DOES MOTION PARALLAX IMPROVE COMMUNICATION EFFICIENCY IN VIDEO CHAT?

ASHLEY FUNKHOUSER

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

APRIL 2025

Abstract

As video communication has become more prevalent in our day-to-day lives, it becomes evident that face-to-face communication vastly outclasses video chat in terms of peer communication. Motion parallax is a perceptual effect that arises when an observer moves relative to their surroundings, or their surroundings move relative to them, causing nearby objects in their visual field to appear to move more quickly than distant objects. This relative motion provides a depth cue that the brain can use to estimate the relative distances and orientations of the objects. Directionality is a mutual understanding of the distance and orientation of people in 3D space (Troje, 2023). Previous studies have found that motion parallax is important in determining the direction of objects (Wang & Troje, 2023). Motion parallax can help provide directionality in day-to-day life, including aiding with nonverbal cues such as pointing or turning one's head.

This study examines whether adding motion parallax to video chat with avatars can enhance communication efficiency, as indicated by performance on an instruction task. Many nonverbal cues like mutual gaze, pointing, and eye contact rely on directionality to function accurately. Video chat can create misleading cues due to the lack of motion parallax, causing misunderstandings (Troje, 2023). This study found that the use of motion parallax while video chatting did not enhance performance on a shared task between two subjects, relative to the control. Further research is required to clarify the relationship between motion parallax and communication efficiency in video chat with avatars.

Acknowledgment

These past four years have been some of my life's most challenging and rewarding. While I expected academic challenges, I also faced personal difficulties. I am a very different person now and am relieved to close this chapter and open new doors. I wouldn't change a single decision – mistakes and all ,I've learned invaluable lessons. I couldn't have made it through without a village of people.

To my dearest wife, Lauren Lagarde: This journey has tested and strengthened our relationship. I'm deeply grateful for your unwavering support through my highs and lows.

To Dr. Suzanne Moeller, Dr. Micheal Wang, and my other academic friends, thank you for your near-unlimited patience with me. Your willingness to help me with a bombardment of questions is beyond appreciated.

To my therapist, thank you for helping me untangle my experiences, forgive myself for my mistakes, learn from my mistakes, and overcome my insecurities and fears.

To Dr. Hajnal, who was my first advisor and who still checked in on me well after graduation from USM, thank you for your confidence and support in me, even when I did not have it for myself.

To my in-laws. Angelina Rushing and Lenny Lloyd, I couldn't ask for better in-laws, and I appreciate your cheerleading for me.

To my parents, Christa and Michael Funkhouser, your unconditional love and support have meant the world to me. I wouldn't be the woman I am today without you.

To Niko, thank you for taking a chance on me. I appreciate everything you've done to fight for me to get through this. It has been a long and arduous task for both of us, but we pulled through. For that, I am thankful.

Table of Contents

List of Figures

List of Tables

Introduction

Video conferencing services such as Skype, Zoom, and Microsoft Teams have become popular during the COVID-19 pandemic to avoid physical contact and to permit colleagues in remote locations to collaborate face-to-face. Despite their convenience, these platforms pose unique challenges. A notable one is “Zoom Fatigue,” a term used to describe the mental exhaustion linked with prolonged video calls (Bailenson, 2021; Webb, 2021). Researchers have suggested that issues with nonverbal cues and eye contact are key contributors, impacting trust, attention, and overall communication efficiency (Bailenson, 2021; Bulu, 2012; Webb, 2021).

These communication breakdowns may be linked to deeper perceptual issues caused by the visual limitations of video conferencing, particularly the loss of *motion parallax* and *deictic consistenc*y, which may affect the *presence* of video chat. *Motion parallax* occurs when an observer or their surroundings move. This causes nearby objects to undergo more retinal movement than objects farther away, allowing the brain to estimate objects' relative distances and orientations. *Directionality* (also called deictic consistency) is the faithful rendering of each participant’s viewpoint and relative orientation in a shared environment. In other words, it ensures that gaze, pointing, and body-turn cues map onto the same 3-D coordinates for both parties. Together, these cues support the experience of presence, the feeling of truly “being there” in a space.

When this consistency is lost, as is often the case in video chat, mutual gaze and spatial references become unreliable, increasing cognitive load and disrupting the fluidity of interaction. This then breaks the overall presence of video chat. Nguyen and Canny (2005) describe this quality as “spatial faithfulness,” emphasizing the importance of maintaining naturalistic spatial cues in mediated communication.

As a result, video communication can sometimes hinder rather than help. Compared to face-to-face communication, and even to strictly audio communication in some situations, it can be less efficient (Hinds, 1999). It can hinder communication efficiency in the areas of the time taken to complete a task, and requires more verbal back-and-forth to establish mutual understanding (A. H. Anderson et al., 1996; Boyle et al., 1994; Hancock & Dunham, 2001; O'Malley et al., 1996). Misleading directionality in video chat may be a contributing factor, which creates a disconnect in the reciprocity needed for mutual gaze and demands more mental effort to convey meaning and pay attention (Hinds, 1999).

Visual Space Versus Pictorial Space

To understand the limitations of video communication, it's crucial to distinguish between two types of perceptual environments: visual space and pictorial space. Video chat distorts mutual gaze and other visual cues because it exists in pictorial space. A pictorial space is defined as the space depicted in a 2D image or on a computer screen. Pictorial space emerges from perceiving a depicted three-dimensional (3D) scene on a flat surface such as a canvas or a screen (Koenderink & van Doorn, 2012). For example, a computer and its physical screen are in visual space, in a manner analogous to a painting canvas and the physical paint on top of it. The scene playing on the computer screen or the picture on the painting canvas are considered pictorial spaces (Figure 1).

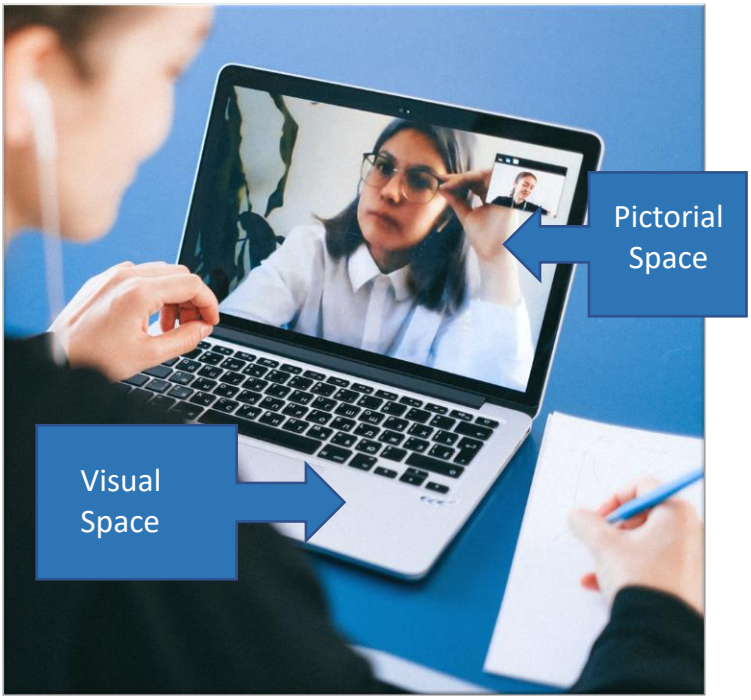

Figure 1

Visual space vs pictorial space

The physical laptop is in physical space while the image of the woman inside is in pictorial space.

Face-to-face communication is in a visual space and has faithful mutual gaze. Visual space refers to the space that is directly visible to the observer (Koenderink & van Doorn, 2008, 2012). In face-to-face conversation, un-distorted depth cues provide observers with information about the spatial and directional relationships between themselves and the objects around them. Studies suggest directional information can help establish rapport and increase communication efficiency (Brennan, 1998; Doherty-Sneddon et al., 1997; Kraut et al., 2003).

Two prominent visual cues for depth perception are motion parallax and binocular disparity. Binocular disparity is defined by differences in the retinal images of an object viewed by the two eyes from slightly different angles. This allows the brain to create a 3D visual space perception by joining the different retinal images from the right and left eye. Motion parallax

refers to the phenomenon where objects closer to an observer go through more retinal movement than those that are further away when the observer moves or changes position.

While motion parallax provides directionally and spatially accurate information about the observer's surroundings, the contribution of binocular disparity to spatial and directional information is more limited. Specifically, binocular disparity provides information about relative depth (e.g., distance between objects), but not egocentric distance (i.e., distance from the observer to an object). In contrast, motion parallax supports more robust spatial and directional judgments. Other binocular cues, such as vergence and vertical disparity, contribute more to absolute distance perception. Ongoing research has shown that motion parallax is particularly important for determining orientation and directionality, whereas binocular disparity is more important for relative depth. Stereopsis is important for distance (Wang et al., 2020; Wang & Troje, 2023).

Importantly, in natural viewing conditions, depth cues like motion parallax and binocular disparity are not processed in isolation. They typically operate in synchrony, integrating to support a cohesive perception of spatial layout and motion. However, in some experimental or mediated contexts, such as screen-based presentations, these cues may become unbalanced, absent, or even contradictory. It is often the conflict or mismatch between available cues, rather than the absence of one alone, that degrades spatial judgments. When this happens, conflicts between depth cues can degrade spatial judgments, potentially impacting how observers interpret the position or motion of on-screen elements.

Directionality and Mutual Gaze

Building on this, it's important to note that some depth cues behave differently in pictorial space than in visual space. The visual information in a visual space constantly updates to reflect

the spatial relationship between a person and their surroundings. For pictorial space, the content displayed on the image plane remains constant regardless of the observer's position (Koenderink & van Doorn, 2008, 2012). For example, in video chat, the vantage point is the camera on the computer. When the user moves in front of the computer, the view of the other person's face does not change, as the camera location of the computer on the other end of the video line does not update accordingly (Figure 2). This can cause distortions of directional cues like gaze and pointing, creating illusions.

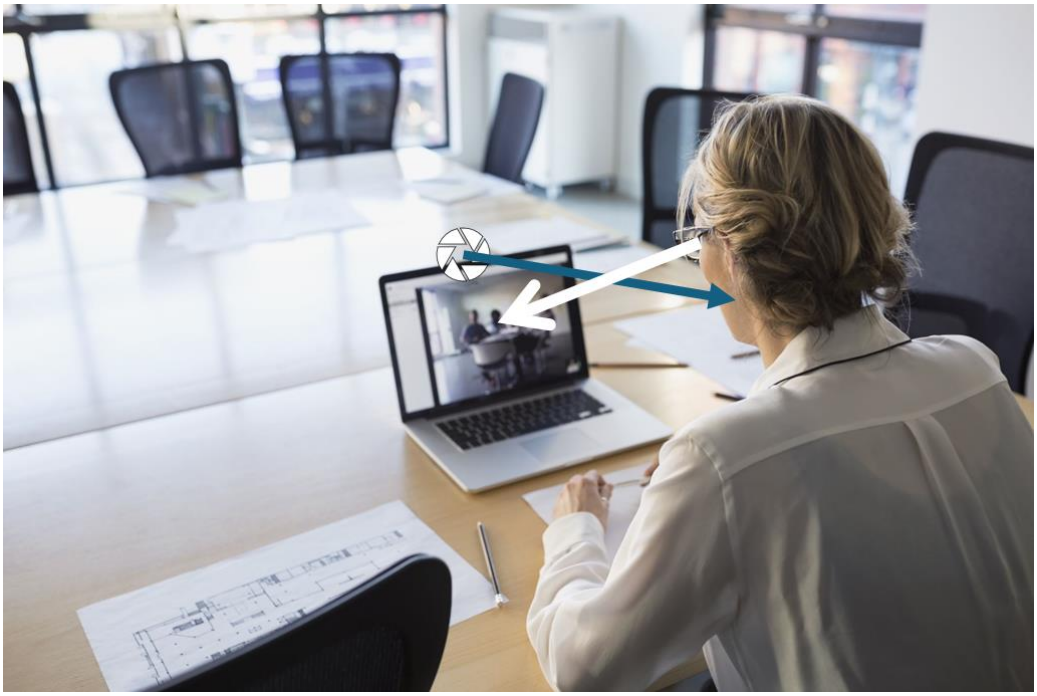

Figure 2

The mismatched vantage point of video chat

The camera's vantage point (blue arrow) versus the person's vantage point (white arrow) is not aligned.

One such example of the distortions of depth cues in pictorial space is the famous Uncle Sam poster, where his finger always points at the observer regardless of where they are in the room. This phenomenon is known as the "Mona Lisa effect," where an object in pictorial space

appears to be oriented toward the observer regardless of the observer's position (Hecht et al., 2014; Maruyama et al., 1985; S. Rogers et al., 2003).

The Mona Lisa and the poster of Uncle Sam are not unique in having this effect, nor is it exclusive to pictures. It can also occur in video chats and other media. In a video chat, if the person on the screen looks straight ahead (into the camera) and the viewer looks at the screen from an oblique angle, the viewer will perceive the person as looking directly at them, this is the Mona Lisa effect in action. However, mutual gaze (where both parties feel they are making eye contact) is more challenging to achieve. If the viewer looks directly at the screen (rather than the camera), the person on the screen will not perceive eye contact, even if they are looking straight into the camera. This makes conveying mutual gaze difficult in video conferencing. In face-to-face communication, the directionality is not distorted, thus allowing for easier and more natural mutual gaze (Figure 3).

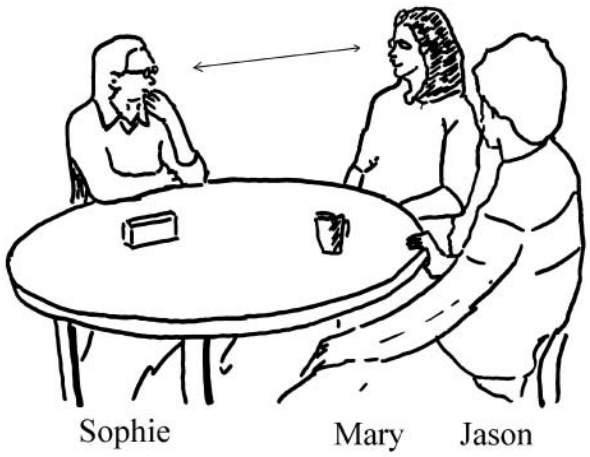

Figure 3

Mutual gaze and directionality in face-to-face communication

Sophie and Mary are sharing mutual gaze. Sophie, Mary, and Jason can tell where everyone is looking at and the orientation of their heads. This is directionality. The directionality affords Sophie and Mary's mutual gaze. Graphic from Haddington (2006)

The pictorial nature of video chat may partially cause "Zoom Fatigue," where users feel unusually exhausted after communicating via video conferencing services (Chen et al., 2021; Troje, 2023). The issue of "Zoom Fatigue" is a new phenomenon, and the causes of such issues are multifaceted. A distortion of directionality, and by extension, a distortion of mutual gaze, may be responsible for some of the known issues that video chat causes for the frequent user (Troje, 2023).

We predict that providing motion parallax in video communication will facilitate more efficient communication. To determine whether providing the user with faithful directionality in video calls can affect communication, we have created a new video communication system that provides motion parallax. The program contains a facial detection algorithm that tracks head location to provide different visual perspectives in video conferencing, thus providing better directionality and alleviating some distortions from regular video conferencing.

Overall, this study aims to investigate whether the introduction of motion parallax can improve communication efficiency and mitigate common issues associated with shared understanding in video conferencing.

Mutual Gaze and Communication

Gaze is the direction or focus of a person's eyes. Mutual gaze occurs when two people make eye contact or look into each other’s eyes (S. J. Rogers, 2013), which is vital for indicating attention and building rapport and intimacy between persons (Argyle & Cook, 1976). Current video communication provides a distorted version of mutual gaze and other directional cues, creating slower response times and the need for more words to reach a shared understanding (Mukawa et al., 2005; Short et al., 1976).

Studies have shown that people tend to have different nonverbal cues in video conferencing, possibly to make up for the differences in how visual cues behave. These differences in nonverbal cues could lead to higher cognitive load, as the standard nonverbal shortcuts are different than what is normal in face-to-face communication. (Croes et al., 2019; Hinds, 1999; Kleinke, 1986). Dalzel-Job (2015) found that mutual gaze was positively correlated with task performance in mediated communication settings. In experiments using avatars in a virtual environment, mutual gaze was most beneficial when implemented naturally, rather than as continuous staring, and was associated with higher social presence and faster task completion. These findings underscore that mutual gaze not only influences social perception but also has a measurable impact on communication efficiency. Importantly, the study also highlighted those behavioral measures, such as response times and task success, were more sensitive indicators of social presence than subjective reports alone.

The current study hypothesizes that mutual gaze would improve if video chat conveyed directionality through motion parallax, thereby increasing efficiency during video communications.

Testing Communication Efficiency

Communication efficiency and cooperation are often tested by having participants perform shared tasks, such as a map task, a maze task, a puzzle task, or a building task (A. Anderson & Garrod, 1987; A. H. Anderson et al., 1996; Garrod & Anderson, 1987; Gergle et al., 2004; Hancock & Dunham, 2001). The basic formula for this type of study is that participants must work together to achieve an end goal or product that requires mutual attention and cooperation. The efficiency metrics that demonstrate the success of the shared task are the

accuracy of the end product versus the target product, the required completion time, and the number of exchanges and interruptions in verbal communication.

Instruction Tasks

Following instructions or cooperating on a task is fundamental in communication. Specific tasks aim to replicate real-life situations where individuals work together to achieve a goal. The focus of these tasks is to evaluate how well participants perform, usually in terms of accuracy and speed, as well as their communication abilities, which may involve analyzing speech patterns, interruptions, and vocabulary. These tasks often require an instructor and an instruction recipient/student, but they can also be cooperative.

Perhaps the most popular of the instruction tasks is the map task. In the map task, two participants work together to find and complete a map route as quickly and accurately as possible (A. H. Anderson et al., 1991). This task was popular among researchers and used in text-based, audio-based, video-based, and face-to-face conditions (A. H. Anderson et al., 1984, 1996; Boyle et al., 1994; O'Malley et al., 1996). The two players were separated unless they were in a face-to-face condition: one must give instructions, and the other must try to draw a route on the map. Route efficiency, communication speed, accuracy of the endpoint, and interruptions in communication (repeating or clarifying instructions) can be efficiency and cooperation measures. Boyle et al. (1994) found that video-based communication was quicker and had fewer interruptions than audio-only conditions. Anderson and colleagues (1996) found that video-based communication was slower and had more interruptions than face-to-face communication.

In contrast, O'Malley et al. (1996) performed a similar experiment to Boyle et al. (1994) but added eye tracking. The conditions were a face-to-face condition, an audio-only condition, and a video chat condition. The group found no significant differences between speed and

accuracy between conditions. However, they found that audio-based communication had far fewer interruptions than video chat. It was noted that participants gazed more at their partner in the video chat condition than in other conditions, and there were significantly more communication breakdowns, such as interrupting the other participant and asking for clarification, as well as more verbal affirmations such as "mhm" in the video chat condition. O'Malley et al. suggests this effect could be related to the novelty of the medium at the time, which could cause higher cognitive load and more mistakes. Video chat could have been less effective at communicating nonverbal cues such as nodding or gazing (O'Malley et al., 1996).

A variation of this task created by the same group is called the "Travel Game," where a participant must work with a "Travel Agent" to create an itinerary across the United States that visits as many cities as possible with minimal changing of airline and backtracking to a previous city. The "Travel Agent" was a confederate with a loose script. Participants used a video conferencing tool with an updating map and travel times, an audio-only tool with a shared screen map and a log of travel decisions, or face-to-face with pencil and paper. Similarly to the map task, the path efficiency, time to the outcome, and the number of verbal interruptions were considered as metrics (A. H. Anderson et al., 1996). The results obtained were similar to those of O'Malley et al. (1996), where video chat did not perform as well as face-to-face interactions. Anderson et al. proposed that subtleties of face-to-face communication were crucial to these shared tasks (A. H. Anderson et al., 1996). Faithful directionality and its subsequent presence could be a significant factor in these results.

Building tasks require people to assemble something mechanical together. These tasks consider how accurately and how quickly the team can build together as a measure of "understanding" and "communication effectiveness." Sometimes participants are given manuals

on achieving the target, while at other times, they are only given the end goal and must produce the instructions themselves (Fussell et al., 2004; Hancock & Dunham, 2001). In an example, one unskilled worker and one expert in the task required two people to repair a bike. In some conditions, the unskilled worker worked alone with a repair manual or had access to an expert through a pure audio connection or a video link. The performance of the novice was improved when an expert watched the novice via video link (Kraut et al., 1996, 2003).

Another study required two participants to build a robot, where one participant had instructions to help the other worker assemble the robot. The study observed the effect of pointing in communication, and various remote pointing methods were observed, such as a cursor or drawing gestures. While the cursor did not seem to improve build speed, drawing gestures did help improve the remote building of the robot (Fussell et al., 2004). Other building tasks are similar, such as a Lego design (Clark & Krych, 2004). They found in that study that participants performed better when a workspace was shared versus an unshared workspace between users or an audio link.

In a similar cooperation task where the "worker" had a puzzle to complete and the "helper" had an answer key, Gergle et al. found that the ability of the helper to see the worker's workspace facilitated faster and more accurate completion of the puzzle. Both conditions had participants perform the task on a computer via an audio link. The condition where the workspace was visible resulted in fewer words needed overall and less back-and-forth between participants, compared to the condition where the helper could not see the worker's screen. Gergle et al. suggested that seeing the workspace can substitute for nonverbal behaviours like pointing or eye gaze, which are directional cues (Gergle et al., 2004).

Clark and Wilkes (1986) and Handcock and Dunham used tangrams. Participants had to collaborate to match tangram figures. One participant had twelve tangram figures to describe to the other participant. The other participant had to identify which of the twelve tangram figures the other was describing from an array of sixteen tangrams. Clark and Wilkes's study compared the verbal conversations between participants to a proposed conversational model where the speaker references the object or process, and the listener then confirms or clarifies. They concluded that their model matched transcripts of participants working together to complete the task (Clark & Wilkes-Gibbs, 1986). Handcock and Dunham's study observed how two participants' personalities meshed between text-based computer-mediated communication and face-to-face conversations. No performance (accuracy) difference was found between the conditions. Face to face was quicker than computer-mediated communication (Hancock & Dunham, 2001). While they did not directly comment on why they thought speed was faster in the face-to-face, the ability to read better nonverbal cues may have aided in speeding up the task.

Challenges with Instruction Tasks

Although many scenarios presented in instructional tasks could have practical applications, they often lack diversity and are not repeated frequently enough. This creates a shortage of replicated paradigms, despite the vast possibilities and room for creativity. Nevertheless, these tasks can help simulate real-world tasks in different environments.

Before selecting an instructional task, one should consider whether to choose a commonly repeated task, a task with limited replication, or an original task. Determining if the task suits the medium used to communicate the instructions is also essential. For example, physical props may not be appropriate for strictly computer-based environments. It is also essential to consider the task's difficulty level and whether it is appropriate for novice users.

Building a bike is likely to be challenging for someone without prior experience, but almost anyone can complete a Lego set or a jigsaw puzzle with proper instructions.

Building a Task to Test Motion Parallax's Effect on Communication Efficiency

Several factors were considered when deciding on the paradigm for testing communication efficiency for video communication with motion parallax. At the time of experimental planning, due to the COVID-19 pandemic and the nature of studying video chat, the ability to perform the task remotely was paramount. We also needed an interesting task to engage participants and encourage others to sign up for additional data points. Puzzles were selected to test communication efficiency, using a methodology akin to Gerle and colleagues' study. However, we selected tangrams like Handcock and Dunham's (2001) for the experiment instead of jigsaw puzzles, as tangrams are a more complex task and require participants to work together more closely (Clark and Wilkes, 1996).

Hypothesis

In the current study, we hypothesize that participants in the motion parallax condition will perform the tangram task faster, and there will be a smaller mean Euclidean distance from the target and participants' solutions.

Methods

Participants

56 adults participated in the present experiment, equaling 28 student-instructor pairs (age mean = 23.76; SD = 6.09; 20 males, 33 females, 3 participants preferred not to answer). Participants were undergraduate and graduate students from York University who participated for course credit or monetary compensation. Participants were recruited via the psychology student participant pool on campus and social media. All participants had normal or corrected-to-

normal vision and spoke proficient English for the experiment. Pairs were not controlled for gender or familiarity with each other, as participant signups were sparse due to COVID-19-related restrictions. The York University Research Ethics Board provided ethics approval for this research (REB # 2021-387). All participants gave informed consent before beginning the experiment via a consent form and a verbal instruction period.

Stimuli

Tangrams

Eight tangram puzzles were utilized (Figure 4**Error! Reference source not found.**). These tangram shapes were located on various teaching websites and were considered easy-level puzzles. (Tangram Channel, 2023)

The tangram task contained seven basic tangram pieces consisting of two large triangles, two small triangles, one medium triangle, a square, and a parallelogram, as shown in Figure 5. All game pieces were white to prevent colour references and encourage additional participant interactions (e.g., describing the shape instead of the colour). Tangram target figures for the student were created using seven black target pieces to obfuscate which shape had been used in what location. Instructor figures were created using black-and-white target pieces for the instructor to use to describe the targets to the students. The shapes of the target and game pieces were the same.

The tangram playing field was ~22 Unity Units (UU) wide and ~16UU tall. The puzzles were around 9UU tall and wide on average Unity units are Unity's in-game measuring system, and one UU is one real-world meter.

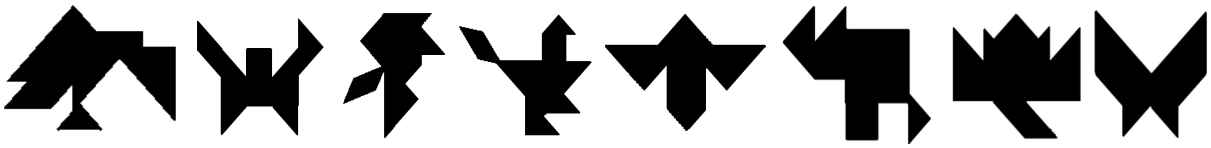

Figure 4

Tangram figures used during trials

Tangram figures used during trials. The puzzles the pairs had to complete during the experiment. From left to right, starting from the top: Puzzle 1: Crow, Puzzle 2: Crab, Puzzle 3: Dinosaur, Puzzle 4: Chicken, Puzzle 5: Bee, Puzzle 6: Rhino, Puzzle 7: Lotus, and Puzzle 8: Butterfly.

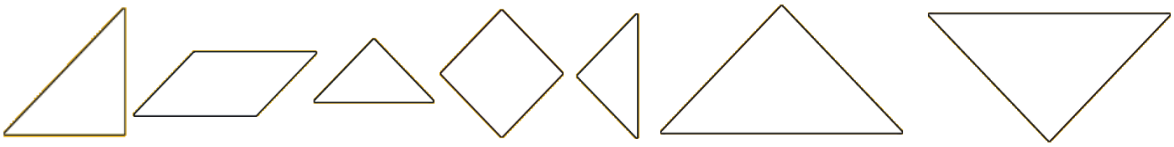

Figure 5

Tangram pieces

Tangram pieces used to create the figures. From left to right: right triangle, parallelogram, small triangle 1, square, small triangle 2, large triangle 1, large triangle 2. These pieces were provided by the Tangrams game from the Unity Assets store (VR Cardboard Buddies, 2016).

Avatars

Participants chose one of two premade face avatars: one that appeared more feminine, and the second that appeared more masculine (Figure 6). Avatars were controlled by the participant's head movements and facial expressions such that the avatar followed the participant's motions. The resulting projection was displayed on the partner's monitor. Simulated motion parallax was either on or off.

Figure 6

Face avatars.

Left: The feminine avatar, Right: The masculine avatar. The avatar was chosen by the participant to best represent them and was controlled by their head and facial movements.

Apparatus

A proprietary application named "MPDepth" was created in the laboratory to simulate motion parallax for screen-based objects and environments to evaluate the effect of motion parallax on the efficiency of shared tasks. The app was implemented using Unity and ARKit.

Two identical setups were utilized for the experiment. Each setup consisted of a large screen to display the stimuli, such as the tangrams, avatar, and workspace, and a sensor mounted on top of the screen. The sensor was an iPhone 12 mini with a "TrueDepth" RGBD camera, placed on an ASUS ProArt Display 27" monitor. The monitor and sensor apparatus were then vertically mounted onto a height-adjustable tripod (Figure 7, left). Participants were directed to stand approximately 75cm away directly in front of the keyboard and mouse and were instructed not to move far from the space (Figure 7, right). Participants were asked to stand as studies have

found that people sway and move more while standing (Fujimoto & Ashida, 2020). The underlying logic is that increased movement would provide greater opportunities for the system to be utilized. The monitor was positioned such that the avatar's eyes were parallel to the participant's eye height. The avatar's head displayed ~19.05cm on the monitor, around the height of the average human head.

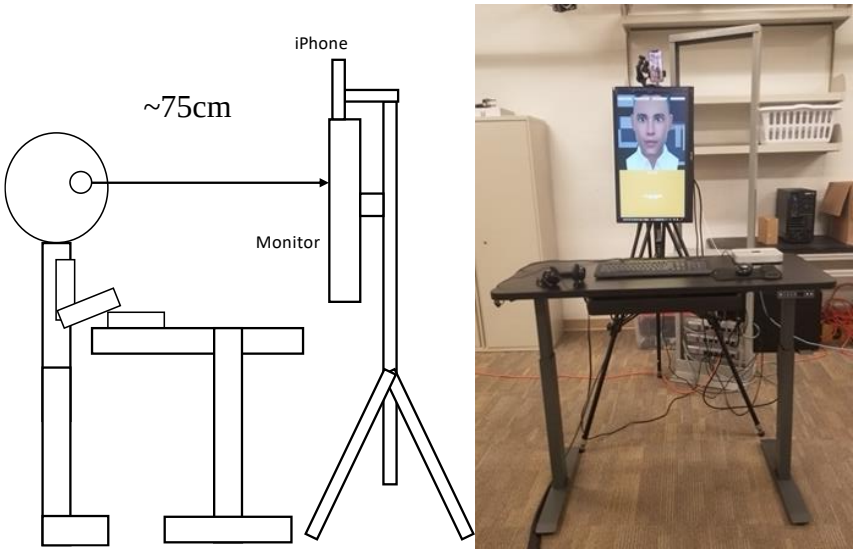

Figure 7

The experimental apparatus

Left: Diagram of the participant and monitor / sensor apparatus. Participants were directed to stand approximately 75cm in front of the monitor. They were instructed not to move far from the keyboard but were allowed to move naturally as they spoke. Right: The participant setup. The monitor is an ASUS ProArt Display 27" with an iPhone "TrueDepth" RGBD camera mounted on top using a tripod. The desk is an adjustable height desk.

A standing desk with a mouse and standard QWERTY keyboard and a headset with a microphone were provided. The mounts were locked into place and calibrated using the Optitrack Motion Capture system. Calibration was performed to determine where the iPhone, and by extension the camera and sensor, was relative to the computer monitor. This would ensure that the parallax effect would be accurate to the subject's movements. The game was

created with an offshoot of the MPDepth app, adding a tangram task modified from the Unity asset store (VR Cardboard Buddies, 2016)

MPDepth

The MPDepth App is an app created in-house to produce motion parallax in video chat. MPDepth was created in Unity and uses the Apple TrueDepth camera and ARKit to track the user's head movements. MPDepth Streamer, which is an app on the phone that connects to the desktop version of MPDepth, creates coordinates to stream to the desktop MPDepth application. These coordinates are then translated to coordinates Unity understands, and a virtual camera from within the app follows those coordinates in the space (Figure 8). If a person translates their head to the left to look at the head and environment from a different angle on the screen, the camera inside Unity shifts left by the same amount within the virtual space, rendering a different perspective of the head and environment on the projection plane as if the screen were a window into the 3D scene. The projection plane is a surface in which a 3D object is projected onto a 2D plane (Figure 8). This creates a dynamic vantage point inside the game playing in Unity that is projected on the screen that mimics the motion parallax found in real space.

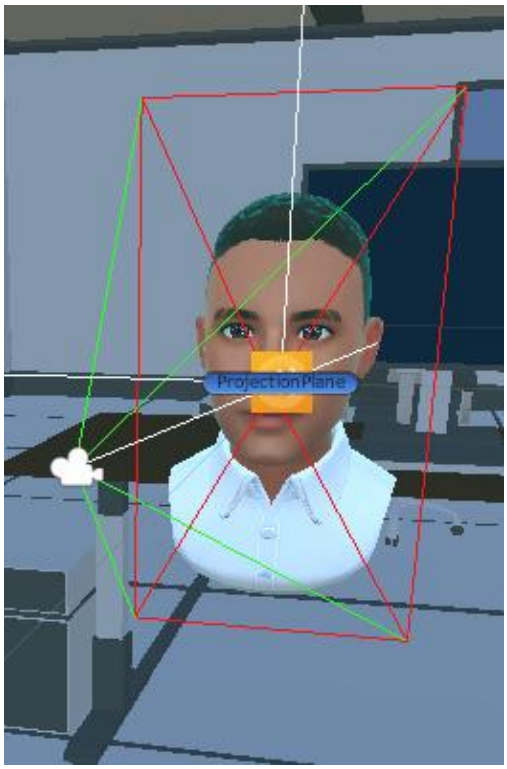

Figure 8

The virtual camera

The virtual camera inside of the MPDepth app showing what would be displayed on the projection plane.

The version of MPDepth used for this experiment has a video chat service with avatars. It adds functionality to allow two players to use the

service as one would use video communication, adding an audio link and ARKit to track facial expressions and mouth movements. Instead of translating around a static object, the virtual camera translates around the other person's avatar, which is driven by their head and facial movements. The participant can see the other participant move their head and face accurately to perceive how they are moving in real space. (Figure 9).

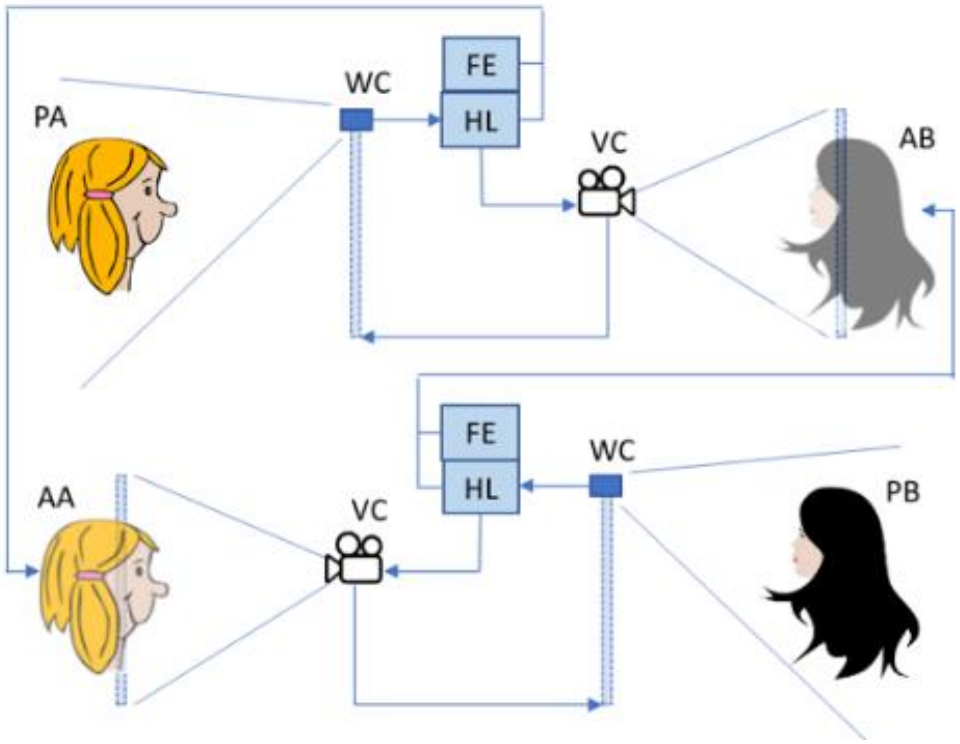

Figure 9

A diagram of the program used to add motion parallax to video conferencing.

PA: person A, PB: person B, AA: avatar A, AB: avatar B WC: webcam, HL: head location, FE: facial expression, VC: virtual camera. The webcam records and tracks the head location and facial expressions of person A. The head location of person A moves a camera in MPDepth2 around an avatar of Person B, creating a vantage point that coincides with Person A's center of projection. Person B's avatar's facial expressions and movements are driven by Person B's actual head movements and facial expressions, which are tracked by their webcam. This effect is symmetrical for both players. Figure provided by Nikolaus Troje.

Tangrams

Participants were required to click and drag the white pieces (game pieces) over the black silhouette target pieces (target pieces), using the D key to rotate a piece clockwise and the A key to rotate a piece counterclockwise. The goal was to move the white game pieces over the target

pieces to make them identical to the target puzzle shape. Users use the mouse to move the pieces around the play area.

Upon completion of a puzzle, a script first checked each type of piece in the puzzle and found the closest of those game pieces that matched the target pieces. The script then compared the distance between the target piece's position and the game piece's position using the vector distance function called Vector3.Distance(a,b) built into Unity. In Unity, the Vector3 class represents a point or direction in 3D space, containing coordinates for position and rotation.

The script retrieves the Vector3 positions of both the solution's puzzle pieces and the player's pieces and calculates the straight-line Euclidean distance between the individual pairs. The formula used was $d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$.

For cases where there are two of the same pieces (two large and small triangles), the game searched for the closer of the two to the solution piece. The Euclidean distance from each pair of pieces was then added up and divided by seven, the number of tangram game and target pieces, to derive the mean distances across all puzzle pieces for the individual puzzle in UU.

In the case of perfect placement of all seven pieces, the Euclidean Distance would equal 0UU. When the pieces are not aligned perfectly, the mean Euclidean Distance of would be greater than 0UU.

Procedure

Upon entering the lab, participants were asked to sign a consent form and fill out demographic forms. After completing the paperwork, participants were taken to one of two rooms with a setup, and the monitor's height was adjusted to be at eye level. Participants were then given instructions on the controls and how to complete the tangram task together. Participants were allowed to ask questions to clarify controls, and the instructions were finished.

Once all questions and concerns were addressed, each participant was instructed to complete a tutorial before going on with the experiment. The tutorial consisted of one puzzle with the solution showing so that the participant could get used to the controls and the motion parallax. The tutorial was always in the parallax on condition (Figure 10).

Figure 10

The tutorial

The tutorial puzzle was given to participants. It contained the controls and a brief overview of the instructions to remind them of the task.

The tangram task required two participants to finish eight tangram puzzles. The screen was split, with the task on the bottom third of the screen and the top two-thirds displaying the

facial avatars (Figure 11). Throughout the tangram task, participants acted either as students or as instructors. Roles were chosen using a randomly generated seed that created four trials labeled "0" and four trials labeled "1," which were then shuffled before the experiment started. "0" trials had Participant A as the instructor and Participant B as the student, while "1" trials had Participant B as the instructor and Participant A as the student. The same seed method was chosen to pick the roles to control the parallax-on and parallax-off conditions. A seed was created for each puzzle, as well as for numbers 1-8, which were then shuffled. Initially, I had intended for there to be an equal distribution of conditions for each puzzle, but this was not achieved.

Figure 11

Game screens

Pictured: screens for the instructor phase. Left: Instructor's screen. The instructor's screen contains the solution to the puzzle that the instructor must describe to the student. Center: Student Screen. The student during this time must listen to the instructor. Right: This is the student screen during the student phase. The instructor screen did not change during the student phase beyond the timer set to 0. The student would have to move the pieces over to the target silhouette to attempt to complete the puzzle.

Each trial was composed of an instructor phase and the student phase. Participants were assigned to the instructor phase during the first 45 seconds of each trial. During this phase, the participant assigned to the instructor role for that trial was given an answer key to the puzzle the student was completing (Figure 11, left). The instructor explained verbally how the various tangram pieces fit together to make that shape to the student, who was also allowed to ask questions. During the instructor phase, participants were asked to stand up straight to ensure that the middle of the video chat portion of the screen remained at eye height.

The instructor had a timer on the screen to inform them how much time they had left. The student was instructed to look at the instructor's avatar and their instructions while the game screen was white (Figure 11, center). After the first 45 seconds, the student phase began, and the participants were no longer permitted to talk. The student was instructed to complete the puzzle by dragging the white tangram pieces over the black silhouette of the puzzle as quickly as possible (Figure 11**Error! Reference source not found.**, right). Upon solving the puzzle, the student would press a check mark to end the trial and move on to the subsequent trial. The student would then tell the instructor to press the checkmark on their screen below the solution to ensure that the instructor was also ready for the next round (Figure 11, left). They would then move on to the next puzzle to complete.

The students' completion time and mean Euclidean distance were recorded for each puzzle they completed. There were no thresholds for accuracy for completion to move on to the next puzzle; however, this aspect was blinded to prevent the groups from skipping puzzles. For half of the puzzles, the face avatars behaved like traditional video conferencing (no-parallax condition). The simulated motion parallax (parallax condition) was enabled for the facial avatars for the other half of the puzzles.

Design and Data Analysis

Each dyad completed 8 trials, taking on the roles of either instructor or student (2 levels) and with or without parallax (viewing condition, 2 levels). Within each dyad, one participant was randomly assigned the role of *instructor* and the other the *student*. Only the students generated data; instructors provided guidance. Thus, all dependent variable measurements were recorded from the student role only. The order of the roles, conditions, and puzzle types was presented at random. The independent variable, viewing condition, was within-subjects.

After one dyad was discarded due to failing to follow instructions, 216 data points (27 pairs × 8 trials) were collected for each dependent variable. Using the interquartile range method of outlier removal, 52 data points were removed, leaving 164 data points.

Puzzle 1 and Puzzle 5 each have 24 instances, representing 14.6% of the total dyad trials. Puzzle 2 has 14 dyad trials (8.5%). Puzzle 3 has 22 instances of dyad trials (13.4%). Puzzle 4 has 25 instances (15.2%), Puzzle 6 has the fewest with 9 instances (5.4%), Puzzle 7 has the most with 26 instances (16%), and Puzzle 8 has 20 instances (12.2%) (see table). A total of 49.40% of the data was in the Parallax-Off Condition, and a total of 50.60% of the data was in the Parallax-On Condition.

A linear mixed model was employed to evaluate the effect of viewing conditions (parallax-on, parallax-off) effects on completion time (in seconds) and Euclidean distance (in UU). This model was chosen because the difficulty range of the puzzles and the students' abilities to complete the puzzles were more variable than initially expected. The linear mixed model (LMM) approach is more appropriate to the experiment due to the random factor of the player, as opposed to the more traditional analysis of variance (ANOVA). An LMM allows for a better way of comparing the dependent variables while excluding the noise of the random effects

and missing data. Using lme4 packages in R version 4.2.2., the following statistical models were designed (Bates et al., 2023; R Core Team, 2023; Singmann et al., 2023):

$$Completion\ Time \sim Viewing\ Condition + (1|Player) + (1|Puzzle)$$

$$Mean\ Eudelian\ Distance \sim Viewing\ Condition + (1|Player) + (1|Puzzle)$$

The fixed effects term, viewing condition, shows the effects of the within factor. The random effect term (1|Player) captures the variance in individual student performance. The random effect term (1|Puzzle) captures the variability in puzzle difficulties. The effect of skill was not directly measured, so it is intertwined with each participant's individual differences. The linear mixed model was estimated using restricted maximum likelihood (REML), which is a method that provides unbiased estimates of variance components. The nloptwrap optimizer, a robust algorithm suitable for complex models, was used to optimize the model parameters.

Analyses were conducted using the R Statistical language (version 4.2.2;(R Core Team, 2023)) on Windows >= 8 x64 (build 9200), using the packages gridExtra version 2.3 (Auguie & Antonov, 2017), lme4 version 1.1.31(Bates et al., 2023), lmerTest version 3.1.3 (Kuznetsova et al., 2017), Matrix version 1.6.0 (Bates et al., 2024), report version 0.5.7 (Makowski et al., 2024) , ggstatsplot version 0.12.2 (Patil, 2021), ggplot2 version 3.4.4 (Wickham et al., 2016), dplyr version 1.1.4 (Wickham et al., 2023), and tidyr version 1.3.0 (Wickham et al., 2024).

Results

The mean puzzle completion time was in the Parallax-Off condition(M=147.32s, SD=80.75) appeared different in the Parallax-On condition (M=142.94s, SD=73.60); however, this difference was not significant. The mean Euclidean distance appeared different in the Parallax-On condition (M=2.03 UU, SD=1.95) than in the Parallax-Off condition (M=1.97 UU,

SD=1.88), but it was not significantly significant (see **Error! Reference source not found.Error! Reference source not found.**).

Measure	Condition	Mean	Median	SD	IQR
Time (Seconds)	Parallax Off	147.32	119.75	80.75	112.79
	Parallax On	142.94	124.09	73.60	64.84
Euclidean Distance (UU)	Parallax Off	1.97	1.28	1.88	0.93
	Parallax On	2.03	1.41	1.95	1.08

Table 1:

Descriptive stats table

Summary of the descriptive statistics for puzzle completion time and Euclidean distance under two conditions: Parallax On and Parallax Off. The mean puzzle completion time was slightly higher in the Parallax-Off condition compared to the Parallax-On condition, with mean times of 147.32 seconds and 142.94 seconds, respectively. Similarly, the mean Euclidean distance was slightly higher in the Parallax-On condition (2.03 UU) compared to the Parallax-Off condition (1.97 UU). The table also includes each condition's median values, standard deviations, and interquartile ranges.

Exploratory searches in the data found no significant trend lines or groups (Figure 12).

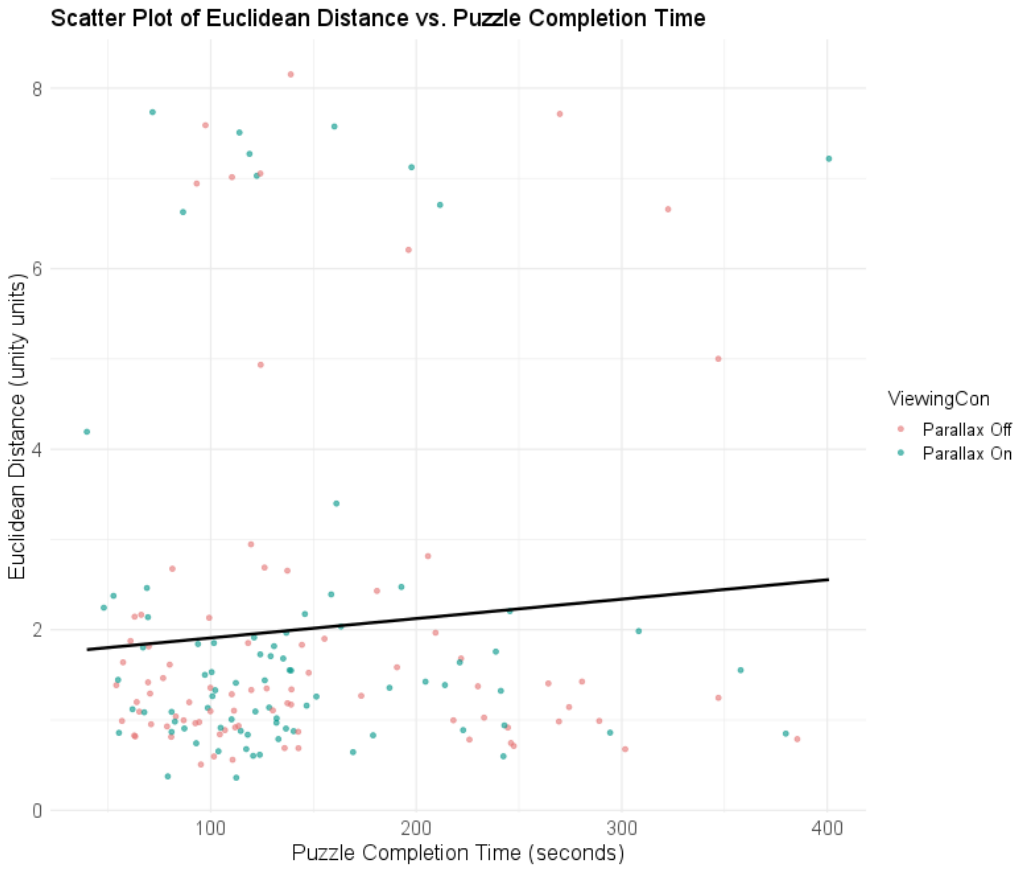

Figure 12

Scatter plot of Euclidean distance versus puzzle completion time.

Each point represents a trial. A fitted trend line was created using ggplot2's geom_smooth function with the linear smooth option for a straight line. Statistical analysis found no significant differences between conditions (Wickham et al., 2016)..

Completion Time

The model revealed that the viewing condition did not have a significant effect on puzzle completion time (Estimate = 7.57s, Std. Error = 10.14, df = 127.14, t = 0.79, p = 0.43), suggesting that the presence of motion parallax did not significantly impact how long participants took to complete puzzles. The intercept was significant (Estimate = 145.88s, Std. Error = 17.44, df = 9.78, t = 8.37, $p < 0.001$), indicating that the overall mean puzzle completion

time for the parallax-off condition is significantly different from zero, confirming that the baseline value is not trivial (Figure 13).

Random effects revealed notable variability between players (Variance = 1914, Std. Dev. = 43.75) and between puzzles (Variance = 1761, Std. Dev. = 41.96) (Figure 13), indicating that individual differences and puzzle complexity contributed substantially to completion time. The residual variance (3186, Std. Dev. = 56.44) further suggests considerable variation in performance that is not explained by the random effects.

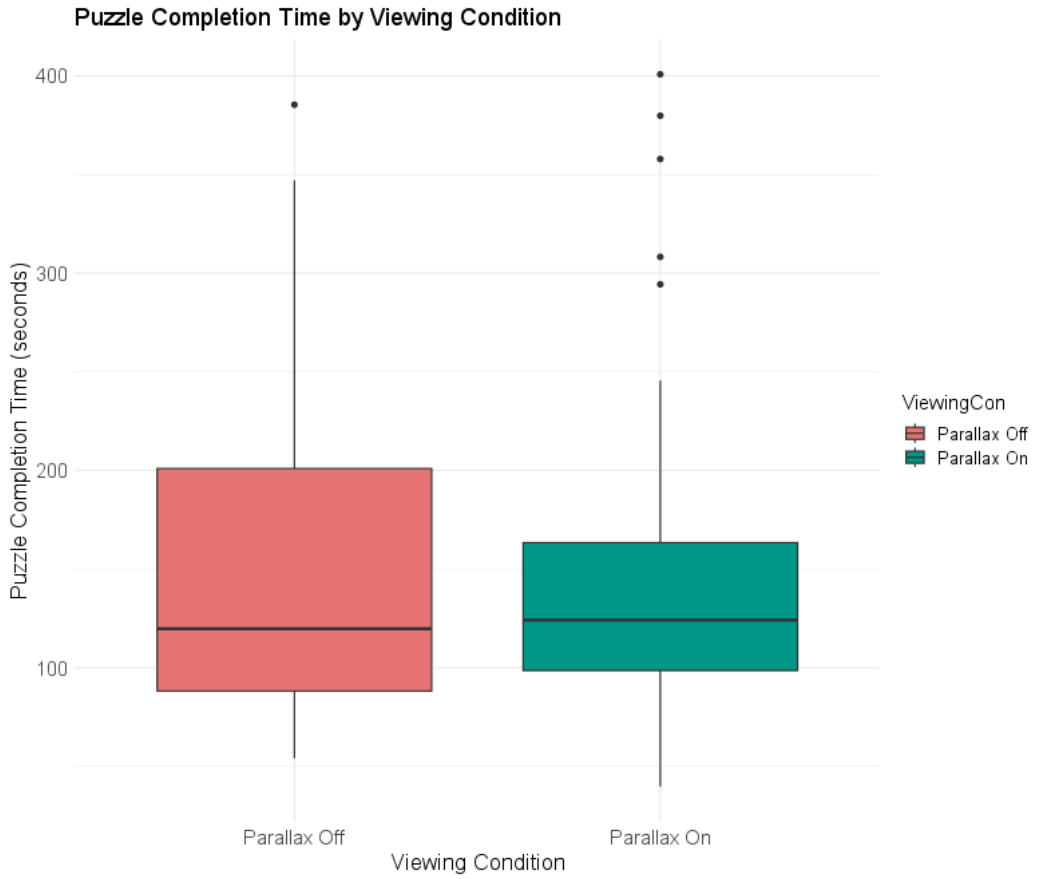

Figure 13

Bar graph of puzzle completion time by viewing condition

Box plot of puzzle completion time by viewing condition. The solid horizontal line represents the median, while black dots indicate outliers. Figures were created using ggplot2 in R (Wickham et al., 2016).

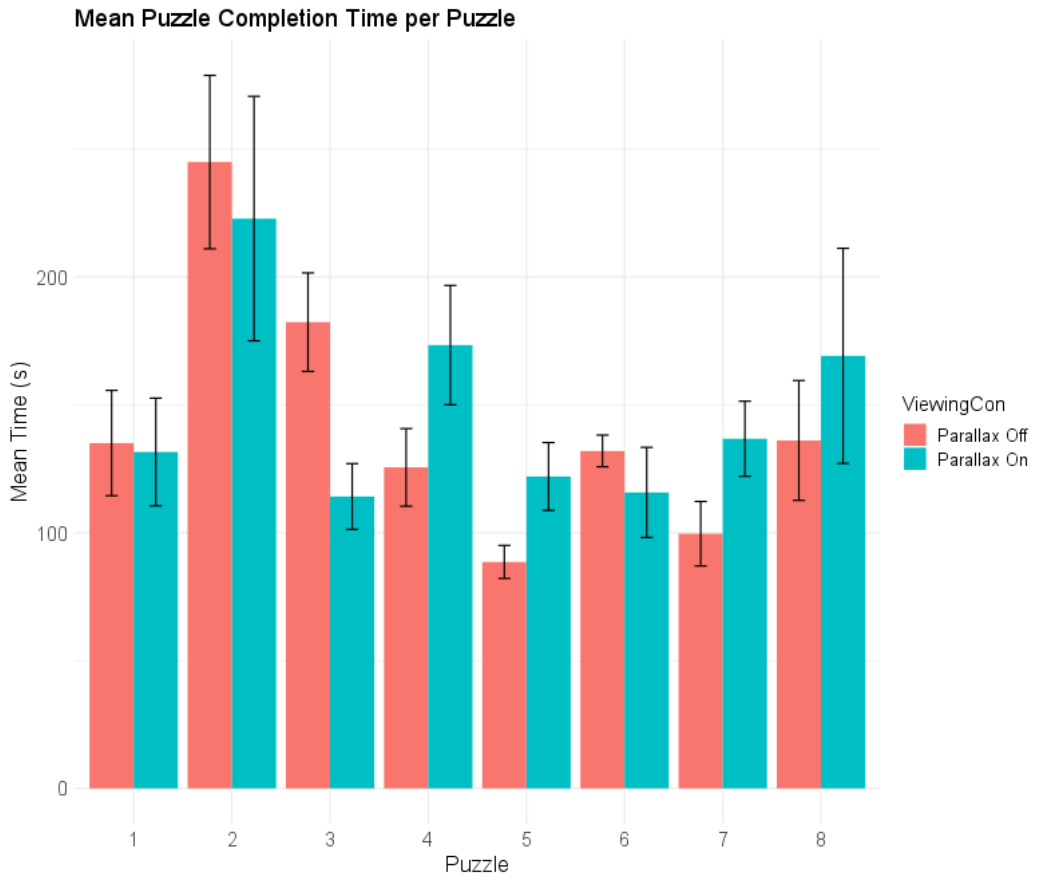

Figure 14

Bar plot of mean times per puzzle

Mean puzzle completion time (in seconds) for each puzzle under two viewing conditions: Parallax On and Parallax Off. Error bars represent the standard error of the mean. Figures were created using ggplot2 in R (Wickham et al., 2016).

Interpretation

Although participants completed puzzles slightly faster on average in the Parallax-On condition, this difference was not statistically significant. This suggests that motion parallax did not meaningfully enhance task efficiency. Instead, variability in puzzle difficulty and individual differences had a greater impact on completion time than the experimental condition.

Euclidean Distance

The model revealed that the viewing condition did not have a significant effect on the mean Euclidean distance of the player solution from the target solution (Estimate = -0.08 UU, Std. Error = 0.29, df = 142.59, t = -0.27, p = 0.79). The intercept was significant (Estimate = 2.13 UU, Std. Error = 0.32, df = 11.54, t = 6.57, p = 3.21e-05), indicating the overall mean Euclidean distance of the player solutions from the target solutions for the parallax off condition are significantly different from zero, indicating that the baseline value is not trivial (Figure 15).

Random effects revealed variability between players (Variance = 0.51 UU^2, Std. Dev. = 0.71) and between puzzles (Variance = 0.45 UU^2, Std. Dev. = 0.67), indicating that individual differences in performance and puzzle characteristics (Figure 16) contributed to variability in Euclidean distance. The residual variance (2.84 UU^2, Std. Dev. = 1.68) further suggests substantial unexplained variation in player performance.

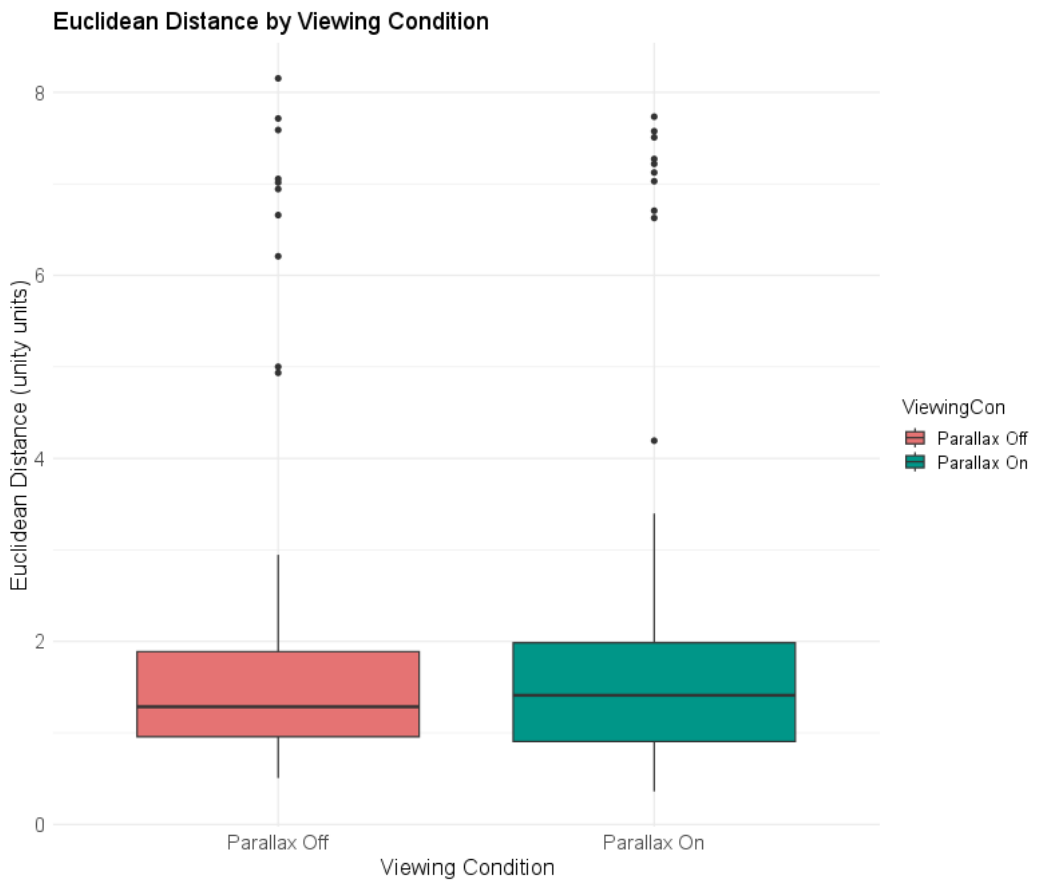

Figure 15

Bar graph of Euclidean distance by viewing condition

Box plot of Euclidean Distance by viewing condition. The solid horizontal line represents the median, while black dots indicate outliers. Figures were created using ggplot2 in R (Wickham et al., 2016).

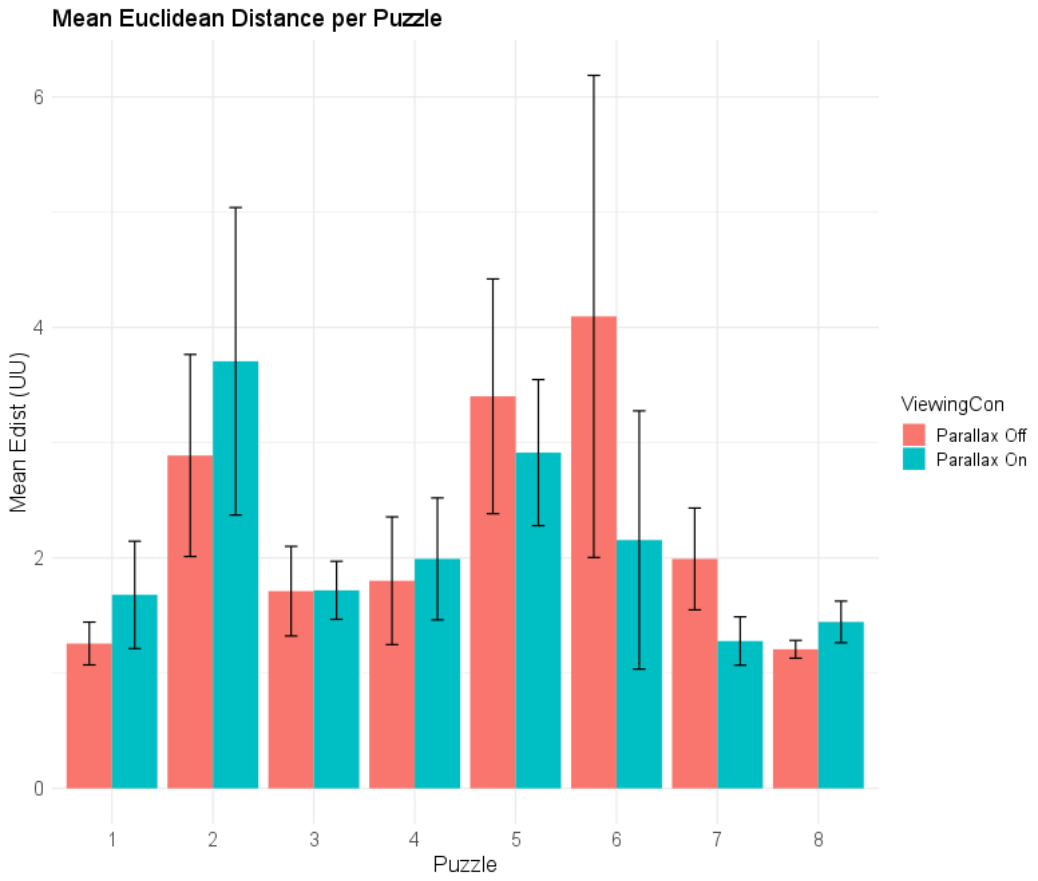

Figure 16

Bar plot of mean Euclidean distance per puzzle

Mean puzzle Euclidean Distance (UU) for each puzzle under two viewing conditions: Parallax On and Parallax Off. Error bars represent the standard error of the mean. Figures were created using ggplot2 in R (Wickham et al., 2016).

Interpretation

While the Parallax-On condition showed a slightly higher average Euclidean distance than the Parallax-Off condition, this difference was negligible and not statistically significant. As with completion time, individual and task-level differences contributed more to variation in performance than the viewing condition did.

Discussion

This study aimed to test whether introducing motion parallax in video communication could enhance communication efficiency by improving mutual gaze and directional cues. The results showed no significant differences between the Parallax-On and Parallax-Off conditions. This outcome raises questions about the theoretical underpinnings of motion parallax in virtual environments. The absence of significant improvement in puzzle completion time and Euclidean distance suggests that while motion parallax may theoretically provide better directionality and spatial cues, its practical impact on communication efficiency may be limited in the context of the tangram task. The large variability among players could hint at differences in experience with the tangram task or a difference in the threshold for novelty affecting participant performance.

O'Malley and colleagues (1996) noted that the novelty of video chat at the time could have created higher cognitive loads and been distracting. Like video chat in 1996, our technology in 2022 might have been more distracting than expected, and more time was needed to acclimate than the tutorial given prior to the trials (O'Malley et al., 1996). Future directions could observe how MPDepth2 affects the cognitive load of participants.

The tangrams themselves could have required more mental imagery due to their complex nature, which could have caused the students to close their eyes to eliminate distractions and focus on building the complex shape. This effect has been shown by Pearson and colleagues (2011), where participants reported needing to close their eyes more to create a more vivid mental image of what was being described. This would detract from the time spent looking at the other user, obscuring the impact of motion parallax. Although participants' experience with tangram puzzles was not explicitly recorded, it is possible that differences in prior exposure or skill level could have influenced performance. Should that be the case, lack of familiarity would

further increase the cognitive load needed to build complex shapes. Future experiments would need to control a set level of skill with the game prior to the task.

The avatars themselves could have been distracting, or their cartoonish nature may have distracted them from the ability to connect to the other player. Future work on the MPDepth app intends to add more realistic avatars using machine learning. Garau and colleagues (2003) experimented with comparing realistic avatars and stick figure avatars in virtual reality, with or without realistic gaze patterns. In their study, participants used a Cave VR system, while others used a traditional VR head-mounted system. Realistic avatars with a natural gaze pattern were found to have a higher perceived communication quality. The task in the current study, explaining a process, was primarily cognitive and did not require physical interaction or awareness of spatial relationships. As such, motion parallax may not have played a significant role in this context. The avatars may have served as a distraction rather than an aid. Future versions of the MPDepth app may add more realistic avatars and observe the potential cognitive effects.

Limitations

The unequal distribution of parallax conditions across the puzzles was also a study limitation. An effort was made to equally distribute the conditions among puzzles, but full distribution was not ultimately achieved. However, this limitation allows future research to use a more controlled design to ensure equal parallax conditions across all puzzles. Adjusting this could contribute to a more robust experimental design. The absence of eye tracking and a post-experiment questionnaire substantially limited the types of insights that could be drawn from the data and constrained the scope of possible conclusions. A condition removing the avatar entirely would have provided insight into if participants used the avatar at all during the experiment.

Future Directions

Observing motion parallax's effects on communication efficiency in video chat is still worth pursuing as the limitations and results from the current study open new doors for further research. However, different protocols and dependent variables would need to be observed. Specifically, the Travel Game might be worth pursuing, especially as the task relies on directional cues and could be a better fit for future work than the Tangram task. Eye tracking is more pertinent considering how the participants communicate with each other and where they look. The map used in the Travel Game would eliminate the need for mental imagery, as the same map is on both the confederate and the participant's screen.

However, during sessions of the Travel Game, it was observed that participants interrupted each other more frequently in the video chat conditions. One possible explanation is that participants may have needed more words to compensate for the missing non-verbal cues (A. H. Anderson et al., 1996). It would be valuable to examine whether participants interrupted each other more in video chat with or without motion parallax. This could provide insight into whether motion parallax provides nonverbal directional cues like face-to-face interactions. Other aspects to observe would be turn-taking behaviour, comprehension via the ability to reach the target and post-experiment questionnaires, and local and remote attention via eye tracking.

Future versions of the MPDepth app will contain more realistic avatars, even going as far as machine learning is concerned, creating motion parallax using images of objects. Observing whether this affects behaviour between realistic and cartoon avatars would be an interesting direction once the technology is complete.

Conclusion

The current study did not fully elucidate the effect of motion parallax on communication efficiency in video chat. Other instruction tasks more suited to completing the data set could pursue additional avenues. Motion parallax alone may not improve communication efficiency, but it may be worth pursuing with other communication paradigms.

References

Anderson, A., & Garrod, S. C. (1987). The dynamics of referential meaning in spontaneous conversation: Some preliminary studies. *Communication Failure in Dialogue and Discourse*, 161–183.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The Hcrc Map Task Corpus. *Language and Speech*, *34*(4), 351–366. https://doi.org/10.1177/002383099103400404

Anderson, A. H., Brown, G., Shillcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. https://www.research.ed.ac.uk/en/publications/teaching-talk-strategies-for-production-and-assessment

Anderson, A. H., Newlands, A., Mullin, J., Marie Fleming, A., Doherty-Sneddon, G., & Van der Velden, J. (1996). Impact of video-mediated communication on simulated service encounters. *Interacting with Computers*, *8*(2), 193–206. https://doi.org/10.1016/0953-5438(96)01025-9

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze* (pp. xi, 210). Cambridge U Press.

Auguie, B., & Antonov, A. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics* (Version 2.3) [Computer software]. https://cran.r-project.org/web/packages/gridExtra/index.html

Bailenson, J. N. (2021). Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior*, *2*(1). https://doi.org/10.1037/tmb0000030

Bates, D., Maechler, M., Bolker [aut, B., cre, Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & simulate.formula),

P. N. K. (shared copyright on. (2023). *lme4: Linear Mixed-Effects Models using "Eigen" and S4* (Version 1.1-32) [Computer software]. https://CRAN.R-project.org/package=lme4

Bates, D., Maechler, M., Jagan, M., libraries, T. A. D. (SuiteSparse, dir(system.file("doc," collaborators listed in, SuiteSparse, package="Matrix"), pattern="License", full.names=TRUE, recursive=TRUE)), library, G. K. (METIS, Minnesota), C. R. of the U. of, onenormest(), J. R. (GNU O. condest() and, California), C. R. of the U. of, nearPD()), J. O. (initial, & implementation), R. C. T. (base R. matrix. (2024). *Matrix: Sparse and Dense Matrix Classes and Methods* (Version 1.7-0) [Computer software]. https://cran.r-project.org/web/packages/Matrix/index.html

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. *Language and Speech*, *37*(1), 1–20. https://doi.org/10.1177/002383099403700101

Brennan, S., E. (1998). The grounding problem in conversations with and through computers. In *Social and cognitive approaches to interpersonal communication.* (pp. 201–225).

Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, *58*(1), 154–161. https://doi.org/10.1016/j.compedu.2011.08.024

Chen, E., Kaczmarek, K., & Ohyama, H. (2021). Student perceptions of distance learning strategies during COVID-19. *Journal of Dental Education*, *85*(S1), 1190–1191. https://doi.org/10.1002/jdd.12339

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Croes, E. A. J., Antheunis, M. L., Schouten, A. P., & Krahmer, E. J. (2019). Social attraction in video-mediated communication: The role of nonverbal affiliative behavior. *Journal of Social and Personal Relationships*, *36*(4), 1210–1232. https://doi.org/10.1177/0265407518757382

Dalzel-Job, S. (2015). *Social interaction in virtual environments: The relationship between mutual gaze, task performance and social presence*. https://era.ed.ac.uk/handle/1842/15792

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, *3*, 105–125. https://doi.org/10.1037/1076-898X.3.2.105

Fujimoto, K., & Ashida, H. (2020). Different Head-Sway Responses to Optic Flow in Sitting and Standing With a Head-Mounted Display. *Frontiers in Psychology*, *11*, 577305. https://doi.org/10.3389/fpsyg.2020.577305

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., & Kramer, A. D. I. (2004). Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human–Computer Interaction*, *19*(3), 273–309. https://doi.org/10.1207/s15327051hci1903_3

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 529–536. https://doi.org/10.1145/642611.642703

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Action as language in a shared visual space. *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work - CSCW '04*, 487. https://doi.org/10.1145/1031607.1031687

Haddington, P. (2006). The organization of gaze and assessments as resources for stance taking. *Text & Talk - TEXT TALK*, *26*, 281–328. https://doi.org/10.1515/TEXT.2006.012

Hancock, J. T., & Dunham, P. J. (2001). Impression Formation in Computer-Mediated Communication Revisited: An Analysis of the Breadth and Intensity of Impressions. *Communication Research*, *28*(3), 325–347. https://doi.org/10.1177/009365001028003004

Hecht, H., Boyarskaya, E., & Kitaoka, A. (2014). The Mona Lisa effect: Testing the limits of perceptual robustness vis-à-vis slanted images. *Psihologija*, *47*(3), 287–301. https://doi.org/10.2298/PSI1403287H

Hinds, P. J. (1999). The Cognitive and Interpersonal Costs of Video. *Media Psychology*, *1*(4), 283–311. https://doi.org/10.1207/s1532785xmep0104_1

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100. https://doi.org/10.1037/0033-2909.100.1.78

Koenderink, J., & van Doorn, A. (2008). The Structure of Visual Spaces. *Journal of Mathematical Imaging and Vision*, *31*(2), 171. https://doi.org/10.1007/s10851-008-0076-3

Koenderink, J., & van Doorn, A. (2012). Gauge Fields in Pictorial Space. *SIAM Journal on Imaging Sciences*, *5*(4), 1213–1233. https://doi.org/10.1137/120861151

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual Information as a Conversational Resource in Collaborative Physical Tasks. *Human–Computer Interaction*, *18*(1–2), 13–49. https://doi.org/10.1207/S15327051HCI1812_2

Kraut, R. E., Miller, M. D., & Siegel, J. (1996). Collaboration in performance of physical tasks: Effects on outcomes and communication. *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*, 57–66.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Makowski (@Dom_Makowski), D., Lüdecke (@strengejacke), D., Patil (@patilindrajeets), I., Thériault (@rempsyc), R., Ben-Shachar (@mattansb), M. S., Wiernik (@bmwiernik), B. M., Siegel, R., & Bock, C. (2024). *report: Automated Reporting of Results and Statistical Models* (Version 0.5.9) [Computer software]. https://cloud.r-project.org/web/packages/report/index.html

Maruyama, K., Endo, M., & Sakurai. (1985). An Experimental Consideration on" Mona Lisa Gaze Effect. *Tohoku Psychologica Folia*, *44*(1–4), 109–121.

Mukawa, N., Oka, T., Arai, K., & Yuasa, M. (2005). What is connected by mutual gaze? User's behavior in video-mediated communication. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, 1677–1680. https://doi.org/10.1145/1056808.1056995

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, *8*(2), 177–192. https://doi.org/10.1016/0953-5438(96)01027-2

Patil, I. (2021). Visualizations with statistical details: The "ggstatsplot" approach. *Journal of Open Source Software*, *6*(61), 3167. https://doi.org/10.21105/joss.03167

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the Mind's Eye: The Metacognition of Visual Imagery. *Psychological Science*, *22*(12), 1535–1542. https://doi.org/10.1177/0956797611417134

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rogers, S. J. (2013). Mutual Gaze. In F. R. Volkmar (Ed.), *Encyclopedia of Autism Spectrum Disorders* (pp. 1966–1967). Springer. https://doi.org/10.1007/978-1-4419-1698-3_628

Rogers, S., Lunsford, M., Strother, L., & Kubovy, M. (2003). The Mona Lisa effect: Perception of gaze direction in real and pictured faces. *Studies in Perception and Action VII*, 19–24.

Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of Telecommunications*. Wiley.

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., Love, J., Lenth, R., & Christensen, R. H. B. (2023). *afex: Analysis of Factorial Experiments* (Version 1.2-1) [Computer software]. https://CRAN.R-project.org/package=afex

Tangram Channel. (2023). *Hundreds of tangram puzzles to solve*. Providing Teachers and Pupils with Tangram Activities. http://www.tangram-channel.com/

Troje, N. F. (2023). Zoom disrupts eye contact behaviour: Problems and solutions. *Trends in Cognitive Sciences*, *27*(5), 417–419. https://doi.org/10.1016/j.tics.2023.02.004

VR Cardboard Buddies. (2016). *Tangram Game Concept | Packs | Unity Asset Store*. https://assetstore.unity.com/packages/templates/packs/tangram-game-concept-73557

Wang, X. M., Bebko, A. O., Thaler, A., & Troje, N. F. (2020). Stereopsis Aids Perceived Distance Based on An Exocentric Pointing Task. *Journal of Vision*, *20*(11), 1171. https://doi.org/10.1167/jov.20.11.1171

Wang, X. M., & Troje, N. F. (2023). Relating visual and pictorial space: Binocular disparity for distance, motion parallax for direction. *Visual Cognition*, *31*(2), 107–125. https://doi.org/10.1080/13506285.2023.2203528

Webb, M. (2021). Zoom Fatigue and How to Prevent It. *Journal of Registry Management*, *48*(4), 181–182.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Posit, & PBC. (2016). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (Version 3.4.2) [Computer software]. https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., Software, P., & PBC. (2023). *dplyr: A Grammar of Data Manipulation* (Version 1.1.4) [Computer software]. https://cran.r-project.org/web/packages/dplyr/index.html

Wickham, H., Vaughan, D., Girlich, M., Ushey, K., Software, P., & PBC. (2024). *tidyr: Tidy Messy Data* (Version 1.3.1) [Computer software]. https://cran.r-project.org/web/packages/tidyr/index.html