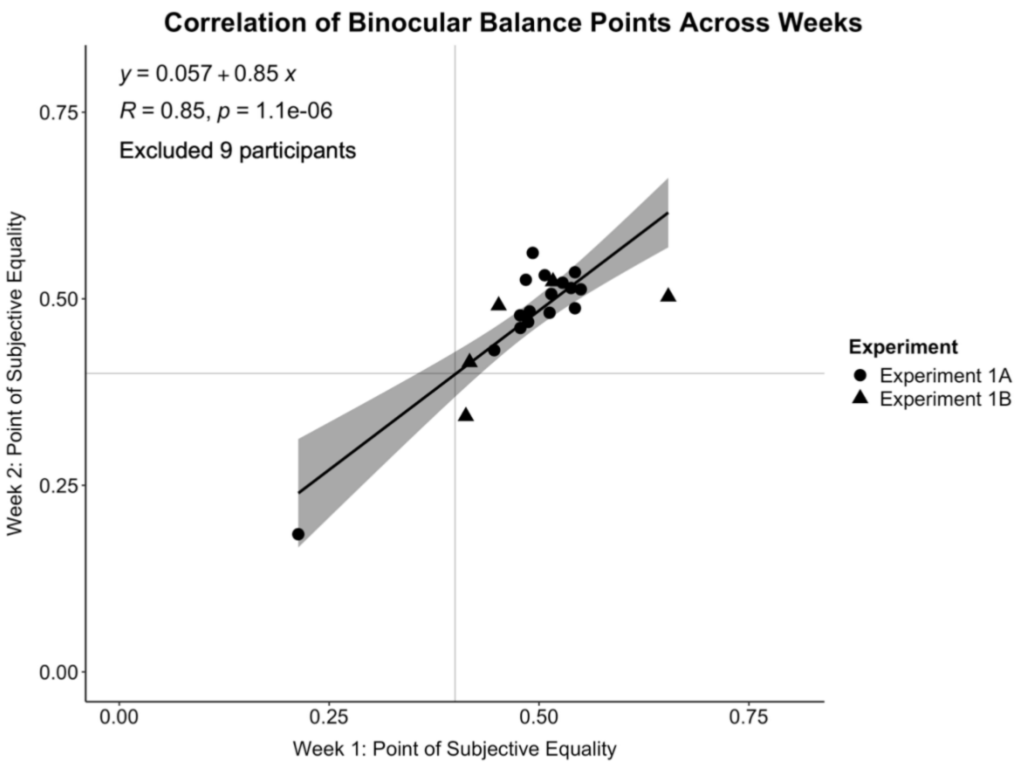
Task	ICC	F	df1	df2	p	lower bound	upper bound	subjects (n)	raters (k)
Letter Dominance	0.85	12.0	20	20	2.7e-07	0.67	0.94	21	2
Plaid Motion	0.71	5.9	32	32	1.2e-06	0.49	0.85	33	2
Pulfrich Disparity	0.70	5.7	31	31	2.8e-06	0.47	0.84	32	2
Pulfrich Response Box	0.69	5.4	31	31	4.9e-06	0.45	0.83	32	2
Motion Parallax	0.14	1.3	13	13	3.1e-01	-0.40	0.61	14	2

Single measures, two-way mixed effects model, consistency definition.

Figure 2

Correlation Plots

A) Letter Dominance



B) Plaid Motion

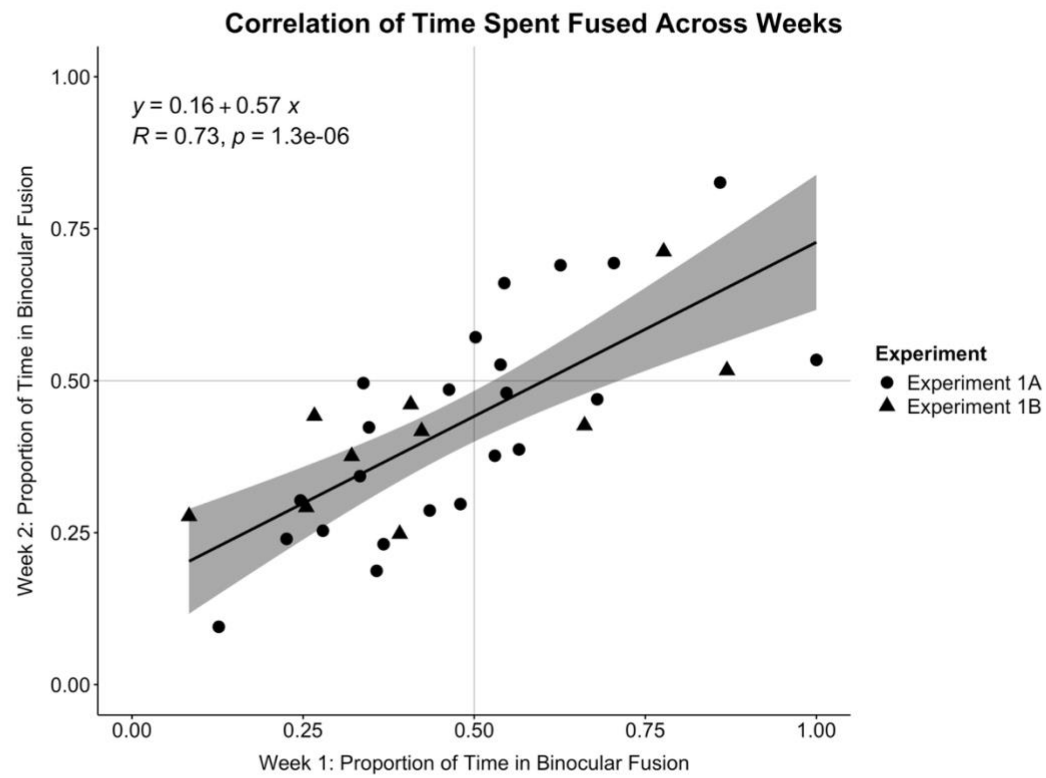
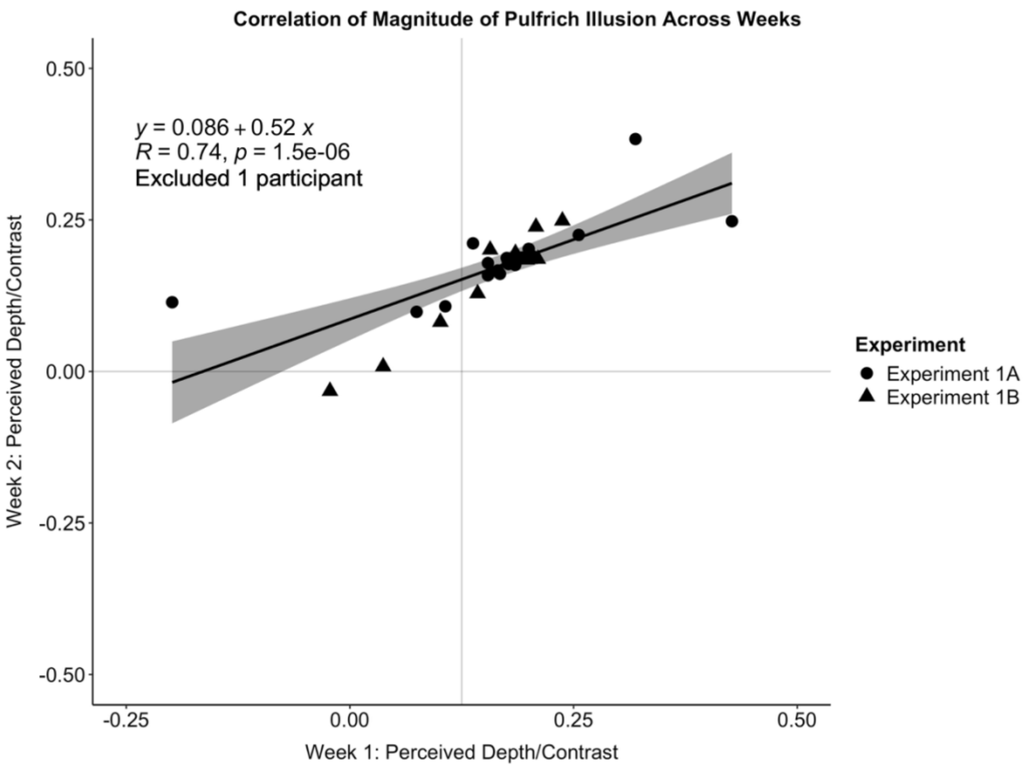


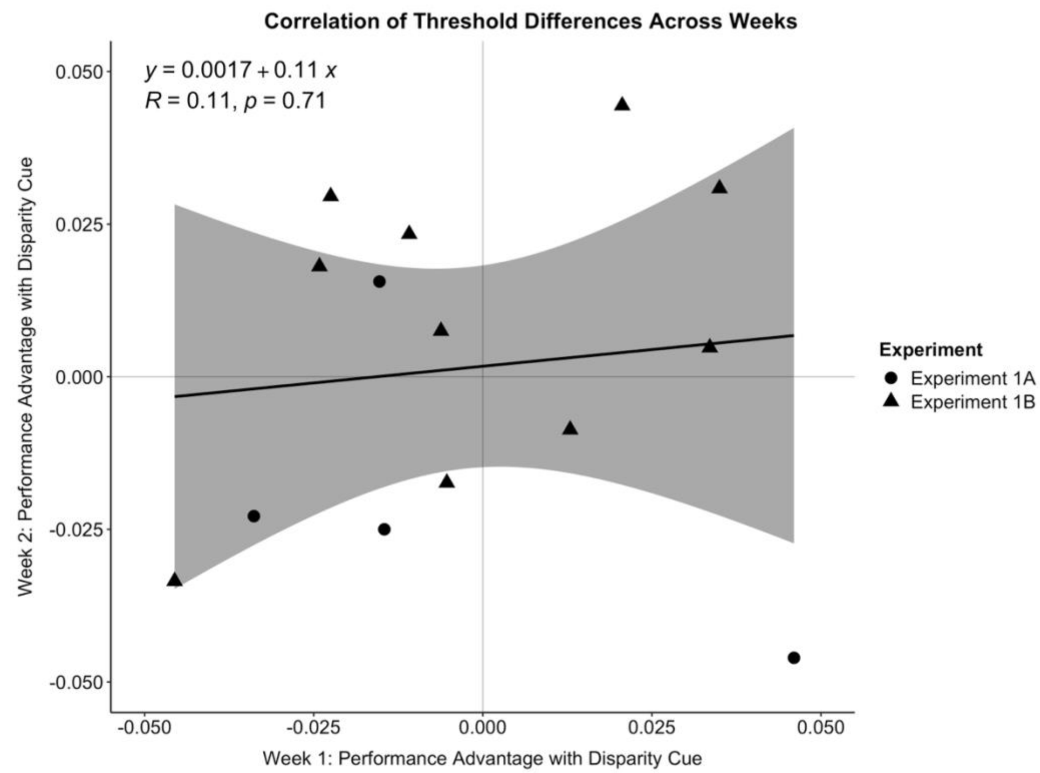
Figure 2 Cont'd

Correlation Plots

C) Disparity Pulfrich



D) Motion Parallax



Discussion

The present study investigated the test–retest reliability of a battery of novel tasks developed to measure binocular vision in patients with amblyopia. Current clinical tests of binocular vision often use static stimuli to obtain measures of binocular vision (i.e., stereoacuity and interocular suppression). Given that patients with amblyopia who lack clinical binocularity can perceive the Pulfrich effect (Maehara et al., 2019), we proposed that stimuli in motion may tap into different aspects of binocular vision than static stimuli, potentially offering additional information. To measure these aspects of binocular vision which we have labelled *hidden binocularity,* we developed a battery of tasks with stimuli in motion. Here we tested the reliability for four of the tasks: Letter Dominance, Plaid Motion, Pulfrich, and Motion Parallax. The first three tasks demonstrated strong reliability ($r = .85$, ICC = .85; $r = .73$, ICC = .71; $r = .74$, ICC = .70 respectively). The Motion Parallax task did not demonstrate reliability ($r = .11$, ICC = .14). These findings highlight the importance of evaluating test–retest reliability *prior* to investigating individual differences in a patient population.

Implications

Overall, results from this study demonstrate that three of our developed binocular vision tests reliably measure aspects of binocular vision. In other words, these tasks demonstrate individual differences in binocular vision that are reliable across time.

Letter Dominance

The Letter Dominance task provides us with a measure of eye dominance. Based on which eye is dominating in an amblyopic patient, we can obtain a measure of the degree of suppression in the amblyopic eye.

Plaid Motion

The Plaid Motion task can reveal fusion abilities retained by an amblyopic patient that may not be captured by static tests such as the Worth 4 Dot test. In the presence of high individual variability between subjects, the proportion of time spent in a fused state was reliable across our two tests.

Pulfrich

Observing a reliable Pulfrich illusion allows us to probe this phenomenon of hidden binocularity more closely with the goal of discovering where in the visual system the integration across the two eyes is coming from. For example, by comparing monocularly and binocularly deprived patients (Lewis & Maurer, 2009), we can explore how no vision versus vision to one eye impacts depth from motion. Moreover, the physical response boxes can provide us with a measure of perceived distance available to adults and children without stereovision .

The Sweeping Pulfrich task will allow us to explore this even further as we can measure a participant's interocular latency threshold for the onset and disappearance of the Pulfrich effect. In other words, we are measuring how much luminance difference between the eyes is necessary to cause a sufficient processing delay in one eye to produce the Pulfrich effect—as well as the amount of luminance difference necessary to eliminate it. Integration of binocular vision does not solely rely on aligning images spatially—it relies on temporal synchronization as well. Exploring individual variability in this task as well as comparing thresholds between control participants and patients can inform us on the temporal sensitivity of binocular integration in amblyopic patients. This may provide critical insight for the rehabilitation of binocular vision in this population.

Motion Parallax

Once reliability is demonstrated, we hope to use the Motion Parallax task to explore whether amblyopic patients with hidden binocularity can detect the disparity advantage in the condition of the task where both the parallax cue and disparity cue are used to create depth in the stimulus. This would provide evidence for the use of motion-in-depth stimuli to measure binocular vision in this population.

We have proposed a few possible explanations to explain the lack of reliability observed in the Motion Parallax task. There are many reasons we initially chose a staircase procedure to measure performance thresholds on this task as opposed to a method of constant stimuli. Firstly, staircases concentrate trials around a participant's threshold requiring fewer trials to obtain a threshold estimate. Staircases avoid presenting trials that are too easy or too difficult for the participant and adapt to the participants' performance in real time. They are also often shorter and adapt to individual performance which is ideal for special populations such as children with amblyopia. Given we had two different stimulus conditions, we specifically used two interleaved staircases. Interleaved staircases reduce response anticipation or strategy use. In tasks where a single staircase is used, participants may be able to predict the pattern of stimulus changes (i.e., every time I make a mistake, the task gets easier). Since the two staircases alternate, performance is less impacted by a single error and the output includes two independent estimates of the threshold providing a more robust final estimate. Although staircase methods offer significant benefits, their implementation in psychophysical tasks requires careful consideration. Our team is still in the process of adapting our staircase procedure to ensure it accurately captures a participant's threshold.

In the first version of our Motion Parallax task, there was an error in the code that created visible disparity when the target had zero depth. In this version of the task, we viewed clear practice effects and staircases were variable even towards the end of trials. Once this error was corrected, we no longer saw such significant practice effects and observed more stable staircases around participants' thresholds. Unfortunately, the threshold advantage in the parallax+disparity condition we anticipated largely disappeared. It's possible that we may have overcorrected for this error and made our task too difficult. Participants often reported that they were guessing on most trials as they could not see a square popping out to them during the brief stimulus presentation. Additionally, participants complete this task three times making it one of the longest tasks in the battery. It is important to note that despite not seeing reliability when conducting a Pearson's correlation on the difference score between the two conditions, when a Pearson's correlation is calculated on the raw thresholds for each condition, the task demonstrates reliability (see Appendix F). This finding highlights the general difficulty in obtaining reliability with difference scores. In order to ensure this task demonstrates reliability, we have modified the task and plan to test participants on each condition separately and then conduct further analyses.

Limitations – Challenges of Employing Tools in Development

All of the tasks tested here are in development, meaning we do not have any well-established protocols for running the tasks. This lead to significant challenges in data collection. Task difficulty provides one challenge. Some tasks are more difficult than others and thus required modified instruction and minor modifications during testing. For example, the Letter Dominance task requires a difficult determination of which of two almost identical images is brighter—many participants did not understand the task. We modified the instructions to ask participants to answer as quickly as possible, which prevented them from overthinking their response and cleaned up the data. The need to respond quickly aligns with results from binocular rivalry studies (Nikiforova, Cowell & Huber, 2024). Over time, the initial percept changes as the processes of rivalry are engaged. This can explain why faster response times aid the participant in completing the task more accurately. Another concern early on in the Letter Dominance task was the range of values tested in the experiment. We had set our range too narrowly so that some participants never achieved a PSE. We increased the range of values tested in Experiment 1b, which resulted in fewer exclusions. However, the nature of the Letter Dominance task—that it taps into rivalry processes—may necessitate that some participants are excluded for not reaching a PSE.

Since the tasks were *in development*, problems arose that were corrected along the way. The first version of our response boxes were not sturdy. The adhesive used was crazy glue and the wires were all loose. Both of the boxes we originally assembled broke multiple times and needed repairs. Ultimately, we redesigned the box with more sturdy connections and better containment of the wires and power source. Since the boxes were hand-made, there was some error in the position of the moons and the stars relative to the scale to read the response. When the moons and the stars are aligned, the scale should read 0 cm; however, each box had an error of ±1 cm. In our redesigned box, we ensured the measurement bar aligned with the stimuli in the box. As we again redesign the box, we are making further improvements based on our experience thus far.

Limitations – Multi-Session Study Designs

There are also several limitations that come with conducting lengthy, multi-session studies. Firstly, multi-session studies introduce intra-individual variability. Given our participants were undergraduate students, a participant may differ on factors such as fatigue or stress between the two sessions impacting their motivation or attention levels during the study. Our study was 1.5 hours per session, which is longer than typical studies that are 1 hour in length. Moreover, our study involved several components (i.e., vision screening, the use of the response box) and it was evident that our study was fatiguing to some participants.

Multi-session study designs may also introduce learning effects. This is evident in the Motion Parallax task as participants often have lower threshold values on both conditions on run 2 compared to run 1. In a separate study, we tested one amblyopic individual with no clinical binocularity and three controls on the battery of tasks over consecutive days (Nichols et al., in preparation). By day 11, the amblyopic individual began to demonstrate performance similar to controls. Additionally, when tested on the clinical Worth 4 Dot test, the amblyopic individual experienced binocular fusion which was not present on every testing day prior. We propose a reduction in interocular suppression contributed to this finding in combination with general learning effects from repeated testing.

Limitations – Control Population

Lastly, it is important to mention that most of our participants were female, and all our participants were undergraduate students meaning our sample is likely from a western, educated, industrialized, rich, and democratic (WEIRD) society (Henrich et al., 2010). This will limit the generalizability of our findings.

Future Directions

Establishing the reliability of these tasks is critical in developing rehabilitation paradigms for this population as these paradigms largely rely on re-establishing binocular vision (Hess & Thompson, 2015). Establishing the reliability of this battery was a necessary first step for our research team as the purpose of this battery is twofold: (1) we aim to assess hidden binocularity possessed by patients with amblyopia and (2) we hope to correlate performance on these tasks with performance on tasks of visuomotor control. Our team predicts that patients with greater hidden binocularity will perform better on tasks of visuomotor control which would inform rehabilitation. We are most interested in the efficient planning and execution of prehension movements since these movements are most needed for effective rehabilitation.

Conclusion

Amblyopia is the leading cause of partial or total vision loss among children and young adults around the world (Fu et al., 2020). There were approximately 99.2 million people in 2019 with the condition, a number predicted to nearly double within the next decade (Fu et al., 2020). It has become evident that current clinical tests of binocular vision may not accurately capture the full extent of the binocular capabilities of an individual with amblyopia. Motion-in-depth stimuli may serve as a useful adjunct for assessing binocular and stereoscopic function in this population. Accurate quantification of binocular vision in this population is of critical importance as rehabilitating binocular vision is at the core of rehabilitation paradigms for this population. Our team developed a battery of binocular vision tasks which may provide more sensitive measures of binocular vision in this population given we use motion-in-depth stimuli.

The current study found that three of the tasks in our battery demonstrate test–retest reliability and can be used to explore hidden binocularity in patients. This was a necessary first step in evaluating whether these tasks were suitable to explore individual and group differences in this patient population. Future studies will involve evaluating the reliability of the Motion Parallax task as well as the four other tasks in the battery which are yet to be tested.

References

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*(5892), 523–525. https://doi.org/10.1038/300523a0

Andrews, T. J., & Blakemore, C. (2001). Integration of motion information during binocular rivalry. *Vision Research*, *42*(3), 301–309. https://doi.org/10.1016/S0042-6989(01)00286-3

Baker, D. H., Meese, T. S., Mansouri, B., & Hess, R. F. (2007). Binocular Summation of Contrast Remains Intact in Strabismic Amblyopia. *Investigative Ophthalmology & Visual Science*, *48*(11), 5332–5338. https://doi.org/10.1167/iovs.07-0194

Birch, E. E., Castañeda, Y. S., Cheng-Patel, C. S., Morale, S. E., Kelly, K. R., Beauchamp, C. L., & Webber, A. (2018). Self-perception of School-aged Children With Amblyopia and Its Association With Reading Speed and Motor Skills. *JAMA Ophthalmology*, *137*(2), 167–174. https://doi.org/10.1001/jamaophthalmol.2018.5527

Birch, E. E., Kelly, K. R., & Giaschi, D. E. (2019). Fellow Eye Deficits in Amblyopia. *Journal of binocular vision and ocular motility*, *69*(3), 116–125. https://doi.org/10.1080/2576117X.2019.1624440

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1998) Statistical Power Analysis for the Behavioural Sciences. Lawrence Erlbaum Associates, Hillsdale.

Chow, A., Silva, A. E., Tsang, K., Ng, G., Ho, C., & Thompson, B. (2021). Binocular Integration of Perceptually Suppressed Visual Information in Amblyopia. Investigative Opthalmology & Visual Science, 62(12), 11. https://doi.org/10.1167/iovs.62.12.11

Daw, N. W. (1998). Critical Periods and Amblyopia. *Archives of Ophthalmology*, *116*(4), 502–505. https://doi.org/10.1001/archopht.116.4.502

Niechwiej-Szwedo, E., Kennedy, S. A., Colpa, L., Chandrakumar, M., Goltz, H. C., & Wong, A. M. (2012). Effects of induced monocular blur versus anisometropic amblyopia on saccades, reaching, and eye-hand coordination. *Investigative Ophthalmology & Visual Science*, *53*(8), 4354-4362.

Fu, Z., Hong, H., Su, Z., Lou, B., Pan, C. W., & Liu, H. (2020). Global prevalence of amblyopia and disease burden projections through 2040: a systematic review and meta-analysis. *British Journal of Ophthalmology*, *104*(8), 1164-1170.

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

H. Schiller, P., L. Kendall, G., C. Kwak, M., & M. Slocum, W. (2012). Depth Perception, Binocular Integration and Hand-Eye Coordination in Intact and Stereo Impaired Human Subjects. *Journal of Clinical & Experimental Ophthalmology*, *03*(02). https://doi.org/10.4172/2155-9570.1000210

Hamm, L., Chen, Z., Li, J., Black, J., Dai, S., Yuan, J., Yu, M., & Thompson, B. (2017). Interocular suppression in children with deprivation amblyopia. *Vision Research*, *133*, 112–120. https://doi.org/10.1016/j.visres.2017.01.004

Handa, T., Mukuno, K., Uozato, H., Niida, T., Shoji, N., & Shimizu, K. (2004). Effects of Dominant and Nondominant Eyes in Binocular Rivalry. *Optometry and Vision Science*, *81*(5), 377. https://doi.org/10.1097/01.opx.0000135085.54136.65

Heng, S., & Dutton, G. N. (2011). The Pulfrich effect in the clinic. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *249*(6), 801–808. https://doi.org/10.1007/s00417-011-1689-6

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, *33*(2-3), 61-83.

Hess, R. F., Mansouri, B., & Thompson, B. (2010). A new binocular approach to the treatment of amblyopia in adults well beyond the critical period of visual development. *Restorative neurology and neuroscience*, *28*(6), 793-802.

Hess, R. F., Mansouri, B., & Thompson, B. (2011). Restoration of Binocular Vision in Amblyopia. *Strabismus*, *19*(3), 110–118. https://doi.org/10.3109/09273972.2011.600418

Hess, R. F., & Thompson, B. (2015). Amblyopia and the binocular approach to its therapy. *Vision Research*, *114*, 4–16. https://doi.org/10.1016/j.visres.2015.02.009

Huang, C.-B., Zhou, J., Zhou, Y., & Lu, Z.-L. (2010). Contrast and Phase Combination in Binocular Vision. *PLOS ONE*, *5*(12), e15075. https://doi.org/10.1371/journal.pone.0015075

Hubel, D. H., & Wiesel, T. N. (1964). Binocular interaction in striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology*, *28*(6), 1041–1059. https://doi.org/10.1152/jn.1965.28.6.1041

Kelly, K. R., Morale, S. E., Beauchamp, C. L., Dao, L. M., Luu, B. A., & Birch, E. E. (2020).

Factors Associated with Impaired Motor Skills in Strabismic and Anisometropic Children. *Investigative ophthalmology & visual science*, *61*(10), 43. https://doi.org/10.1167/iovs.61.10.43

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Levi, D. M., Knill, D. C., & Bavelier, D. (2015). Stereopsis and amblyopia: A mini-review. *Vision Research*, *114*, 17–30. https://doi.org/10.1016/j.visres.2015.01.002

Levi, D. M., McKee, S. P., & Movshon, J. A. (2011). Visual deficits in anisometropia. *Vision Research*, *51*(1), 48–57. https://doi.org/10.1016/j.visres.2010.09.029

Lewis, T. L., & Maurer, D. (2009). Effects of early pattern deprivation on visual development. *Optometry and Vision Science*, *86*(6), 640-646.

Maehara, G., Araki, S., Yoneda, T., Thompson, B., & Miki, A. (2019). Suprathreshold Motion Perception in Anisometropic Amblyopia: Interocular Speed Matching and the Pulfrich Effect. *Optometry and Vision Science*, *96*(6), 434–442. https://doi.org/10.1097/OPX.0000000000001381

McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficits in amblyopia. *Journal of Vision*, *3*(5), 5. https://doi.org/10.1167/3.5.5

Nawrot, M., & Joyce, L. (2006). The pursuit theory of motion parallax. *Vision Research*, *46*(28), 4709–4725. https://doi.org/10.1016/j.visres.2006.07.006

Nikiforova, M. S., Cowell, R. A., & Huber, D. E. (2023). Gestalt formation promotes awareness of suppressed visual stimuli during binocular rivalry. *Visual cognition*, *31*(1), 18–42. https://doi.org/10.1080/13506285.2023.2192991

Pulfrich C. (1922) Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. *Naturwissenschaften*, *10*(553–564). doi: 10.1007/BF01571319.

R Core Team. (2021). R: a language and environment for statistical computing.

Tidbury, L. P., Brooks, K. R., O’Connor, A. R., & Wuerger, S. M. (2016). A Systematic Comparison of Static and Dynamic Cues for Depth Perception. *Investigative Ophthalmology & Visual Science*, *57*(8), 3545–3553. https://doi.org/10.1167/iovs.15-18104

Verghese, P. (2023). The utility of peripheral stereopsis. *Frontiers in Neuroscience*, *17*. https://www.frontiersin.org/articles/10.3389/fnins.2023.1217993

Wang, Y., He, Z., Liang, Y., Chen, Y., Gong, L., Mao, Y., Chen, X., Yao, Z., Spiegel, D. P., Qu, J., Lu, F., Zhou, J., & Hess, R. F. (2019). The Binocular Balance at High Spatial Frequencies as Revealed by the Binocular Orientation Combination Task. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00106

Appendix A: Clinical Vision Screening Data

Clinical Vision Screening Data
Experiment 1a

Subject	Age/Sex	Week	Acuity (OD)	Acuity (OS)	Stereoacuity	Binocular Fusion	Eye Dominance
1	18F	Week 1	20/20 -1	20/20 -1	10/10	Fused	OD
1	18F	Week 2	20/20 -1	20/20 -1	9/10	Fused	OD
2	18F	Week 1	20/20 -1	20/20	9/10	Fused	OD
2	18F	Week 2	20/20 -2	20/20	10/10	Fused	OD
3	18F	Week 1	20/20 -2	20/20 -2	10/10	Fused	OS
3	18F	Week 2	20/20 -1	20/20 -1	9/10	Fused	OS
4	18F	Week 1	20/20	20/20	7/10	Fused	OD
4	18F	Week 2	20/20	20/20	7/10	Fused	OD
5	18F	Week 1	20/20	20/20	8/10	Fused	OD
5	18F	Week 2	20/20	20/20	9/10	Fused	OD
6	18F	Week 1	20/20 -2	20/20 -1	8/10	Fused	OS
6	18F	Week 2	20/20 -1	20/20 -1	10/10	Fused	OS
7	18F	Week 1	20/20 -2	20/20 -2	10/10	Fused	OS
7	18F	Week 2	20/20 -1	20/32 -1	8/10	Fused	OS
8	18M	Week 1	20/20 -2	20/20	9/10	Fused	OD
8	18M	Week 2	20/20 -2	20/20 -2	8/10	Fused	OD
9	18F	Week 1	20/20 -2	20/32 -2	8/10	Fused	OD
9	18F	Week 2	20/20 -1	20/20 -1	9/10	Fused	OD
10	19F	Week 1	20/20	20/20	7/10	Fused	OD
10	19F	Week 2	20/20	20/20	9/10	Fused	OD
11	19F	Week 1	20/32 -1	20/20 -2	9/10	Fused	OS
11	19F	Week 2	20/20 -2	20/20 -2	10/10	Fused	OD
12	18M	Week 1	20/20	20/20 -1	9/10	Fused	OS
12	18M	Week 2	20/20 -1	20/20	9/10	Fused	OS
13	16F	Week 1	20/20 -1	20/20 -1	8/10	Fused	OS
13	16F	Week 2	20/20	20/20	8/10	Fused	OS
14	18F	Week 1	20/20	20/20	10/10	Fused	OS
14	18F	Week 2	20/20	20/20	10/10	Fused	OS
15	25M	Week 1	20/20	20/20	9/10	Fused	OD
15	25M	Week 2	20/20	20/20	10/10	Fused	OD
17	18F	Week 1	20/20 -1	20/20	9/10	Fused	OD
17	18F	Week 2	20/20	20/20	9/10	Fused	OD
18	18M	Week 1	20/20 -1	20/20 -2	6/10	Fused	OS
18	18M	Week 2	20/20 -1	20/20 -1	8/10	Fused	OS
19	19F	Week 1	20/20 -1	20/20 -1	8/10	Fused	OD
19	19F	Week 2	20/20 -2	20/20 -2	8/10	Fused	OD
20	18F	Week 1	20/20	20/20	10/10	Fused	OS
20	18F	Week 2	20/20	20/20	8/10	Fused	OS
22	19F	Week 1	20/32 -1	20/20 -1	4/10	Fused	OS
22	19F	Week 2	20/20 -2	20/20	8/10	Fused	OS
23	19M	Week 1	20/20 -1	20/20 -1	9/10	Fused	OD
23	19M	Week 2	20/20	20/20 -1	10/10	Fused	OD
24	18M	Week 1	20/20	20/20	9/10	Fused	OD
24	18M	Week 2	20/20	20/20	8/10	Fused	OD
25	19F	Week 1	20/20 -1	20/20 -1	10/10	Fused	OD
25	19F	Week 2	20/20 -2	20/20	10/10	Fused	OD

Note: OD: Right eye; OS: Left eye, VA measured using Lighthouse chart at time of test, Stereoacuity measured using Randot at time of test, Binocular Fusion measured using Worth 4 Dot at time of test, Eye Dominance measured using Miles Test at time of test, Subjects denoted with * did not demonstrate normal stereoacuity

Table A1 Clinical vision screening data for participants in experiment 1a.

Clinical Vision Screening Data

Experiment 1b

Subject	Age/Sex	Week	Acuity (OD)	Acuity (OS)	Stereoacuity	Binocular Fusion	Eye Dominance
1	20M	Week 1	20/25	20/20 -1	9/10	Fused	OS
1	20M	Week 2	20/20 -2	20/20 -1	10/10	Fused	OS
3*	21F	Week 1	20/25	20/25	5/10	Fused	OD
3*	21F	Week 2	20/20 -1	20/20	6/10	Fused	OD
5	18M	Week 1	20/25 -2	20/25 -2	9/10	Fused	OS
5	18M	Week 2	20/25 -1	20/20 -1	10/10	Fused	OS
6	18M	Week 1	20/20 -2	20/20 -1	10/10	Fused	OD
6	18M	Week 2	20/20 -2	20/20 -2	9/10	Fused	OD
7	20M	Week 1	20/20 -2	20/32 -1	8/10	Fused	OD
7	20M	Week 2	20/20 -1	20/20 -1	8/10	Fused	OD
8*	21F	Week 1	20/25 -2	20/25	5/10	Fused	OD
8*	21F	Week 2	20/25 +2	20/25	6/10	Fused	OS
9	22F	Week 1	20/25 -1	20/25	10/10	Fused	OD
9	22F	Week 2	20/20 -1	20/20	9/10	Fused	OD
10	18F	Week 1	20/20	20/20	10/10	Fused	OD
10	18F	Week 2	20/20 -1	20/20	10/10	Fused	OD
11	21F	Week 1	20/20	20/20	8/10	Fused	OD
11	21F	Week 2	20/25 -2	20/20 -1	9/10	Fused	OD
12	19F	Week 1	20/50 -2	20/25	9/10	Fused	OD
12	19F	Week 2	20/40 -2	20/40 -2	8/10	Fused	OD

Note: OD: Right eye; OS: Left eye, VA measured using Lighthouse chart at time of test, Stereoacuity measured using Randot at time of test, Binocular Fusion measured using Worth 4 Dot at time of test, Eye Dominance measured using Miles Test at time of test, Subjects denoted with * did not demonstrate normal stereoacuity

Table A2 Clinical vision screening data for participants in experiment 1b.

Appendix B: Apparatus and Stimuli

Figure B1 Pictured on the left is the participant seating and chin rest setup. Pictured next are the innermost mirrors on the mirror stereoscope. The metal edges of the mirrors have been covered with white paper. On the right are the two large outer mirrors of the stereoscope. The mirrors have been covered with sheets of paper with a rectangle cut out in the centre to restrict participants' view to the monitor.

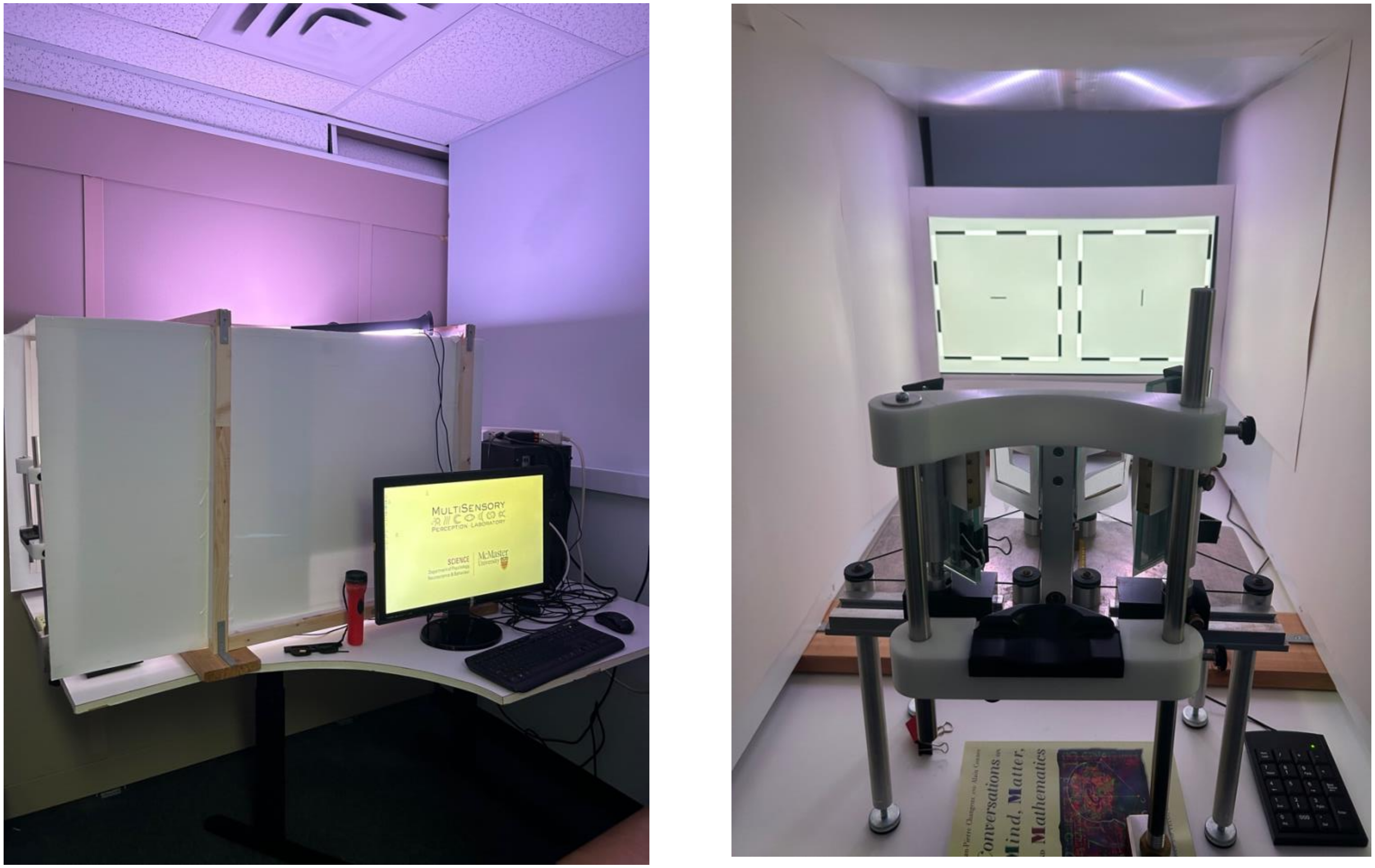


Figure B2 The first version of the enclosure was built with a wooden frame. White corrugated plastic sheets were placed on the inside to create an enclosure surrounding the setup. White Bristol board was adhered onto the plastic sheets to eliminate reflections from the monitor.

Figure B3 The second version of the enclosure was built with plywood that was then painted black and attached using L brackets.

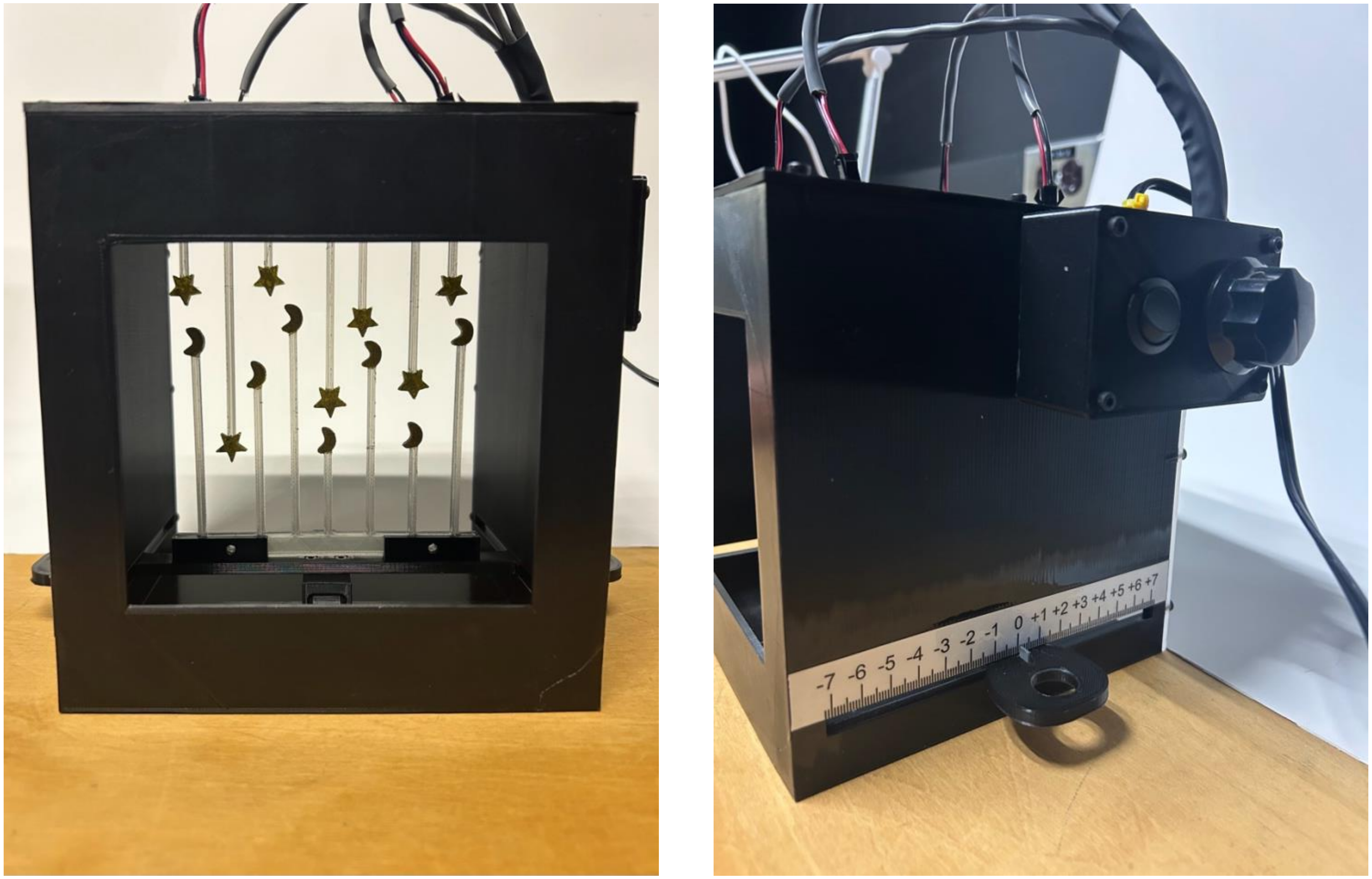


Figure B4 Pictured is the modified response box. Plastic stars hang from the ceiling of the box and a plastic set of moons are attached to a slider. Using the slider, the participants can change the location of the moons relative to the stars. The scales on the sides of the box are in cm. The switch and dial on the side of the box adjust the LED lights on the ceiling of the box

Appendix C: Participant Instructions

Letter Dominance:

"Right now, you should see 2 letters on the screen, one on the top and one at the bottom separated by a cross in the middle."

"For this task, you need to indicate which letter appears brighter to you. You have been provided with a keypad to your right. Please use the 8 key if you feel the top letter is brighter and 5 for the bottom."

Give the participant a moment to get oriented.

"One letter may be a dark shade and one may be a light shade. Please do your best to ignore the shades of the letters but instead, response with which one looks brighter to you."

"Remember to look at the central cross the entire time and do not look at the letters directly."

"If you are ever unsure, please take your best guess. It is best not to hesitate during this task and to answer as quickly as possible."

Note: Participants normally respond instantly to the letters. If a participant seems to take long to respond, remind them that they don't need to overthink and simply select whichever letter appears brighter to them overall.

Use the Subject Checklist to complete practice interpretation. If the results do not look good, reinstruct the participant and practice again. Press **Enter** to leave the practice.

Ask the participant if they are ready -> start actual test "3 = Letter Dominance Beta"

Make notes of any factors that could affect test results and any feedback from the participant using the BV data collection sheet.

Plaid Motion:

Before you instruct the participant, open the practice trials and press **Enter** to present the demo.

Note: The first trial of the practice trials is a demo, and you must press **Enter** to move on to the actual practice trials.

"Right now, you should see gratings moving in a particular direction. This direction will keep changing during the test or it may stay the same. There are five numbers on top which are the possible movement directions."

"You will need to tell me the movement direction you are seeing using the numbers above the circle. You do not need to use the keypad for this task, you can just state the numbers out loud. Please remember, you are not reporting which way the lines are facing but instead, which direction they are moving in.

Show the participant the diagram now.

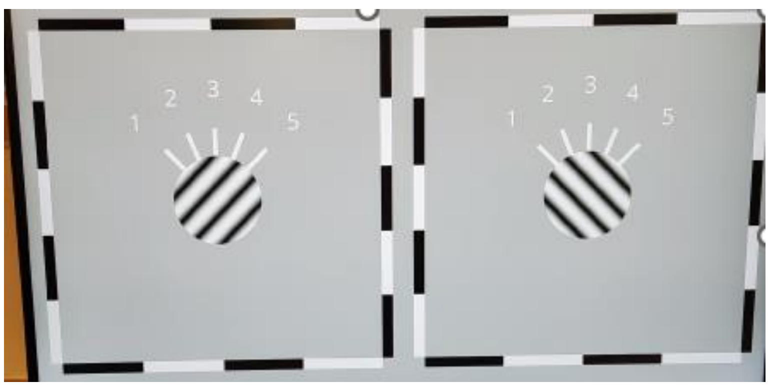
"Sometimes you may see gratings moving consistently upwards or downwards. For this, you would say number 3 out loud."

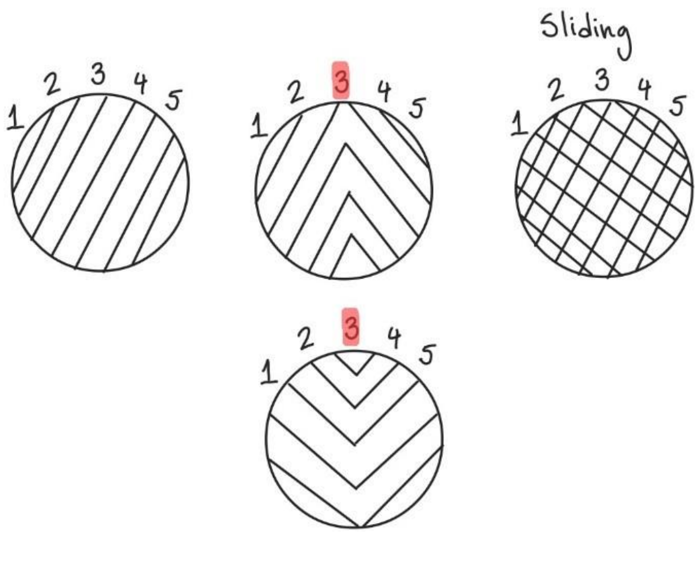
"Sometimes you may see parts of gratings moving in different directions at the same time. OR you might see a plaid design. In this case, you can say "sliding" out loud."

"This is a continuous test. It is a time-sensitive measurement, and there are no right or wrong answers so please don't hesitate to tell me what movement direction you're seeing whenever it changes."

Press **Enter** to leave the practice. Ask the participant if they are ready -> start actual trial. Note observation and feedback.

Plaid Motion Diagrams:





Motion Parallax:

"In a moment, you will see one large square with one small square in each corner. When a trial starts, the large square will tilt very quickly. During that brief time, one of the four small squares will appear to be floating compared to the others."

"The numbers 1, 2, 3, and 4 correspond to each of the small squares in the corners of the larger square. Your task is to say the number out loud for the small square which you saw floating."

"Please make sure you are looking at the center of the cross the entire time."

Use the Subject Checklist to complete practice interpretation. If the results do not look good, reinstruct the participant and practice again. Press **Enter** to leave the practice.

Ask the participant if they are ready -> start actual test "2 = Motion Parallax"

Note observation and feedback.

Important Note: This task is repeated 3 times. You do not need to exit the program.

Pulfrich:

Practice

Press "2" on your keyboard to select Match Practice first. Press "Enter" and skip the response box alignment. Press "Enter" once more and you should see moons and stars on the screen.

Participants will only complete ONE trial in this practice session. This is just a demo so that they know what the Pulfrich effect looks like during the Luminance Sweep.

"Right now, you should see moons moving across the screen, stationary stars, and one stationary moon at the bottom of the screen. Does it look like the moving moons are in front of/behind the stars?"

"Now I would like to direct you to the stationary moon at the bottom of the screen. Is the stationary moon on the same level as the moving moons?"

Participants should say no.

"Using the slash on your keypad, you can move the stationary moon forward. Using the star, you can move it backwards. Your task is to move the stationary moon to the same depth as the moving moons. Please note you cannot press and hold the keys to make the moon move faster."

Note: Make sure the participant understands that they are not moving the stationary moon left or right. Make sure they understand they are not moving the moving moons.

Once the participant finishes one trial, exit this demo.

Luminance Sweep

Press "1" on your keyboard to open "Luminance Sweep." Based on the subject checklist, select the non-dominant/amblyopic eye on screen.

"For this task, you are going to press either 4 or 6 on your keypad (please tell the participant the correct number based on their eye dominance) and report any changes in what you are seeing as you press down on this key. Please start by pressing the key slowly and stop and let me know the very first point where you begin to see depth between the moons and the stars. This can look like the moon in front or behind of the stars as you previously viewed in the demo."

Note: You should only be sampling FOUR points for control participants. 2 over each eye.

Once the participant reports depth, click the + on your keyboard (use the main keyboard, not the keypad). On the screen, it will say "Sampled!" That is how you will know the ND filter was recorded.

"Now, please keep pressing the key and stop and let me know when there is no depth between the moons and stars. You can press and hold the key as well to speed up the process."

Take another sample using the + on your keyboard.

"Now, please keep pressing the key and stop and let me know when you begin to see depth between the moons and stars again. The moons and stars may get really flat before you start to see depth again."

Take another sample using the + on your keyboard.

"Finally, please keep pressing the key and stop and let me know when you see no depth between the moons and stars. The depth between the moons and stars may get larger before it completely goes away."

Take the last sample using the + on your keyboard.

Click Enter to end the task.

Match Actual Task (18 trials)

Press "4" on your keyboard to select Match Actual Task Beta. Instruct participants.

"Similar to the practice trial you completed, your task is to use the slash and star on your keypad again and match the depth of the stationary moon to that of the moving moons."

After the participant completes one trial, instruct them to use the response box.

"Now, please remember the depth between the moving moons and the stars on your screen. Using the slider on the response box, please slide the moving moons so that they are at the same depth away from the stars as you viewed on the monitor."

Important Note: Always use the reading on the right side of the response box.

Type that in and press the + key on the keypad and Enter to move onto the next trial.

The task will exit once all trials are completed.

Appendix D: Pulfrich Response Box Correlation Plot

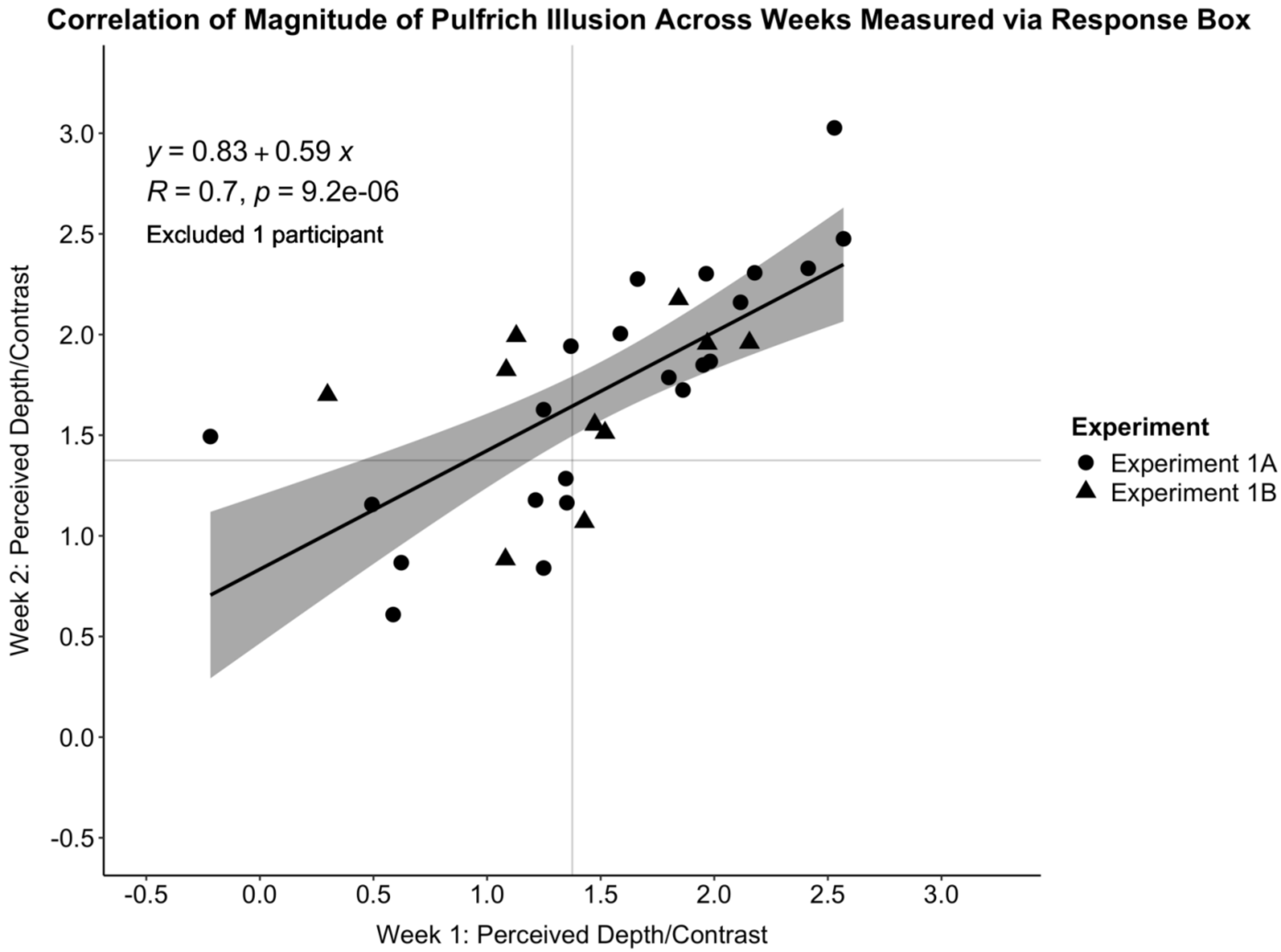


Figure D1 Pulfrich Response Box Task Correlation Plot

Appendix E: Individual Data Plots for All Tasks

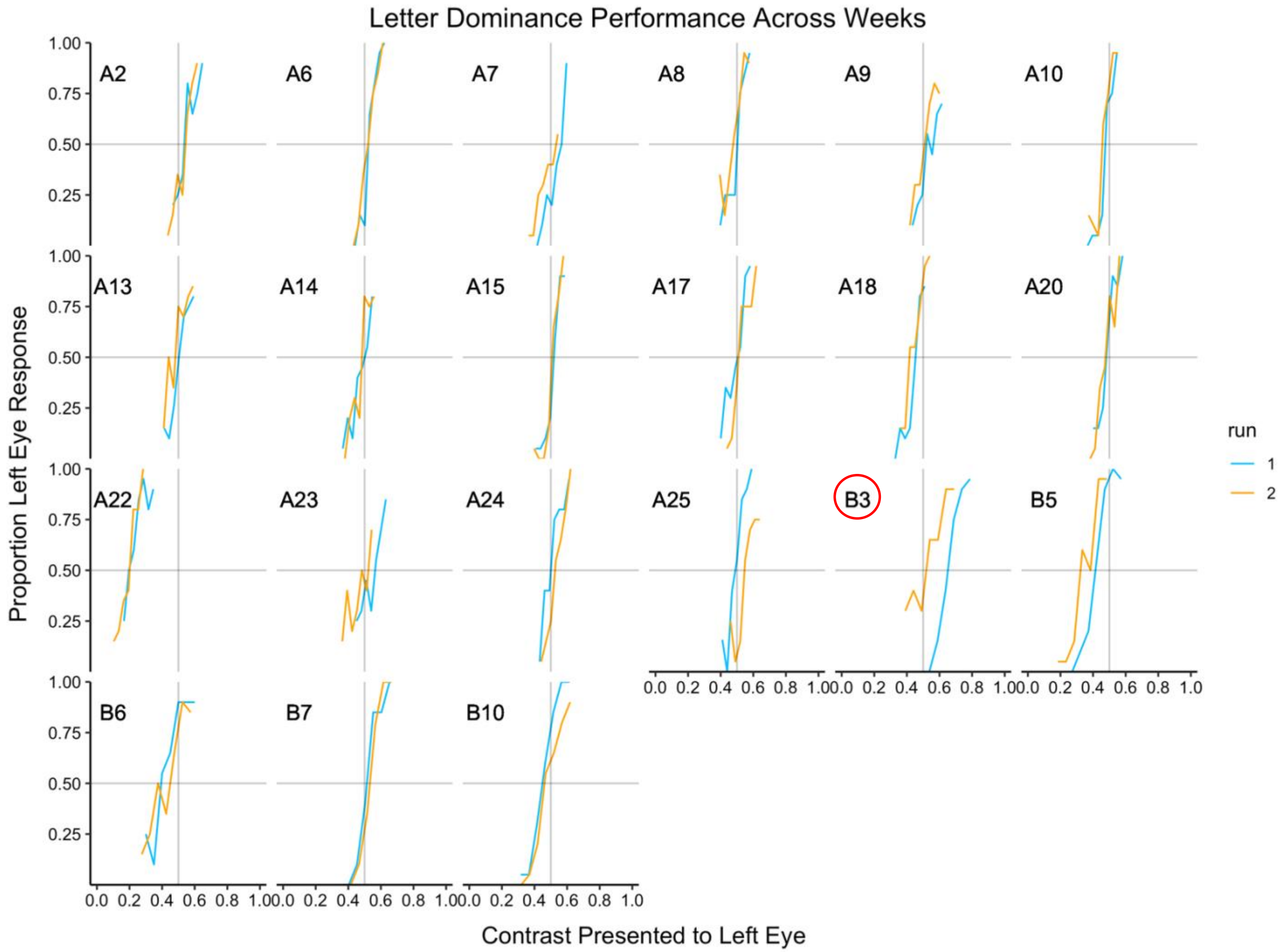


Figure E1 Individual data plots for the Letter Dominance task. Subjects identified with a red circle did not pass vision screening.

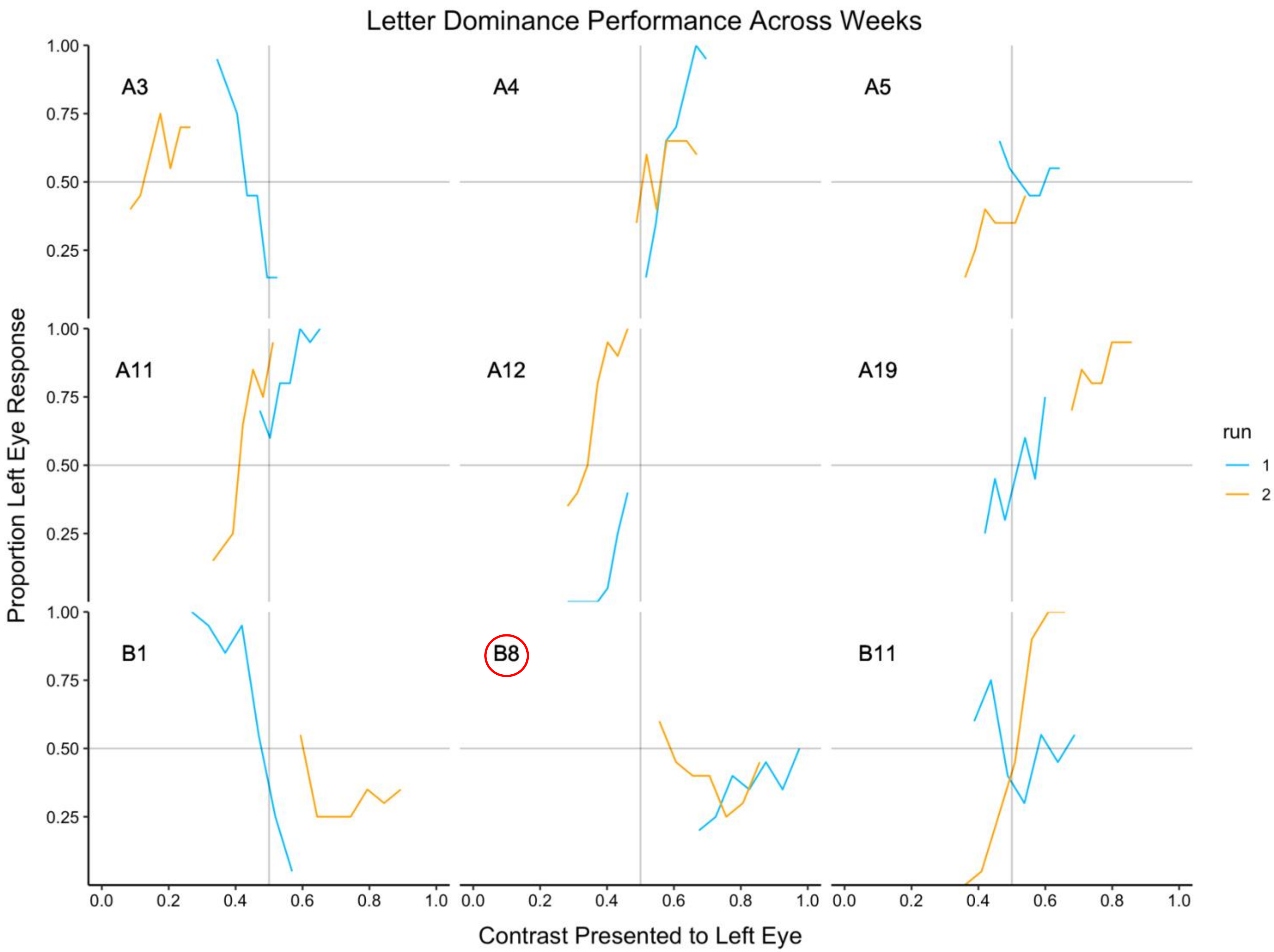


Figure E2 Individual data plots for participants excluded from the Letter Dominance analysis.

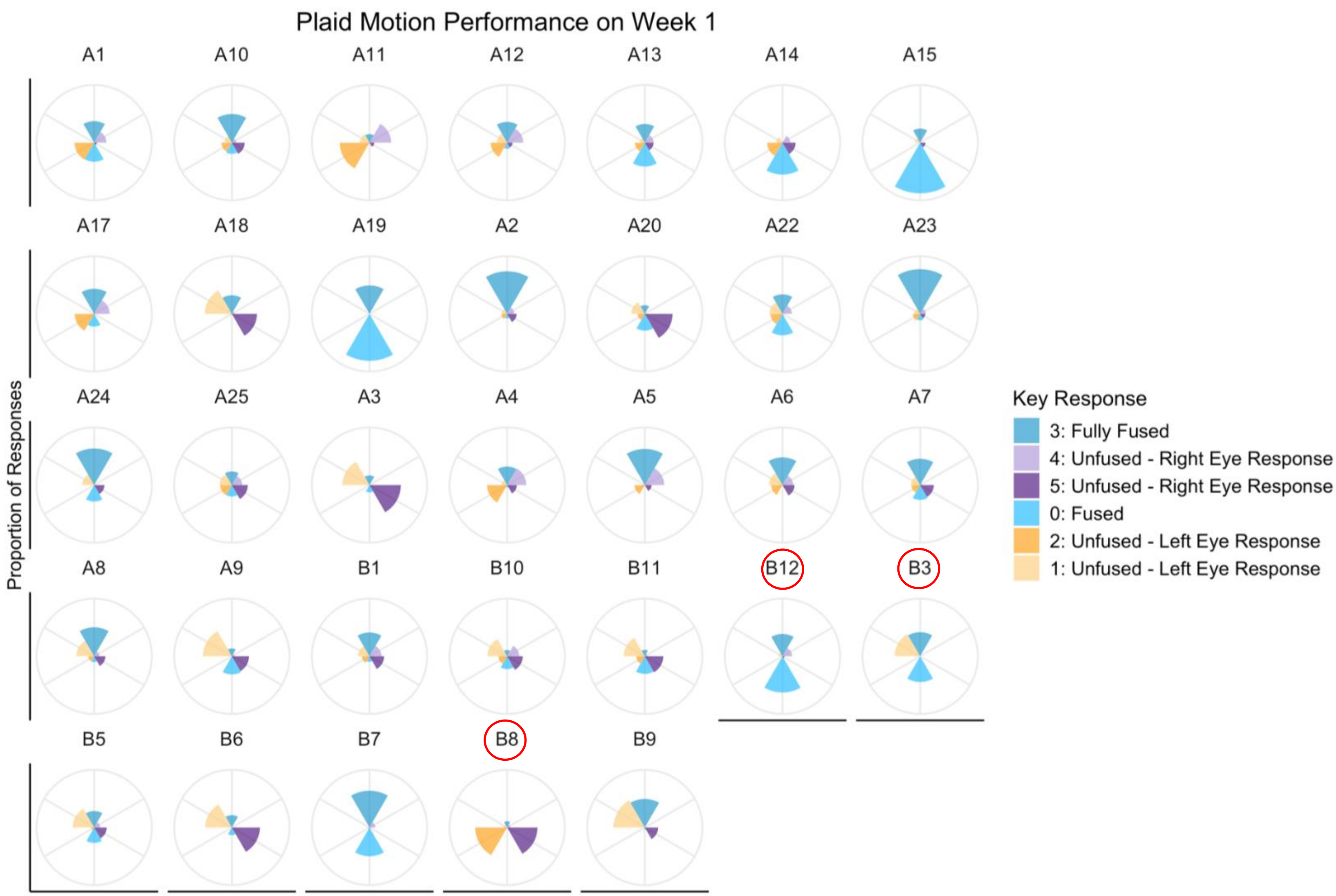


Figure E3 Individual data plots for the Plaid Motion task on week 1.

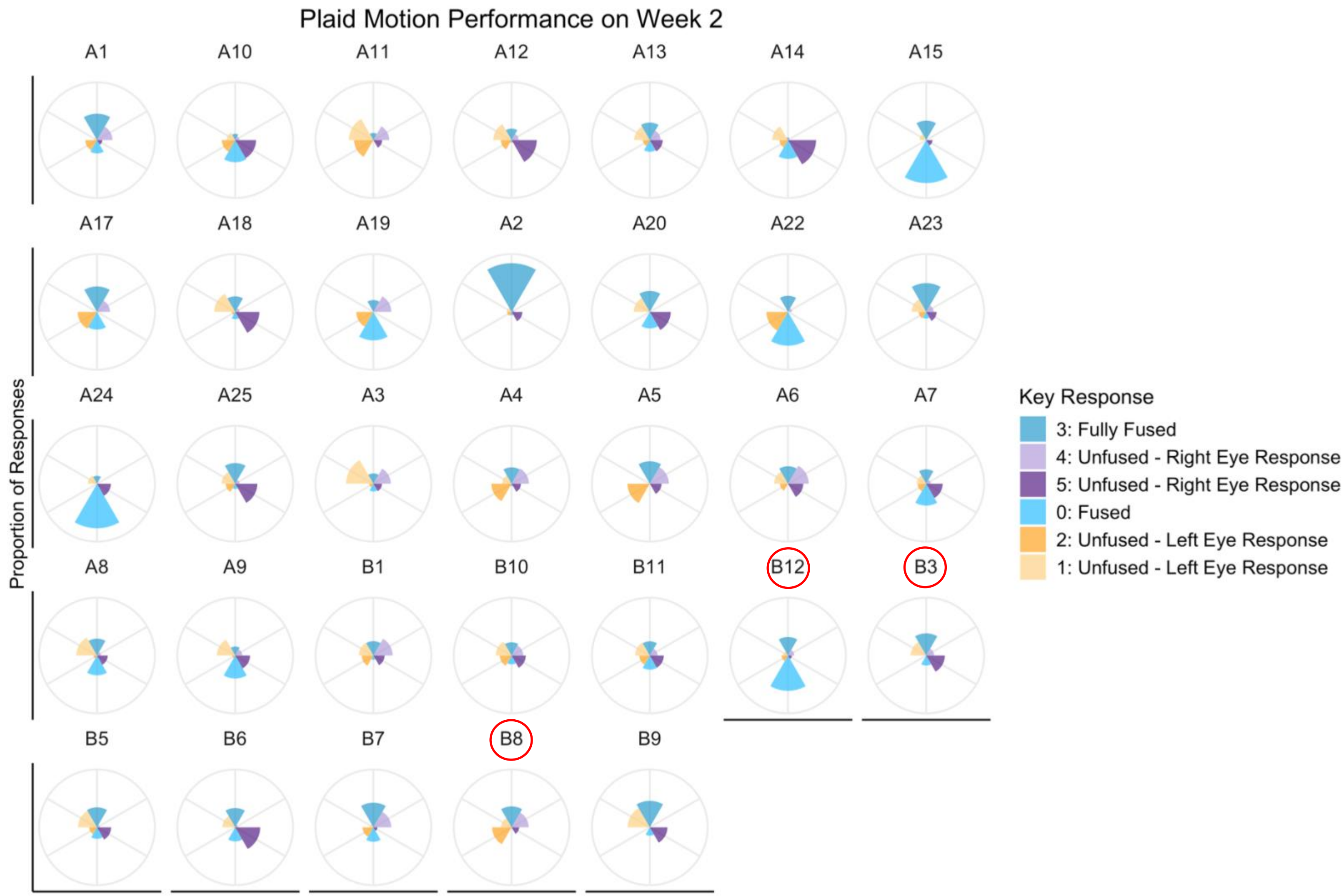


Figure E4 Individual data plots for the Plaid Motion task on week 2.

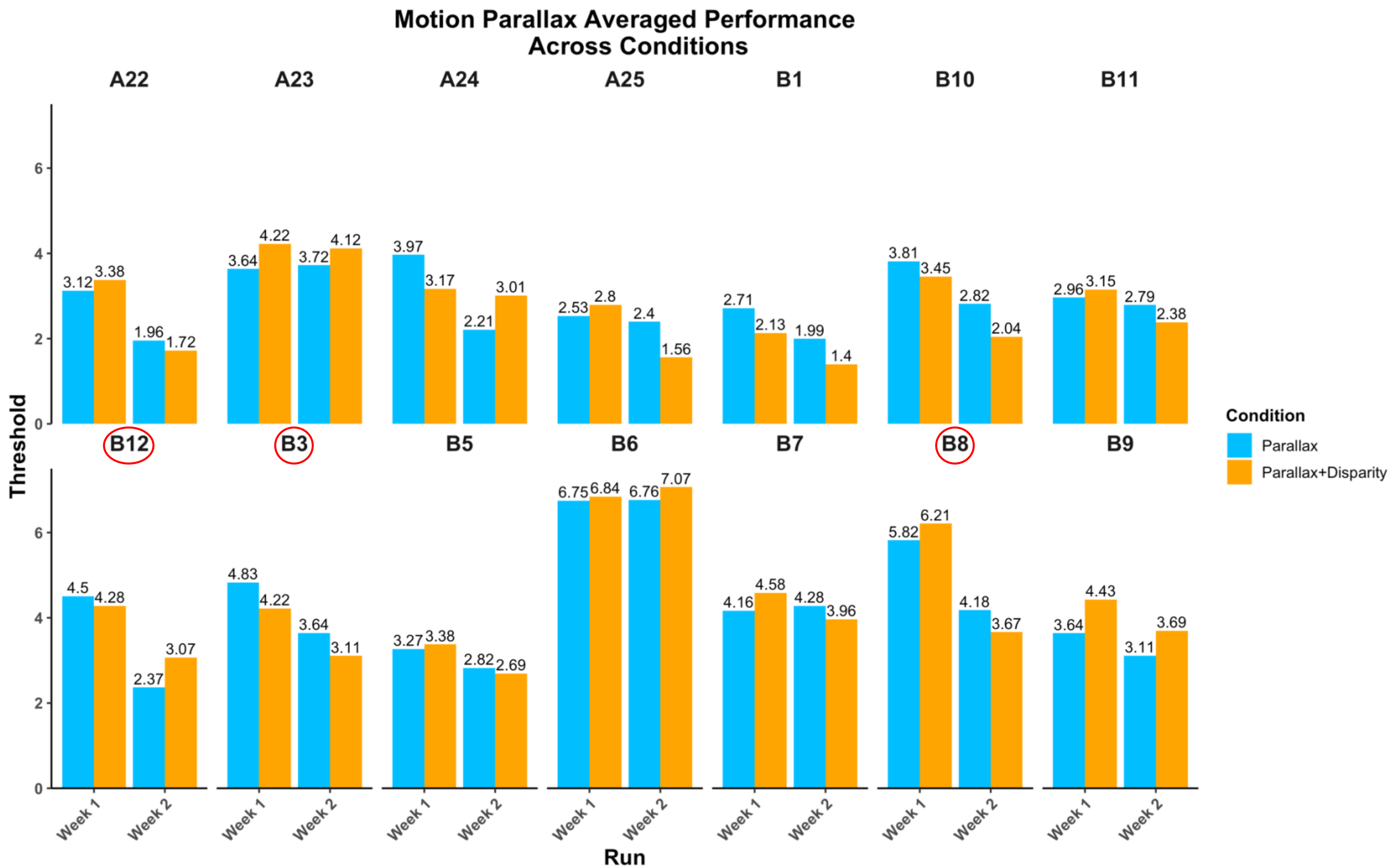


Figure E5 Individual data plots for the Motion Parallax task.

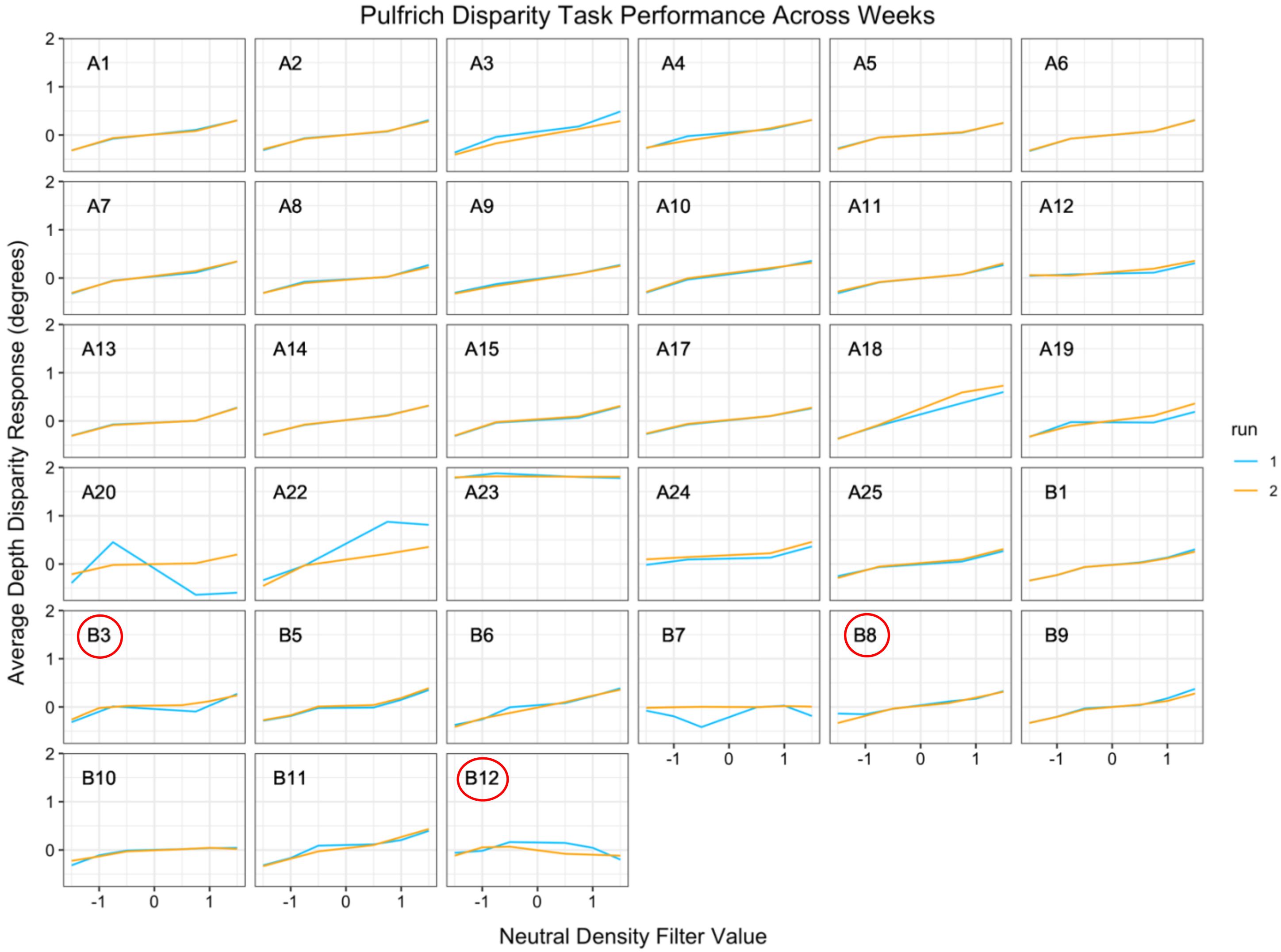


Figure E6 Individual data plots for the Pulfrich Disparity task. Subject 23 was excluded from the correlation analysis due to a clear misunderstanding of the task.

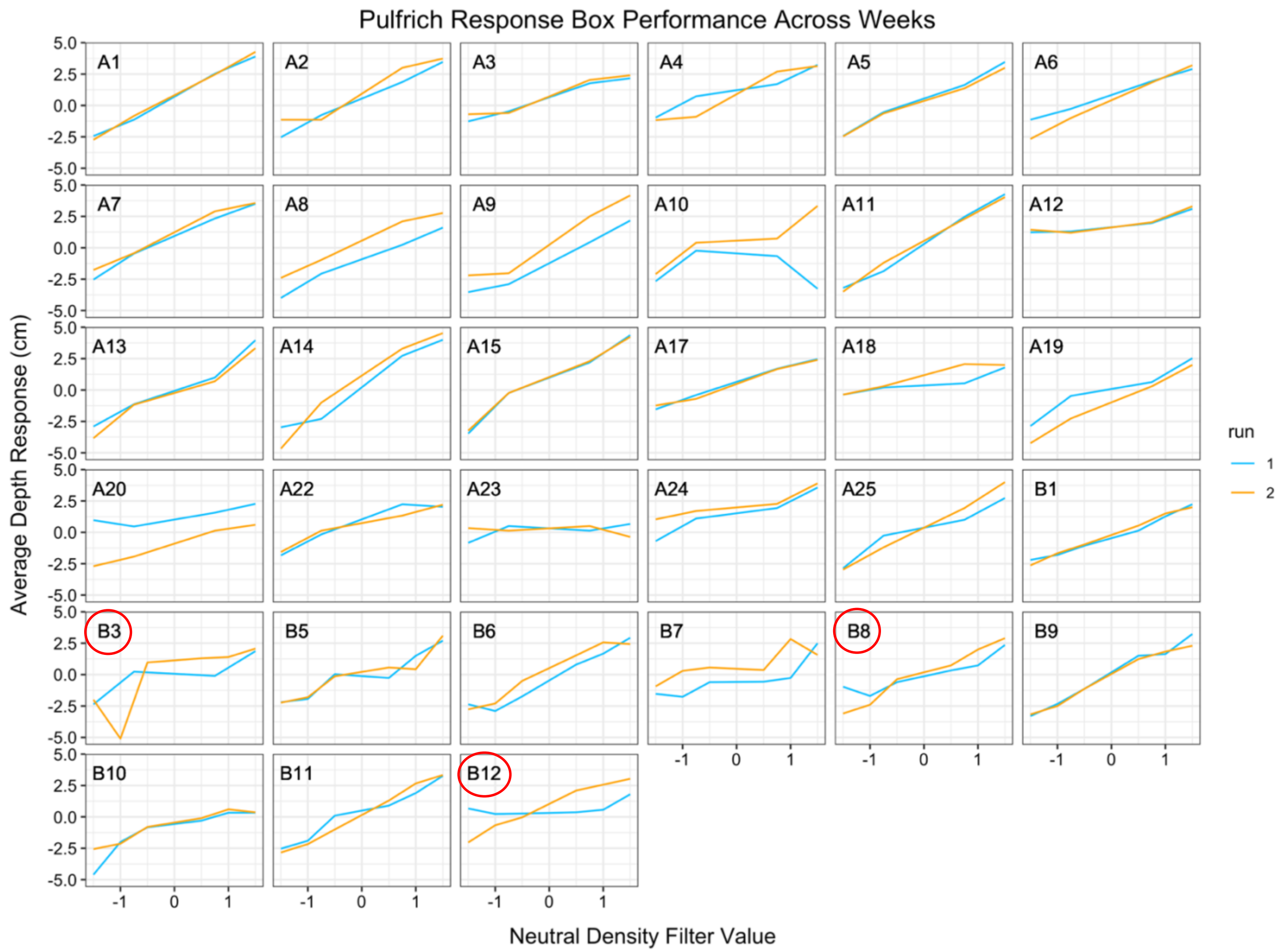


Figure E7 Individual data plots for the Pulfrich Response Box task. Subject 23 was excluded from the correlation analysis due to a clear misunderstanding of the task. Subjects A8, A9, A19, and A20 used two different version 1 response boxes between week 1 and week 2. Subjects A22-25 and B1-B12 used version 2 of the response boxes.

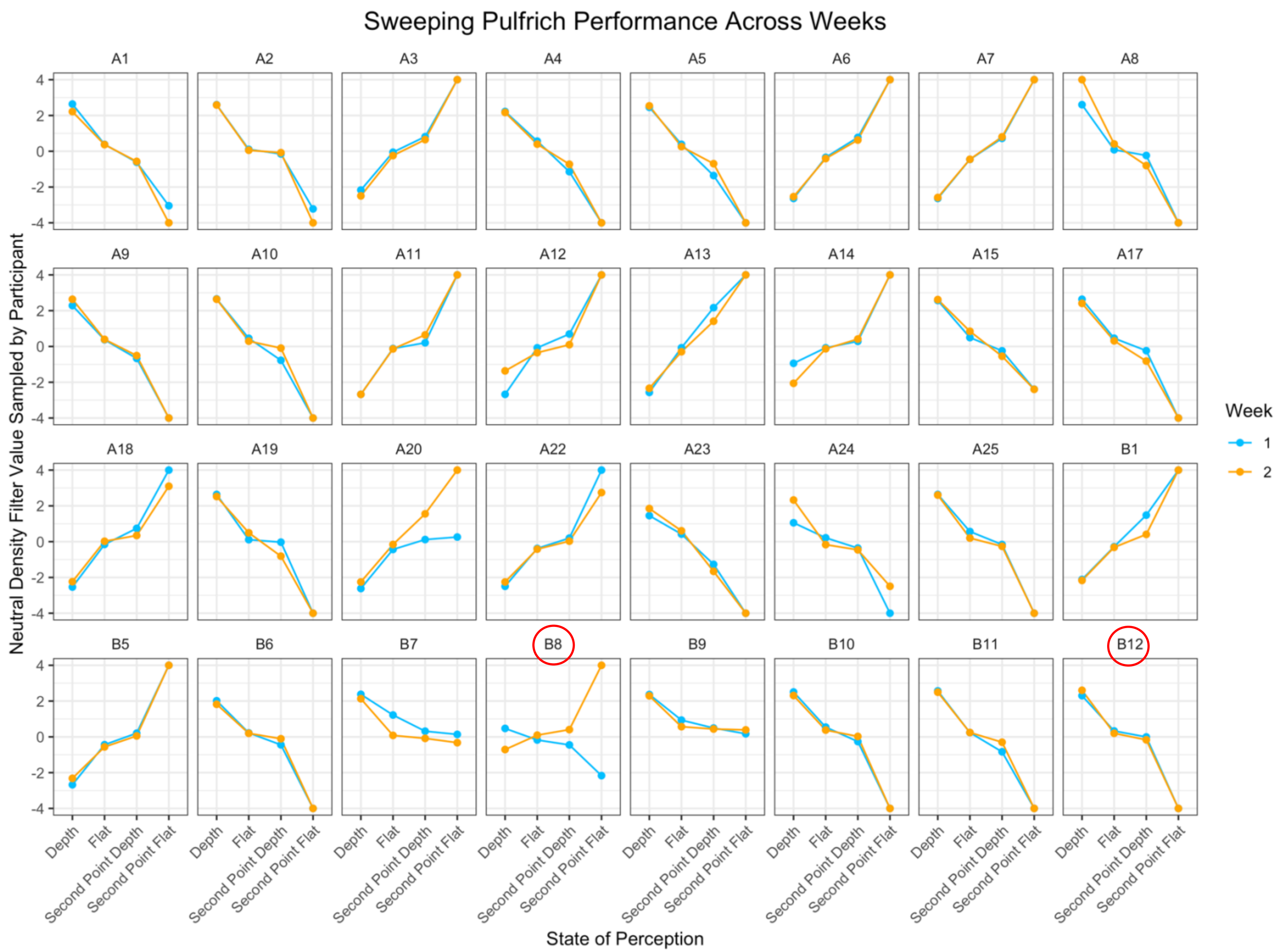


Figure E8 Individual data plots for the Sweeping Pulfrich task.

Appendix F: Motion Parallax Other Correlation Plots

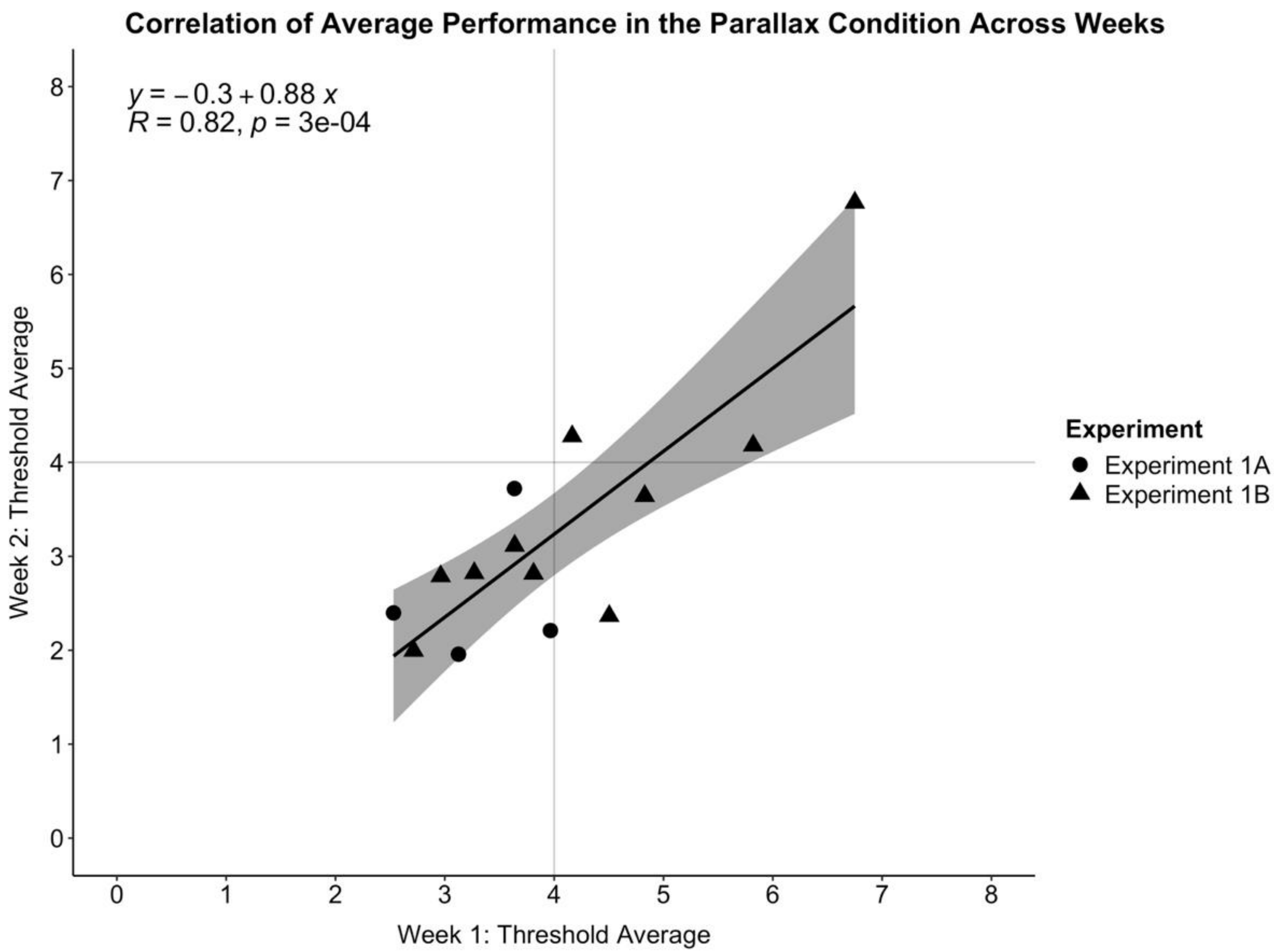


Figure F1 Correlation plot for the Motion Parallax task calculated with parallax trials only.

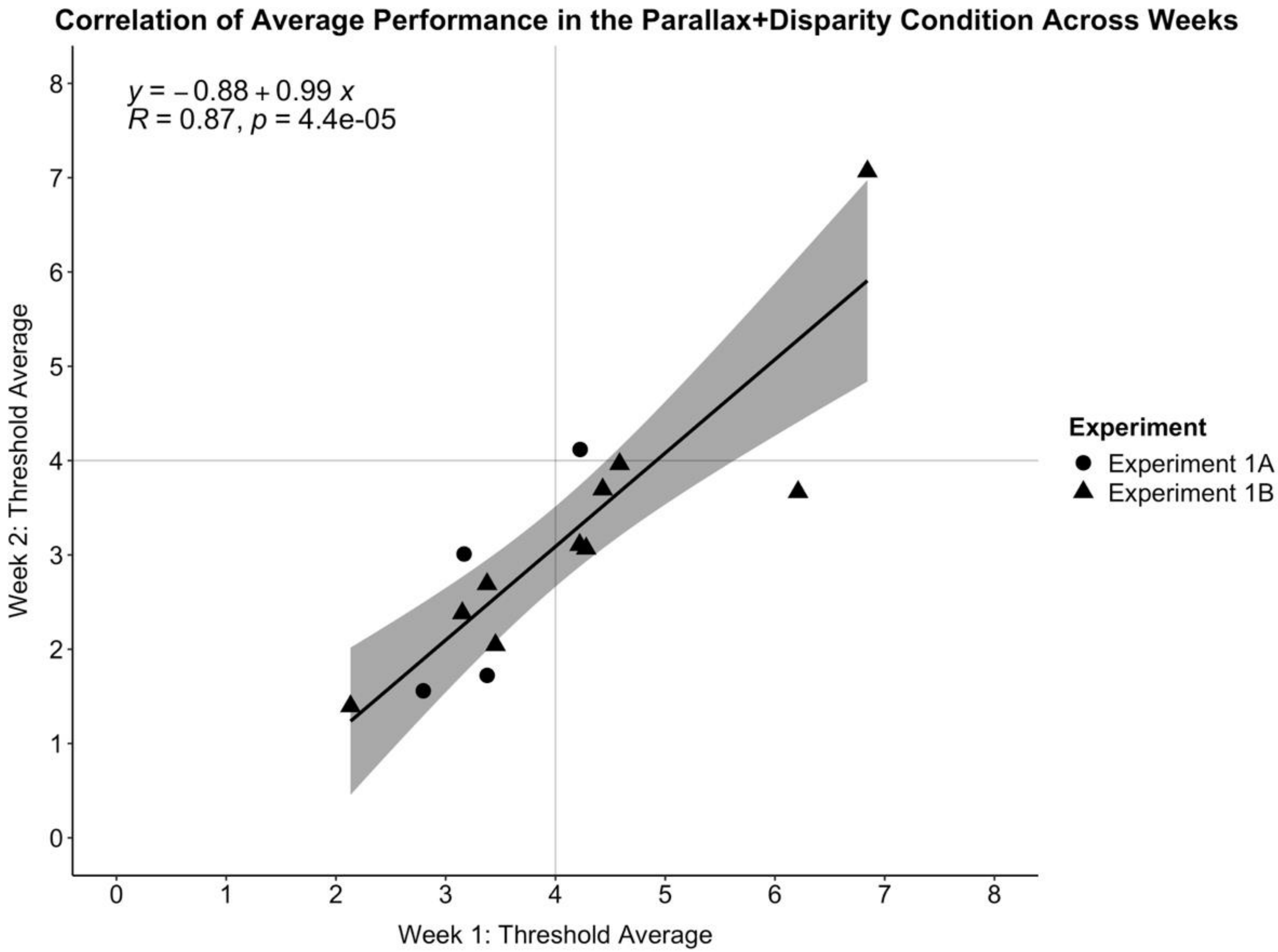


Figure F2 Correlation plot for the Motion Parallax task calculated with parallax+disparity trials only.