Leveling the Playing Field: A Comparative Reevaluation of Unmodified Eye Tracking as an Input and Interaction Modality for VR

Ajoy S. Fernandes *
Meta Reality Labs Research

T. Scott Murdison†
Meta Reality Labs

Michael J. Proulx‡
Meta Reality Labs Research

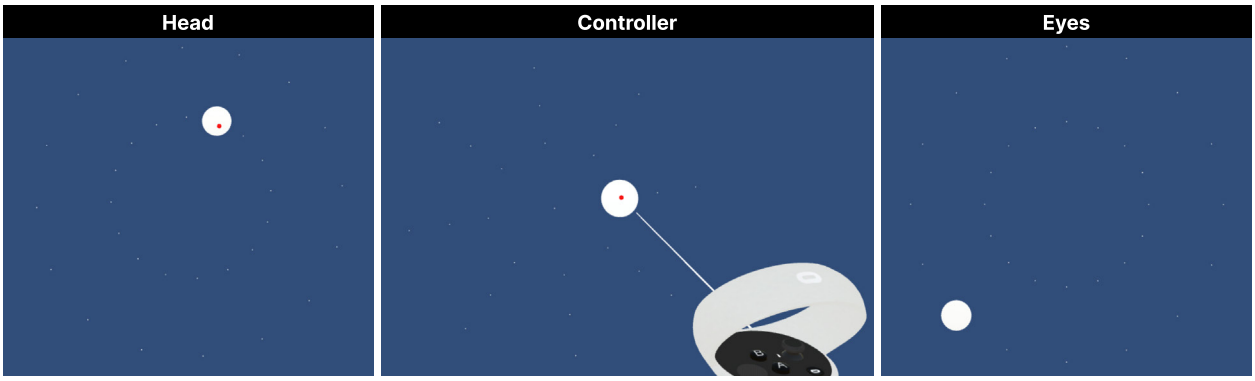

Fig. 1: Head input modality (left) with red cursor on target using ISO 9241-9 style double rings target positioning. Faint target placeholder dots show where targets can appear. Controller input modality (center), using the same red cursor, with a 0.5m thin white rod, using Random Web target positioning. Unmodified Eye input modality (right) with no cursor feedback.

Abstract— In this study, we establish a much-needed baseline for evaluating eye tracking interactions using an eye tracking enabled Meta Quest 2 VR headset with 30 participants. Each participant went through 1098 targets using multiple conditions representative of AR/VR targeting and selecting tasks, including both traditional standards and those more aligned with AR/VR interactions today. We use circular white world-locked targets, and an eye tracking system with sub-1-degree mean accuracy errors running at approximately 90Hz. In a targeting and button press selection task, we, by design, compare completely unadjusted, cursor-less, eye tracking with controller and head tracking, which both had cursors. Across all inputs, we presented targets in a configuration similar to the ISO 9241-9 reciprocal selection task and another format with targets more evenly distributed near the center. Targets were laid out either flat on a plane or tangent to a sphere and rotated toward the user. Even though we intended this to be a baseline study, we see unmodified eye tracking, without any form of a cursor, or feedback, outperformed the head by 27.9% and performed comparably to the controller (5.63% decrease) in throughput. Eye tracking had improved subjective ratings relative to head in Ease of Use, Adoption, and Fatigue (66.4%, 89.8%, and 116.1% improvements, respectively) and had similar ratings relative to the controller (reduction by 4.2%, 8.9%, and 5.2% respectively). Eye tracking had a higher miss percentage than controller and head (17.3% vs 4.7% vs 7.2% respectively). Collectively, the results of this baseline study serve as a strong indicator that eye tracking, with even minor sensible interaction design modifications, has tremendous potential in reshaping interactions in next-generation AR/VR head mounted displays.

Index Terms—Eye tracking, User experience, Input devices, 3D user interaction, Human factors and ergonomics, Gaze targeting

1 Introduction

Eye tracking is crucial in enabling next-generation AR/VR interactions and applications. Over the years, advancements in hardware, sensors, AI, and compute capability in commercially available technology have progressed from enabling reliable head-tracking to controller-tracking and hand-tracking. Advancements in each have allowed us to go from head targeting, as seen in the HoloLens 1, to rich hand-tracked manipulations, as seen with the Meta Quest 2 today [17, 52–54]. Eye tracking sensors have shown up in devices like the HTC Vive Pro Eye, Magic Leap 1, HoloLens 2, Meta Quest Pro, and Pico Neo 3 Pro Eye, and will be a foundational component of future AR/VR devices. As it advances in capability, eye tracking has the unique potential to unlock intuitive and efficient interaction schemes alongside head, controller, hand tracking, and many schemes that would otherwise be impossible without it. Some examples include eye tracking for hands free interaction paradigms [34], for contextualized interactions [6], gaze accelerated interactions [35], or as an interaction disambiguator [40]. Yet, surprisingly, the foundational work which allows exploring the advantages of eye tracking in the unique contexts afforded by AR/VR has not been firmly established.

Given multiple possible input modalities in a head-mounted display (HMD; e.g., controller, head, eyes; see Fig. 1), it can be difficult to predict the relative performance of one versus another. How do we compare human performance across these input modalities? Comparing inputs is not straightforward and often depends on the context in which the particular input is used. For example, voice inputs might be promising within the privacy of one's room but might be challenging in more public settings. Similarly, carrying a controller or using one's hands for direct manipulation away from one's home or office poses similar challenges. Eye tracking typically provides the privacy and portability lacking in most other inputs available today. There are many more trade-offs to consider, but the groundwork to effectively compare

*e-mail: ajoyferns@meta.com
†e-mail: smurdison@meta.com
‡e-mail: michaelproulx@meta.com

Manuscript received 14 October 2022; revised 13 January 2023; accepted 30 January 2023. Date of publication 22 February 2023; date of current version 29 March 2023. Digital Object Identifier no. 10.1109/TVCG.2023.3247058

trade-offs between input methods and interaction techniques in a fair and ecologically valid manner lacks a consistent basis.

While targeting and selection are only one of many potential interaction combinations enabled in AR/VR today, it remains prevalent through commercial headsets as a method of interacting and navigating through UIs in virtual environments. The 3D head raycast enables convenient and reliable targeting and selection without the additional risks and costs associated with controller or hand tracking hardware. The 6 DOF (degrees of freedom) controller functions as a low-risk intuitive input device, enabling direct and indirect interactions in 3D [7]. These two input methodologies have found their way into many AR/VR products today, like the Meta Quest 2, HTC Vive, HoloLens 1, and Magic Leap 1. Hand tracking technologies continue to evolve and improve [17, 52–54]. Meanwhile, when considering eye tracking in AR/VR, it brings up an immediate reaction of words like "eye-fatigue [36]," "Midas Touch, [20]" or "High Risk." Eye tracking as an input selection modality is not new, and was first explored decades ago as a potentially empowering tool [55]. The reputation of eye tracking has been hurt, though, by some strict recommendations on its use in the absence of clear evidence. An oft-cited example [19] notes without citations or evidence that people only position the eye within one degree of accuracy, with the suggestion that this is due to the width of the fovea; this has led others to conclude incorrectly that eye tracking cannot target anything smaller than this due to an anatomical limitation [32], without any consideration of eye tracking fidelity. In contrast, the original work proposing gaze as an input modality for targeting noted that the eye can select objects accurately as 10 minutes of visual angle [55], so the misconception that the eyes can only select targets of 1 degree or larger is holding back the promise of gaze.

Some past reports have found that within targeting and selection tasks, particularly in VR, eye tracking has sub-par performance relative to head-tracking [39] using the ISO 9241-9 reciprocal selection task [59], and other methods [25,31]. We see similar results when comparing eye tracking to mouse input and again for head-tracking [18]. There is a prior study [48] that found a higher throughput compared to these prior experiments when using gaze; however, it can be challenging to make comparisons across studies and this work did not have any comparison to common modalities, which we utilize in our study. Unlike what we provide in our work, many prior studies, including these three discussed here [18, 39, 48] do not measure or report eye tracking calibration statistics in their reports; at best they report the manufacturer's stated eye tracking quality, but even that is rare. In cases where the eye tracker performance is poor, but not reported on a participant level, it is impossible to know whether poor interaction efficiency is due in earnest to the experimental manipulation or simply the result of tracking errors. This perhaps gives rise to the mistaken assumptions about the limits of eye tracking, such as the misconception noted above that targets must subtend one degree of visual angle due to the size of the fovea [32], even though there is no such limitation. Perhaps this is why many studies of targeting and selection in VR have primarily focused on hand, wand, or controller modalities [3, 4, 8].

Other studies use the eyes to drive a visible cursor [48]. Given that eye movements are at least partially driven by low-level retinal inputs [41] and that the oculomotor system is not evolved to physically manipulate objects in the environment [42], the feedback provided to the observer is extremely important to ensure consistency with visuomotor expectations [33]. This is one reason why we chose not to provide any cursor or online visual feedback in our study's eye tracking condition. To date, errors that arise in most eye tracking pipelines (e.g., latency, bias, noise) have made eye tracking practically unusable without significant workarounds in the user interface. Unfortunately, these workarounds tend to detract from the design and intuitive use of the system (e.g., larger target sizes, longer dwell times [47]). Noting that poor eye tracking performance has been a limitation in the past, we will use a newer and higher-performing eye tracker to look at the effects of sub 1 degree mean accuracy errors running at an effective frame rate of 90Hz [57]. These eye tracking performance characteristics will enable us to effectively measure human performance on designs closer to those within many UI constraints prevalent in next-generation devices.

It is essential to establish a baseline of comparison for targeting and selection with eye tracking, most closely representing next-generation AR/VR hardware, interaction models, and user interfaces. This study attempts to lay that foundation by first comparing unmodified eye tracking, using only the signal as provided from the eye tracker's API without any visual feedback or adjustment, and a controller trigger press for selection. We compare eye tracking to other heavily used inputs for AR/VR, including head tracking (e.g., HoloLens 1) and controller 6-DOF tracking (e.g., Magic Leap, Meta Rift and Quest products, HTC Vive). We will use controller tracking as an indirect proxy to hand targeting and selection methods, as controller tracking often offers higher fidelity tracking and selection, because it is simply less prone to computer vision errors and physical and behavioral idiosyncrasies associated with users. For example, a review of past works found that the mouse was more efficient than a touchpad for this reason [49]. This study does not look into optimizing eye tracking based human performance and thus does not provide a general conclusion on eye tracking for interactions, but instead sets the baseline for one specific but highly prevalent targeting and selection model. As we provide no visual feedback to the user, and we do not intend to, we will likely not obtain optimal human performance in subjective and objective terms. We also provide no form of manipulation to the mechanics of the interaction, such as performing a nearby search and match of targetable elements or adjustment to target colliders. Forming such a baseline and understanding of using unmodified eye tracking for gaze interaction will allow future refinement of interaction models containing additional affordances and feedback methods.

Many eye tracking studies have looked at targets locked to a display or monitor [46, 55]. This target placement, however, does not represent many interactions now capable in AR/VR devices where virtual objects can maintain their positions in world space, and users are free to roam around as they wish [16]. This study proposes a method of targeting and selection that remains ecologically valid for numerous targeting and selection tasks available in AR/VR apps today and in the near future. Targets remain locked in world space for targeting and selection to occur, and this layout remains consistent for each of the three modalities presented.

How do we effectively compare eye tracking with other input methods for world-locked targeting and selection tasks? Fitts's Law has historically provided a rigorous basis by which novel interaction schemes can be quantifiably tested [11, 37, 49]. Interestingly, even Fitts attempted to study the role of eye movements in complex tasks, though the technology was not yet at a level where it could be assessed using the law that is attributed to him [12]. However, throughput comparisons in isolation may not be the sole indicator of the success of an input method. That is, not all interactions in AR/VR require a user to navigate through a UI quickly. Furthermore, some input methods could produce a high throughput but are incredibly fatiguing over a longer duration. Others might take a long time to learn, while others are prone to a high error rate. Regardless, alongside many other measures, throughput does serve as one effective means of comparing inputs but should not be evaluated in isolation.

Fitts's Law traditionally assumes that targets are set on a 2D plane [49]. Because targeting and selection in AR/VR typically involve a form of raycasting in 3D, the targets with the same metric size will have different angular sizes based on the distance off-center they are from the user, despite being on the same plane perpendicular to the user's forward direction. The actual targetable area is smaller despite having the same metric size at further distances. Furthermore, a target that rotates toward a user will, from the user's point of view, have a larger targetable area than one that does not. That brings to question that in targeting tasks in AR and VR, rather than having UI elements flat on a planar surface, UI should instead be rotated towards the user and perhaps be laid out radially relative to the user to maximize the angular targetable area. Given depth and fully navigable space in VR, we argue that the use of targetable areas in angular space are more meaningful for VR-based interaction comparisons than targetable areas fixed in metric space. Our study considers this by placing targets, both on a plane with a z-distance 1 meter away from the user and by placing

targets radially at a euclidean distance 1 meter from the user while rotating towards them (see Fig. 3 and 4).

Satisfaction in targeting and selection tasks might be linked to the error rate. What percentage of intended hits end up being hits, and what number end up being misses? Current standards, like the ISO 9241-9 reciprocal selection task, leave current eye tracking inputs at a disadvantage. Eye tracking accuracy in many of today's systems typically falls off at larger eccentricities from a tracker's center [10], likely attributing to higher error rates. This study explores targeting and selection with an alternative random web layout of targets and successions while also using the ISO 9241-9 as a condition.

Finally, despite evaluation from all objective measures, the user subjectively determines if they prefer a particular form of input over another. We capture subjective evaluations from users about their preference of input methodologies and their perceived fatigue as they complete blocks.

Putting this all together, this is the first study that sets a baseline for comparing eye tracking to other key interaction modalities in an ecologically valid manner in VR. Specifically this paper makes the following contributions:

1. Unlike many studies which use head locked targets, we systematically form a baseline of unmodified eye tracking performance as an interaction modality, with a button press for selection, using world-locked targets. That is, how does relatively capable VR eye tracking running at (approx. 90Hz and $<$1.0 degree average accuracy error) perform as a targeting and selection modality?
2. This is the first study that we are aware of that compares eye tracking to a combination of other well known interaction modalities such as controller and head inputs, as used in AR/VR today. We establish that eye tracking with trigger selection, even without any treatments, can serve as a targeting and selection modality that is comparable to controller targeting and selection, and in most measures performs better than head tracking using both objective and subjective measures.
3. We question ISO 9241-9 standard as an appropriate method of performance, specifically for eye tracking. We propose another layout for targeting and selection tasks in AR/VR.
4. We compare the effects of laying targets in spherical vs planar coordinates.

In the remainder of this paper, we will dive deep into the experiment design decisions and their corresponding implementation. Next, we present a user study that compares eye tracking with controller and head targeting and selection in planar or spherical coordinates, using either a variant of the ISO 9241-9 reciprocal selection task or an alternative we provide. We then discuss the study's results and present our conclusions and directions for future work.

2 Equipment

We used a modified off-the-shelf VR HMD, the Meta Quest 2 headset, with corresponding controllers and a display refresh rate set at 120Hz. The display has an approximate per-eye field of 92x96 degrees of visual angle (deg), resolution of 1832x1920 pixels, with an approximate pixel density of 21 pixels per degree [22]. We used an HMD-integrated binocular XR eye tracking platform from Tobii (Tobii AB, Sweden). This eye tracker has a sampling frequency of 240 Hz and is based on Tobii's latest generation off-axis (direct to eye) solution for VR and AR optical designs – including 'pancake' lens designs common in newer VR products. We used an MSI GS66 Stealth laptop to power this experience, running an NVIDIA GeForce RTX 3070 graphics card. The experiment application was developed in Unity, and runs at an average measured framerate of 87 frames per second.

3 User Study

3.1 Experiment Design

We tested the following conditions (some illustrated in Fig. 1):

1. Input Modality (Eye, Controller, Head)
2. Standard (Double-ISO, Random-Web)
3. Geometry (Planar, Spherical)
4. Target Diameter (3, 4, 5 degrees)

These result in [3*2*2*3 = 36 unique blocks] per participant. Given the large number of blocks, we randomly counterbalanced the experiment by shuffling the order of blocks for each participant before they began the study. A Latin Square or similar method was not used.

In each block, participants are presented with a sequence of 25 or 36 targets for Random-Web or Double-ISO style target sequences, described in detail below. Following every block, users completed an in-app survey. Participants went through 1098 targets (366 per input condition) through the course of the experiment.

Fig. 2: Eye calibration quality utility, showing typical results, ran before each eye tracking block. Eye calibration keyboard button presses are managed by the experimenter.

3.2 Practice Round

Each participant went through a practice version of the experiment. They are shown, in a fixed order, the three experiment blocks, representing controller, head and then eye input modalities. Participants are calibrated before performing the eye tracking block. We provided only the Double-ISO standard, Spherical geometry, and 5-degree targets across each practice block (we describe conditions below). We used only the Double-ISO standard as we wanted all participants to learn the fixed pattern associated with the progression of targets for that standard before the main experiment. We provided 5-degree targets and the Spherical standard as these would be easier to target, learn, and generalize to the experiment mechanics.

3.3 Participants

We recruited 32 participants within our general organization, and 30 were successfully able to complete the experiment. The two unsuccessful attempts were a result of wire disconnection when the two physically adjusted their headset, which compromised their attempts. Out of the 30 participants who completed the experiment, 29 completed the survey used to provide the stats in this section (10 female; age range 24–53, average 37 years old); 14 participants reported that they had corrected vision, out of whom 9 wore contact lenses during the experiment, 1 wore glasses, 4 wore neither, and 1 had corrective refractive surgery; 12 participants indicated that they spend no time in VR per week, while 10 participants indicated that they spend between 0.5-3 hours, and 4 participants indicated they spend $>$ 3 hours.

3.4 Eye Calibration

Participants were required to calibrate the eye tracking system at the beginning of the experiment. Immediately following this, they were required to check the quality of the calibration through a test sequence. Participants were allowed to proceed assuming that their mean calibration error was less than 1.5 degrees, or otherwise recalibrated.

After the first block, participants were programmatically prompted to check the quality of the calibration (see Fig. 2) before every eye tracking block (for a total of 12 blocks). If the experimenter accompanying participants during the study believed that participants could

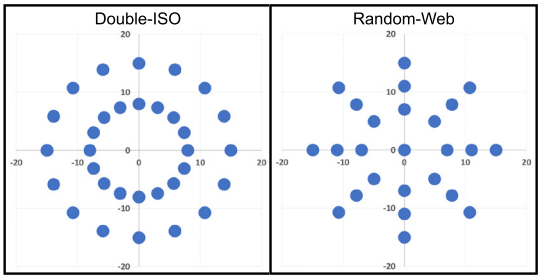

Fig. 3: Target placed in Double-ISO style standard (left) vs. Random-Web style alternative (right). Angular placements were consistent in both Planar vs. Spherical coordinates, but targets were rotated towards the dummy-camera in Spherical, at 1m Euclidean distance, whereas targets were placed at 1m z-distance away from the dummy-camera in Planar. In VR the targets were actually white on blue backgrounds.

improve the calibration quality, they would recalibrate and retest calibration quality before continuing the study. We recorded calibration statistics every time the participant completed the calibration quality check.

3.5 World-Locked Target Design and Input Modalities

Target Appearance: It is known within the literature for eye tracking that particular target designs are more effective than others in encouraging stable and accurate fixations [50]. We use such designs in our eye calibrator. However, most AR/VR UI elements do not practically fit within these constraints and are not necessarily suited for controller or head-based selections. Hence, we used plain white, circular, flat targets with no texture or apparent visual markers to enable a baseline comparison. Past work [39] varied size and luminance contrast, which is known to capture attention [38], so that was avoided here, and was consistent across all trials.

Target Placement: Before each block, we place targets in front of the user, in world coordinates, with the centroid of the group of targets set at the user's head height, regardless of head orientation. We do this to ensure that the targets will be centered directly in front of the user at the start of each block, representing many UI panels in AR/VR headsets today. We accomplish this placement by using a "dummy camera," which remains static in world coordinates through each block. We place targets in the dummy camera's coordinate frame before each block. The user remains seated throughout the session, meaning that the dummy camera coordinate frame approximates an upright, primary head position.

Right before each block, we set the dummy camera to match the position and yaw of the main camera. The dummy camera then maintains that position and yaw throughout that block. The dummy camera, however, will hold a pitch and roll of 0. That is, the forward vector of the dummy camera will be parallel to the ground and perpendicular to gravity through all trials. We do this to ensure that targets are perpendicular to the ground while maintaining their position in world coordinates across all trials per block.

From a user's perspective, they will see a set of target placeholders appearing at the start of each block, representing where targets will be positioned, but they will only see one target at a time (see Fig. 1). Target placeholders are world locked, set to 0.1 degrees at the start of the block, but then maintain a constant size in meters – that is, targets and target placeholders will appear static as participants rotate and move their heads.

Target Diameters: We set target diameters at the start of each block to an intended angular size of either 3, 4 or 5 degrees per block. All targets/trials per block will maintain the same metric size, position, and rotation through the block. The user can move their head positionally or rotationally during a block. Because of this, the visual angle of targets can appear larger or smaller if the user moves closer or further away from the targets.

3.5.1 Standard: Double-ISO vs Random-Web

Double-ISO: To allow us to compare our results to studies in the past, we maintained standards from the ISO 9241-9 reciprocal selection task, with a ring of 18 targets.

An issue with this task, in particular, is that if we use only a single diameter for our ring of targets, especially if we use a large diameter ring, we will likely limit eye tracking inputs. This occurs because most eye trackers have more significant accuracy errors towards fringe angles [10]. To offset this issue, rather than having a single ring per block, we had the users go through two rings, an outer ring followed by an inner ring, totaling 36 targets per block.

We have users start targeting from the rightmost point, going to the leftmost point, maintaining this alternating sequence rotating counter-clockwise until all targets in the outer ring have been selected. The outer ring is completed first before the inner ring. Participants will see the same sequence of 36 targets/trials for each block. For each of the corresponding 18 blocks, targets follow the same alternating pattern, starting and ending on the same target.

As we show in the left box of Fig. 3, from the dummy camera, we placed the 18 targets comprising the inner ring at a 16-degree diameter (8 degree radius). We set the 18 targets compromising the outer ring at a 30-degree diameter (15 degree radius).

Random-Web: For each of the 18 Random-Web blocks, a participant will see a shuffled set of targets appearing in one of 25 predefined locations without repetition. As illustrated in the right box of Fig. 3, predefined locations include one target at the center of the dummy camera's FOV. Next, 8 targets are equally spaced and placed within an inner ring with a radius of 7 degrees. We placed another 8 at a radius of 11 degrees and the final 8 at 15 degrees from the center.

3.5.2 Target Geometry: Planar vs Spherical

Planar: As shown in Fig. 4, Planar placement is akin to placing targets of the same metric size on a wall. Target positions (X,Y) are in degrees relative to the dummy camera, and Z is the z-distance of the targets from the dummy camera in meters. Again this is not the radius or Euclidean distance from the dummy camera. Targets are placed onto the z-plane and are not rotated towards the camera. Consequently, targets further from the center will have smaller angular sizes than those closer to the center, while the metric size remains constant. In this experiment, all targets in planar coordinates are placed at z=1 away from the dummy camera, meaning that the euclidean distance to the target will be > 1 for those off center from the dummy target.

Spherical: In this placement (see Fig. 4), from the user's perspective, if all targets were visible at the start of each block, they would appear to form a sphere around the user's head. When viewed from the dummy camera, the centers of the targets will appear to be placed in the same angular coordinates as "Planar Coordinates." However, there are some key differences: First, we place targets at a fixed radius rather than a fixed z-plane. That is, the euclidean distance to the target from the dummy camera will now always be = 1, whereas previously, distances in Planar Coordinates could be $>=1$.Second, we rotate targets towards the dummy camera. Because of this, the target's perceived angular size will match the target's intended angular size.

3.5.3 Input Modalities And Feedback

Head: We raycast from the center of the head (main) camera. A flat cursor is visible on the plane's surface for Planar Geometry. Spherical Geometry has a cursor at either a fixed radius of 1m, or the distance of the point of intersection of the rotated target. The simple flat cursor represents those used in many past commercially available head-based UIs (like the HoloLens 1). The cursor appears just in front of the raycasted surface (rather than a fixed distance from the head) to minimize vergence/focus issues [24].

Controller: We raycast in the direction of the forward vector of the Meta Quest 2 controller. The cursor visual here is identical to that of the "Head" input condition. In addition, a thin white rod, starting

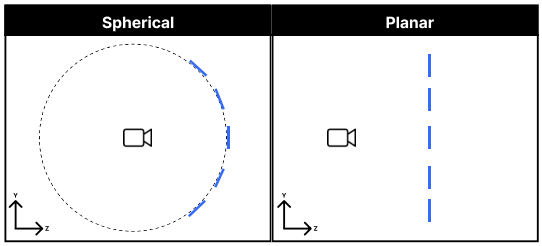

Fig. 4: Simple schematic showing targets tangent to sphere in Spherical geometry, vs flat on a plane in Planar geometry.

from the controller tip, of a fixed length of 0.5m was visible to users, visualizing the controller's forward direction.

Eye: We fire a raycast from the Tobii gaze origin point in the direction of the gaze direction vector provided. We provide no cursor or target feedback. Participants simply look at the target before selection.

Rationale of Cursor and Feedback Design: Each input cannot be fairly compared simply using the same feedback methods. We chose to maintain existing cursor and head/controller targeting and selection standards to establish a comparison relative to standards today. The thin white rod was used to replicate controller cursor and feedback designs used in highly adopted VR headsets (Meta Quest 2 or HTC Vive). A simple red dot was used to signify head cursor location, similar to many world locked cursors (HoloLens). A rod was not used in the case of head input, as stereoscopic parallax causes one to see double rods (unless one eye is closed [21]) which is uncomfortable and does not provide a fair comparison.

Cursor feedback as shown with the head or controller inputs cannot be simply replicated as feedback for the eye: Adding a similar cursor for eye trackers does not offer the same value, and can often be prone to the well known issue of cursor chase when accuracy offsets are present [29]. There are several other methodologies available aside from cursor feedback, each with unique advantages and disadvantages (highlighting, adding a border etc), but none are widely adopted in current AR/VR. Furthermore, given a certain eye tracking accuracy, we can expect that items will be targeted just by looking at them, within a certain accuracy bound. In contrast, one could be quite off target without a cursor for controller and head tracking. To avoid these risks and set a baseline to be built upon in future paradigms, it made the most sense to start with unmodified eye tracking rather than adding additional variables from visual feedback to our evaluation. Thus, we showed no visible cursor or visual feedback to the participant in the eye tracking condition. That said, it is important to note that implicit feedback - whether the target progression timed-out or continued after a button press - indicated when they successfully selected the target.

3.5.4 Target Progression

If the raycast intersects with the target and the user presses the Meta Quest 2 controller trigger, the target progresses immediately to the subsequent target/trial. If a user fails to hit and activate the target, the target will automatically advance to the next one in the sequence after a 5-second timeout. That is, targets progress after 5 seconds whether the user attempted to select it or not. Given the large number of trials (1098 targets in total) each user goes through, 5 seconds, which was longer than $> 99\%$ of all selections in pilot tests, would not unnaturally rush users.

4 Measurements

4.1 Quantitative Data

In addition to recording information about the conditions being run (Input Modality, Standard, Geometry, and Target Diameter), at a frame level, we recorded each target's position and rotation in world coordinates and dummy camera coordinates. We also recorded the position and rotation of the main camera (participant's head), the controller's position and rotation, relevant eye tracking data, and information relevant to transform these values into any of the coordinate systems described above.

We recorded every time a controller trigger press occurred and indicated whether it was a hit or miss, and by extension of the data mentioned previously, the metric and angular distance from the target to the corresponding cursor or interaction point (in the case of eye tracking).

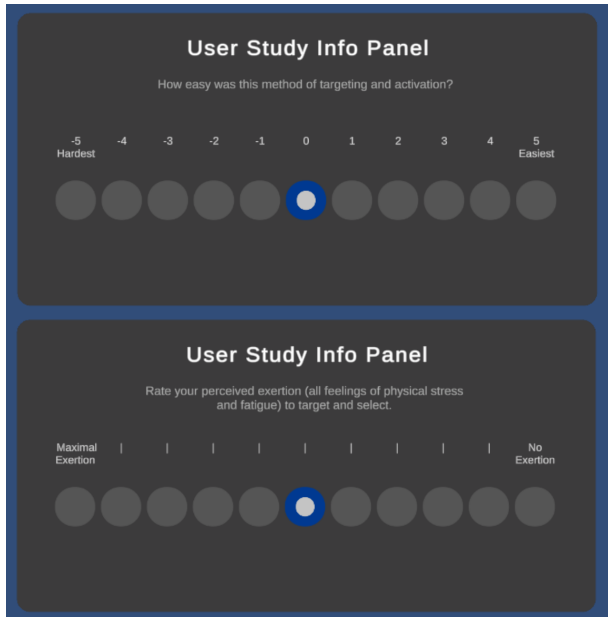

Fig. 5: Question 1 and 3 of the in-app Subjective Evaluation user interface shown after each block.

4.2 Subjective Evaluation

Following each of the 36 blocks, within the virtual environment, we prompted users to complete 3 questions displayed on a panel (see Fig. 5). We presented questions in a Likert scale format, with 11 potential responses, and default values were median values. Participants provided their input using the Meta Quest 2 controller, using the thumbstick to increase or decrease values and would press the trigger to select a value. A left or right thumbstick push would move the cursor one unit to the left or right. This means that a participant would have to push the thumbstick left or right 5 additional times to go to the minimum or maximum values. Although implemented for efficient user experience, this method did start the cursor at a potential response; however, it was a neutral score and the user had to exert more effort for extreme values, and so potentially underestimated the evaluation, but was not biased positively or negatively.

Q1 Ease of use: How easy was this method of targeting and activation? **A1**: -5 hardest, 5 easiest, Default value 0. All integers in between were visible.

Q2 Adoption: How much do you agree with the statement: "I am willing to use this as a method of targeting and selecting in an AR/VR device"? **A2**: -5 Strongly Disagree, 5 Strongly Agree, Default value 0. All integers in between were visible.

Q3 Fatigue: Rate your perceived exertion (all feelings of physical stress and fatigue) to target and select. **A3**: We presented 11 options with the left most element titled "Maximal Exertion" and the right most titled "No Exertion." Options in between were simply lines.

4.3 User Preferences, Experience and Feedback

Each participant had to fill out a questionnaire through Google Docs for demographic information and the following questions regarding their experience:

1. Rank your preferred methods of aiming at the objects (1 being the best, 3 being the worst): [Controller] [Head] [Eye Gaze]
2. Why did you prefer that method of aiming?

3. Did you use any particular strategy for carrying out each task, and if so please describe it briefly.

4. How many hours per week on average do you spend doing any sort of gaming (console, cell phone, etc.)?

5. How many hours per week on average do you spend in Virtual Reality (head-mounted display, such as Quest 2)?

6. Do you have any other observations, comments, or suggestions you would like to share?

5 Data Analysis

We used Fitts' Law to evaluate human performance across all conditions. This section will describe the parameters we used to calculate throughput in bits per second.

Movement Time: We calculated movement time (MT) as the time from when a target is displayed to when the user successfully selects that target. We ignore all misses. We use this value in our per-target/trial throughput calculations.

Index of Difficulty: We used the Shannon Formulation Fitts's Index of Difficulty for our calculations, most frequently used in targeting and selection tasks [28].

$$ID = \log_2(D/W + 1)$$

We realize that because we use world-locked targets, there are multiple possible approaches to calculate the distance to the target (D) and the target size (W), given that we have units in both meters and degrees. We also have targets positioned in world coordinates, dummy camera coordinates, or head coordinates. If we used degrees from the main camera's coordinate space, participants could move their heads from the start to the end of any particular trial/target. From the participant's perspective, this would change the angular distance between targets in degrees relative to the head, from when the target is displayed, to the time it is selected. Similarly, if the participant moves closer to the target during the trial period, that target's visual angle will also increase from the participant's point of view.

We reduced these ambiguities in calculating ID by using the dummy camera and the target positions relative to that camera in degrees. We chose this method because targets remain constant in this coordinate space throughout the trial. We also used the target size relative to the dummy camera in degrees. The actual size of the targets, if they were in planar coordinates, would be <= their intended size (3,4 or 5 degrees), away from the center, whereas this is not the case with spherical coordinates.

Note on Effective Index of Difficulty: In future studies, we would like to measure the performance of advanced eye tracking signal manipulation [45], with contextual information provided from the scene, or AI, and subsequent collider or cursor manipulation. The accuracy of the eye tracker and the specific (x-y) points of selection required to calculate the "Effective" Index of Difficulty, which is dependent on this, is less relevant as a metric. As the intended goal of this study is to set up a foundation to calculate the rate at which participants were able to progress through targets, rather than validate the accuracy of our eye tracker, we chose to omit using the Effective Index of Difficulty in our statistical analysis [49].

Throughput Calculations: Throughput in bits per second (TP), was calculated on an individual target/trial basis using ID and MT, calculated above. We only calculated throughput and considered MT on "Hits." Furthermore, as the first target of each block doesn't have a prior target to calculate ID, we also omitted the first target shown in each block from our calculation.

$$TP = ID/MT$$

Miss Calculations: We calculated miss percentages in the following manner: # miss trigger presses / (# miss trigger presses + # hit trigger presses) * 100, per block.

6 Research Hypotheses

Given that this is a baseline study, we expect that participants using unmodified eye input will not necessarily perform better than using the controller, but will perform similarly. Specifically:

H1: *Participants will have similar throughput (bits/s) and movement time (s) with controllers and eyes in both ISO and Random Web conditions. Eyes will perform better than the head throughout.* The eyes will lead the head and hand in selecting targets of unknown locations [9, 13, 15]. Higher quality eye tracking should be able to capture this, but given the limitations of our eye tracker, especially at fringe angles, we expect misses to slow down eye tracking throughput on average to that of controller tracking.

H2: *We expect eyes to have the most significant error rate, as we provide no form of feedback before selection. We expect 3 degree targets to have the most errors, specifically for eyes, with decreasing error rates as target size increases.* Misses are a function of calibration error. We expect that 3 degree targets (which represent a 1.5 degree error threshold), will be closer to the limits of our eye tracker's accuracy for participants in fringe angles. Furthermore, as we provide no feedback, we expect that users will press the controller trigger when not necessarily looking straight at the target.

H3: *We expect controller and eye to be similar in terms of preference for the Subjective Evaluation questions 1-3, as well as subjective questions provided in our survey. We expect participants to rate those two input modalities higher than head in all dimensions.* Continuous head movements require the most effort [1], especially when compared to 6-DOF wrist controller movements. The eyes require even less effort [27], but maintaining gaze on a target prior to a trigger press is slightly unnatural. The misses accrued with eye tracking can be tiring over time which is why we do not explicitly state the eye will be rated better than controller tracking in subjective evaluations.

H4: *We expect that participants will have a higher throughput for Double-ISO tasks than for Random-Web tasks, as participants can learn the pattern of ISO tasks over time.* We expect users to have learnt (implicitly or explicitly [30]) the pattern in Double-ISO tasks over time, and thus will naturally move to succeeding target locations.

H5: *We expect no difference in throughput in Planar vs. Spherical Geometry when comparing input modalities.* For the angular positions we are testing (max 15 degrees from the center) we will see that target angular sizes as viewed from the user remain similar despite target location and rotation towards users in spherical coordinates.

7 Eye Tracking Performance

All 30 participants successfully went through eye calibration and passed the criteria for calibration accuracy (noted below). Participants had their calibration quality tested before each eye tracking block. In other words, calibration quality was tested 12 times per participant, and they were allowed to recalibrate at any point during those tests.

During calibration quality tests, participants were shown 5 targets locked in display coordinates (as shown in Fig. 2). These consisted of 1 target positioned at the center of the display and 4 targets forming a cross, each 10 degrees from the center. On their first calibration quality check attempt, participants were required to have a mean calibration accuracy of < 1.5 degrees. All participants were able to meet this requirement.

If we were to average every single block a participant ran the calibration test, and then take the average of that number across all participants, we see that mean calibration error is 0.781 degrees, SD 0.348. Repeating for p50 error values, we have mean = 0.704, SD = 0.358. For p75 error values, we have mean = 0.942, SD = 0.466, and finally for p95 error values, we have mean = 1.588, SD = 0.811.

We recognize that these values represent the 10-degree radius covered by the calibration quality test, and we expect these accuracy values to fall off further away from the center.

8 Statistical Assumptions

We first confirmed that the data satisfied the assumptions of a repeated-measures analysis of variance (ANOVA). We used repeated-measures ANOVAs to compare the effects of our design variables in terms of the

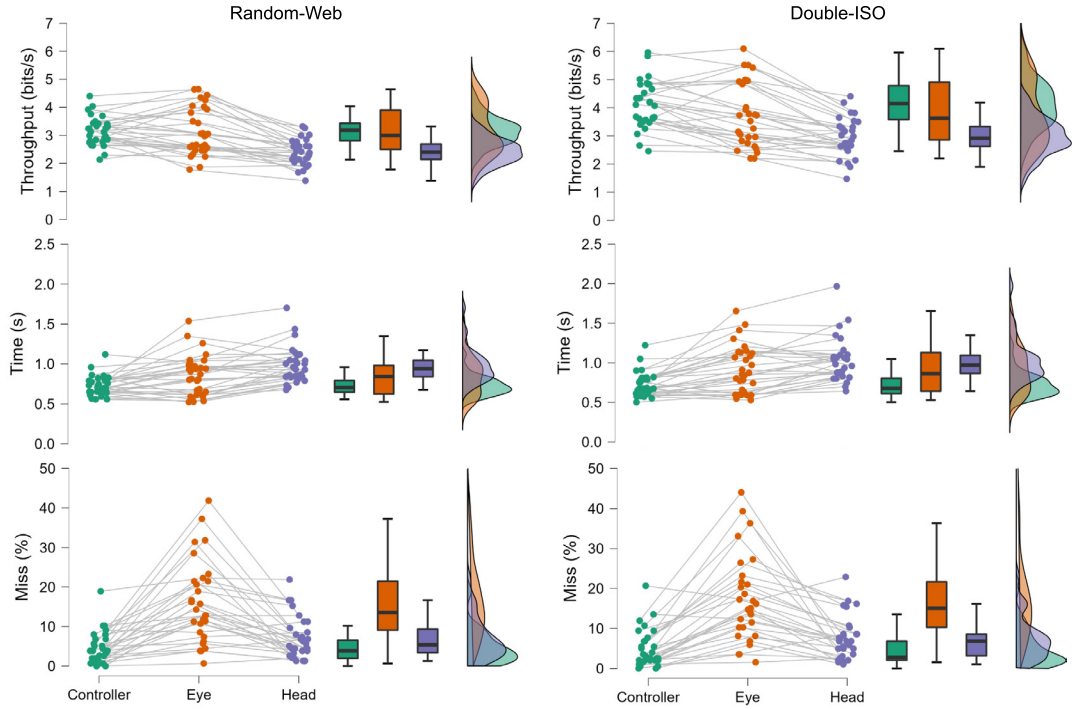

Fig. 6: Chart showing how individual participants compared across different input modalities and Standard target placement.

		Throughput (bits/s)		Movement Time (s)		Misses (%)	
Input	Standard	Mean	SD	Mean	SD	Mean	SD
Controller	Web	3.16	0.50	0.73	0.12	4.46	3.98
	ISO	4.12	0.82	0.72	0.15	4.86	4.67
Eye	Web	3.13	0.83	0.85	0.25	16.2	10.11
	ISO	3.74	1.12	0.92	0.30	18.38	12.1
Head	Web	2.41	0.45	0.98	0.22	7.02	5.23
	ISO	2.96	0.67	1.03	0.27	7.44	5.29

Table 1: Throughput, movement time and misses. N=30.

primary measures (throughput in terms of bits per second, movement time, and accuracy in terms of misses). ANOVA has been repeatedly found to be robust to violations of normality [5, 44], therefore that was not a primary concern; Shapiro-Wilk tests never showed a significant departure from normality across all conditions for the primary dependent variable, throughput ($p > .05$), and for most conditions for misses and response times. We provide visualizations of the distributions for our results. The survey data were also analyzed by repeated-measures ANOVAs, as theoretical developments [23, 51] and simulations [56] support the consideration of what might be termed ordinal data on interval scales. Note that there were some analyses where Mauchly's test of sphericity was significant, indicating that the assumption of sphericity was violated. In those instances, when ε was greater than 0.75 we then used the Huynh-Feldt correction, and when it was less then we used the Greenhouse-Geisser correction [14]; these cases are apparent due to the degrees of freedom not being integers. Paired t-tests with Holm correction were used for all pairwise comparisons between conditions of interest to correct for family-wise error rates. All tests for significance were made at the $\alpha = 0.05$ level. The figures provide raincloud plots [2] for extensive visualization of the data. **Timeouts:** We observed a total of 114 total timeouts across all participants, 98 (eye), 10 (head), 6 (controller). As this accounts for only 0.35% of total targets displayed for all participants, we decided to omit timeouts from our calculations, to simplify our analysis.

9 Results

The descriptive statistics for throughput, movement times and miss percentage are given in Table 1, and illustrated in Fig. 6.

9.1 Assessing Throughput and Movement Time

To assess **H1** and **H4**, we analyzed throughput as a function of condition. The full ANOVA results reveal that the interaction of Input Modality by Standard was statistically significant ($F(2,58) = 33.728$, $p < .001$, $\eta_p^2 = .538$) as were the main effects for Standard ($F(1,29) = 160.348$, $p < .001$, $\eta_p^2 = .847$) and Input Modality ($F(1.721,49.902) = 43.994$, $p < .001$, $\eta_p^2 = 603$). The contrasts between Input Modality in terms of throughput revealed that there was no significant difference between Eye and Controller inputs (mean difference of 0.208 in favor of Controller; $p = .058$), but Controller and Eye were both significantly different than Head (mean differences of 0.961 and 0.753, respectively, with Head the inferior input in both cases; $p < .001$). The impact of Input Modality and Standard presentations on movement time are largely consistent with throughput, with the exception that post-hoc tests revealed significant differences across all Input Modalities ($p < .003$). The results largely support **H1** in that Eye and Controller inputs were largely the same in terms of throughput and movement time, with both being significantly better than Head, and only different in terms of movement time. The analysis of throughput also supports **H4**, in that the Standard conditions (Double-ISO vs Random-Web) were significantly different in the ANOVA results, and with the ISO condition having a higher throughput than the Web condition (mean difference of 0.705; $p < .001$).

In assessment of **H5**, we found that there was no impact of Geometry (Planar versus Spherical) presentations on throughput (bits/s), as seen in the lack of any statistically significant results in the three-way interaction (Input Modality* Geometry* Standard), in the two-way interactions (Geometry * Standard; Input Modality* Geometry), and the main effect of Geometry (all $F < 1.0$ and $p > .05$). There was also no statistically significant impact of Geometry on movement time; however, there was a significant main effect on misses (planar: 9.71%; spherical: 8.97%; $F(1,29) = 4.864$, $p = .036$, $\eta_p^2 = .144$), but no significant interaction effects of Geometry on misses (all $F < 1.0$ and $p > .2$). Overall, this supports **H5**, that the presentation of stimuli in either spherical or planar coordinate does not impact targeting and selection in terms of throughput and movement time, and only a small difference in misses.

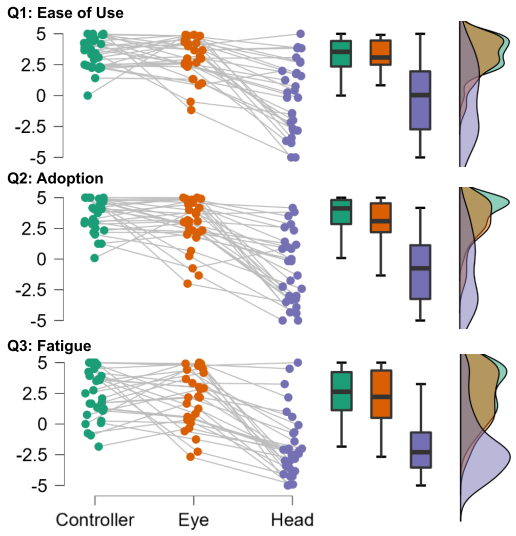

Fig. 7: Responses to Questions 1-3 of the Subjective Evaluation. Controller and Eye are comparable, and both have more favorable responses relative to head.

	Q1 Ease of use		Q2 Adoption		Q3 Fatigue	
Input	Mean	SD	Mean	SD	Mean	SD
Controller	3.42	1.23	3.65	1.34	2.50	2.02
Eye	3.07	1.60	2.88	1.90	2.11	2.25
Head	-0.15	2.84	-0.85	2.77	-1.71	2.61

Table 2: Subjective Evaluation results for the questions shown after each of 36 blocks with values ranging from [-5,5], N = 30.

9.2 Assessing Error Rates

The descriptive statistics for misses are also in Table 1, and illustrated in the lower panel of Fig. 6. We examined the misses (error rates) to assess **H2**, with a primary interest in Input Modality and Target Size, though the full repeated-measures ANOVA included Geometry and Standard as well. For misses, there were significant main effects of Input Modality ($F(1.151,33.392) = 33.528$, $p < .001$, $\eta_p^2 = .536$), Target Size ($F(2,58) = 71.528$, $p < .001$, $\eta_p^2 = .712$), Geometry (see paragraph on **H5**), and Standard ($F(1,29) = 5.933$, $p = .021$, $\eta_p^2 = .170$), but with no significant interactions (all $p > .2$) except for that of Input Modality as a function of Target Size ($F(2.947,86.258) = 10.45$, $p < .001$, $\eta_p^2 = .265$). In support of **H2**, the post hoc comparisons revealed that the Eye had more misses than the Controller (12.02, $p < .001$) and the Head (9.511, $p < .001$), but there was no significant difference between the Head and Controller (2.51, $p > .104$). Moreover, for Target Size, the 3 degree target had more misses than the 4 degree (3.91, $p < .001$) and 5 degree (5.37, $p < .001$) targets, and the 4 degree had more misses than the 5 degree, too (1.46, $p = .003$), and the interaction between Input Modality and Target Size is due to the greater number of misses for the Eye across all target sizes in comparison to the Controller ($p < .001$) and the Head ($p < .001$), but no significant difference between the Controller and Head ($p > .3$) for each paired target size.

In regards to **H4**, the Standard condition had significant differences noted for throughput, movement time, and misses, but perhaps not in the manner always expected. The notable exception is that there were more misses for the Double-ISO condition (9.816%) than the Random-Web condition (8.857%, $p = .023$). Yet, as anticipated by **H4**, the ISO outperformed Web conditions for throughput (3.605 and 2.901, respectively, $p < .001$) and movement time (.891s and .853s, respectively, $p = .003$), likely due to certainty of target placement that aids performance in speed, but not accuracy.

9.3 Survey and Qualitative Results

9.3.1 Subjective Evaluation

We examined preferences predicted in **H3** as a function of Input Modality after each block with the in-app questions and overall at the end of the experiment, with the results provided in Table 2 and illustrated in Fig. 7. First, considering the in-app questions, the first question was regarding how easy they found the method of targeting and selection. We used a repeated-measures ANOVA to assess the impact of Input Modality, Geometry, and Standard on responses. Only the main effect of Input Modality was statistically significant ($F(2,58) = 36.114$, $p < .001$, $\eta_p^2 = .524$). Post hoc tests revealed that there was no significant difference between Controller and Eye (mean difference 0.353, $p = .449$), but both were rated significantly easier than Head (mean difference compared to Controller 3.572 and compared to Eye 3.219, $p < .001$).

The second question queried how willing they were to use the method, and there were significant main effects of Input Modality ($F(2,58) = 49.434$, $p < .001$, $\eta_p^2 = .599$) and Standard ($F(1,29) = 5.075$, $p = .032$, $\eta_p^2 = .001$)). Post hoc tests revealed a similar pattern of results as the first question, with no significant difference between Controller and Eye (mean difference 0.769, $p = .118$), but both were rated significantly more willing to be used than Head (mean difference compared to Controller 4.506 and compared to Eye 3.736, $p < .001$). Interestingly for the Standard, the participants were more willing to use the Random-Web than the Double-ISO by a small but significant margin (mean difference 0.185, $p = .032$).

Finally, the third question measured the degree of exertion the participants felt they used for each method. Like the second question, there were significant main effects of Input Modality ($F(2,58) = 43.144$, $p < .001$, $\eta_p^2 = .560$) and Standard $F(1,29) = 8.484$, $p = .007$, $\eta_p^2 = .002$). Post hoc tests revealed a similar pattern of results as the other questions, with no significant difference between Controller and Eye (mean difference 0.386, $p = .444$), but both were rated significantly less exerting than Head (mean difference compared to Controller 4.208 and compared to Eye 3.822, $p < .001$). Again, interestingly for the Standard, the participants found less exertion for the Random-Web than the Double-ISO by a small but significant margin (mean difference 0.204, $p = .007$). Overall, these results support H3, that the participants would equally favor the Controller and Eye over the Head as an input.

9.3.2 User Preferences, Experience and Feedback

In the post-experimental survey (n = 29, as 1 did not complete the survey), participants were asked to rank each Input Modality as the Best, Middle, or Worst, and the Eye was found to be superior to the other two modalities, in support of **H3**. The Best ranking was given the most to the Eye (14), then the Controller (13), and finally the Head (2). The Middle ranking most frequently went to the Controller (16), then the Eye (12), and then the Head (1). The Head was most often rated as the worst (26), compared to the Eye (3), and none thought the Controller was the worst. In qualitative explanations for why the Eye was chosen as the best, participants noted: "Once I got used to it, it felt natural and easy;" "It was the quickest;" and "Easiest." The Controller was favored due to: "It's easy and very accurate" and "Accuracy and easier to move the wrist." Although the Head was generally not favored, it was noted that it was preferable due to: "High reliability" and "More accurate." Note that in this study, the Eye condition included natural head movements as there would be hardly any applications where a user's head would be immobilized; indeed this natural connection was noted by a participant: "I used a combination of moving the headset and my eyes. When the target was too low or high it would cause eye strain so I had to use my head to help out." Finally, we noted in our design decisions that, while a cursor is natural in use cases for Controller or Head targeting, it is a nuisance for Eye targeting; indeed, a participant noted this due to how the experiment begins with a cursor driven by unmodified eye-signal shown before calibration, but never shown again, that indicated the common issue of cursor-following behavior with gaze: "At the beginning I could see a dot following my eye movement. It was annoying because the dot would jump a lot."

10 Discussion

In this paper, our goal was to set the groundwork for evaluating eye tracking as an input and interaction modality. We designed an experiment comparing unmodified eye tracking with other established input modalities available in commercial headsets, particularly those using the controller or head, specifically for world-locked targeting and selecting tasks within AR/VR. We included several factors in our experimental design, including a Double-ISO or Random-Web target presentation standard, planar or spherical target geometry, and various target diameters. We intentionally did not modify the eye tracking signal provided through the device's API in any way.

Specifically, in targeting and selection tasks, a person's natural tendency is not to look at a target, maintain their gaze on that target, and then select. Rather, in natural gaze explorations, we tend to move our eyes quickly from one point of interest to the next (e.g., at a frequency of up to 3 times per second [26, 58]), and hand-eye coordination is also subject to different sensory-motor latencies [43], making it difficult for the user to estimate exactly when, relatively, gaze and hand events occur. However, as we expected in **H1**, we see that people perform similarly, though slightly worse, in throughput and movement time using unmodified eye tracking relative to the controller. Still, eye tracking consistently outperforms the head in the same measures. People also tend to have a similar preference for eye tracking and controllers, both being better than the head, in the in-app questions and our post-experiment survey, which is consistent with **H3**, and some past research that found eye tracking was not perceived as more effortful than using a controller [27]. However, eye tracking seems to perform relatively poorly with target misses.

We hypothesized eye tracking to have the worst selection error rate (**H2**), consistent with past work [18, 27, 39, 48]. We did not provide any online visual feedback to the participant for the eye tracking condition, and this is more so an issue when calibration errors are large, as targets of intent might not be selected. In contrast, participants had constant feedback on where they were pointing with their head or hand through the red cursor we provided. Indeed, if we removed the cursor from the controller and head targeting conditions, performance would likely suffer, because unlike eye tracking, users do not know where precisely they are pointing with their head or hand. Such a cursor will also not give the same effect if we drive it using raw gaze. Specifically, it is well-known that the availability of online cursor-style feedback in a gaze-interaction system will result in the phenomenon of "chasing," where users attempt to place their gaze on an eye tracked cursor with inherent errors. "Chasing" occurs because retinal position errors partially drive saccades [41]. This highly undesirable scenario results from the unexpected consequences of saccades in some UI feedback designs [39], as the visual system is evolved to expect objects in the world to remain stable, regardless of where the eye is pointed. However, we can avoid the negative effects of an erratically moving cursor by providing a visual indicator that the target is selected through an outline, color change, haptic feedback, or other similar means.

Because we chose to provide no online feedback to the participant (for reasons described in section 3.5.3), there was no way for them to know if the system indeed registered that they were looking at a target or not before selection. This issue of misses is only exacerbated when participants have poor calibration, targets are small, and select targets are in fringe angles. With no feedback, in this case, a missed selection due to system error is far more likely. If participants did not know that the system failed to recognize their intended target of choice, they would press the trigger with no success. Instead, participants were required to use their best guess on whether the eye tracker estimated their gaze to be on the selected target, increasing the probability of a measured "miss" and thereby increasing the overall miss percentage, and possibly also subjective frustration. This effect is akin to the scenario of flipping a coin, intending to press the button on "heads," but not knowing if it is "heads" or "tails" before selection. Appropriate online feedback that does not introduce new retinal motion, such as a color change or target highlight, lets the user know if the button press will be successful before selection. Some of our participants noticed this effect, e.g.: "Also, there isn't a targeting feedback in this case so I worry if I might make false clicks and not know what I clicked. Otherwise, it is still fairly intuitive and easy to use!"

Participants know they are missing the target when a target does not progress on a controller press. A question for future work is: If we provide users with visual feedback and they are shown that the system does not register their target of intent – how would this affect their subjective evaluation of eye tracking? One hypothesis is that it could reduce the experience for low-accuracy systems, as users now visually know the system is working poorly. However, the opposite could be true as it also reduces ambiguity from the user's perspective and allows them to correct their gaze until their target of intent is registered.

We expect providing feedback will reduce misses (participants will not press the trigger if they know the intended object is not selected). That said, any adaptation that the user might make to account for system error in such a scenario (or one without feedback), may induce visual fatigue due to unnatural eye movements. Indeed, some participants reported that they attempted to adjust their gaze: "For eye Gaze method, aiming target in the center of FOV is helpful, however in one session maybe due to miscalibration, I had to aim the target a little lower [than] the center" (sic). While providing feedback can aid participants in learning how to compensate for this error, this is a design trap. We should not be training participants to move their eyes unnaturally, but rather we should adapt the system to account for the natural use of gaze. Future work could avoid scenarios where users try to learn the system's limitations by eliminating their need to adjust for errors altogether. One example of this would be to adaptively modify the UI scheme, or visual elements of a scene to account for eye tracking calibration quality.

In support of **H4**, we saw better throughput and movement time performance in Double-ISO over the Random-Web Standard. This is likely because participants know where targets are positioned beforehand as they learn the pattern over trials. These two layouts represent tasks where users know the position of the intended element beforehand and others where they would have to navigate to an element at an unknown location. In support of **H5**, we saw that participants performed similarly in planar vs. spherical Geometries, at least for the range of target positions we provided. These results suggest that the spherical Geometry presented could be adopted in future similar experiments using similar target positions. However, as we increase the target position range to greater than 15 degrees away from the center of the dummy camera, we cannot conclude that this effect will hold, because the magnitudes of target depth differences increase with presentation eccentricity.

We use throughput and movement time as human performance measures, but beyond a certain threshold, it is not always necessary to optimize for them, and they are not necessarily indicative of what an application in AR/VR requires. In other words, having a higher throughput or lower movement time does not necessarily make an input method better. It depends on the context these measures are applied within. Most applications do not require users to progress through scenes and UIs non-stop at high throughputs with a known sequence (like in the ISO-style layout). In such scenarios, marginal increases to an already high throughput provide little to no additional value. For many system UI navigation tasks, we suspect that navigating at a throughput similar to our results for unmodified gaze will likely be sufficient.

In conclusion, eye tracking could be valuable in scenarios where portability and privacy are required, which might not be best suited for other inputs. Even out of the context of these scenarios, our post-experimental survey showed that participants still preferred unmodified eye tracking. Our novel experimental paradigm has helped lay the groundwork for improving eye tracking based inputs and interactions in AR and VR HMDs. We can expect eye tracking errors to reduce over time as investments in eye tracking technologies continue. Taken together, this study provides a strong indicator that eye tracking has tremendous potential, not just in its role in targeting and selection tasks but generally in reshaping future AR/VR interactions.

Acknowledgments

The authors wish to thank Joseph Zhang, Ken Koh, Duane Sawyer, Joel Shook, Carmen Wang, and Carina Thiemann for their contributions to this study.

REFERENCES

[1] J. Ahn, S. Choi, M. Lee, and K. Kim. Investigating key user experience factors for virtual reality interactions. *Journal of the Ergonomics Society of Korea*, 36(4):267–280, 2017. 6

[2] M. Allen, D. Poggiali, K. Whitaker, T. R. Marshall, and R. A. Kievit. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, 4, 2019. 7

[3] A. U. Batmaz, M. D. Barrera Machuca, J. Sun, and W. Stuerzlinger. The effect of the vergence-accommodation conflict on virtual hand pointing in immersive displays. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2022. 2

[4] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 585–592. IEEE, 2019. 2

[5] M. J. Blanca Mena, R. Alarcón Postigo, J. Arnau Gras, R. Bono Cabré, R. Bendayan, et al. Non-normal data: Is anova still a valid option? *Psicothema*, 2017. 7

[6] R. A. Bolt. Gaze-orchestrated dynamic windows. *ACM SIGGRAPH Computer Graphics*, 15(3):109–119, 1981. 1

[7] D. A. Bowman. Interaction techniques for immersive virtual environments: Design, evaluation, and application. *Journal of Visual Languages and Computing*, 10:37–53, 1998. 2

[8] L. D. Clark, A. B. Bhagat, and S. L. Riggs. Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies*, 139:102413, 2020. 2

[9] S. de Vries, R. Huys, and P. Zanone. Keeping your eye on the target: eye–hand coordination in a repetitive fitts' task. *Experimental Brain Research*, 236(12):3181–3190, 2018. 6

[10] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1118–1130, 2017. 3, 4

[11] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381, 1954. 2

[12] P. M. Fitts, R. E. Jones, and J. L. Milton. Eye fixations of aircraft pilots. iii. frequency, duration, and sequence fixations when flying air force ground-controlled approach system (gca). Technical report, Air Materiel Command Wright-Patterson AFB OH, 1949. 2

[13] J. H. Fuller. Head movement propensity. *Experimental Brain Research*, 92(1):152–164, 1992. 6

[14] E. R. Girden. *ANOVA: Repeated measures*. Number 84. sage, 1992. 7

[15] A. Gopal, S. Jana, and A. Murthy. Contrasting speed-accuracy tradeoffs for eye and hand movements reveal the optimal nature of saccade kinematics. *Journal of Neurophysiology*, 118(3):1664–1676, 2017. 6

[16] J. Hadnett-Hunter, G. Nicolaou, E. O'Neill, and M. Proulx. The effect of task on visual attention in interactive virtual environments. *ACM Transactions on Applied Perception (TAP)*, 16(3):1–17, 2019. 2

[17] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T.-H. Yu, C.-J. Tai, M. Akbay, Z. Wang, et al. Megatrack: monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graphics (ToG)*, 39(4):87–1, 2020. 1, 2

[18] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, pp. 1–5, 2018. 2, 9

[19] R. J. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)*, 9(2):152–169, 1991. 2

[20] R. J. Jacob. Eye tracking in advanced interface design. *Virtual Environments and Advanced Interface Design*, 258:288, 1995. 2

[21] R. Jota, M. A. Nacenta, J. A. Jorge, S. Carpendale, and S. Greenberg. A comparison of ray pointing techniques for very large displays. In *Proceedings of Graphics Interface 2010*, GI '10, p. 269–276. Canadian Information Processing Society, CAN, 2010. 5

[22] C. Kim, A. Klement, E. Park, J. Han, L. Rao, and J. Zhuang. 6-2: Invited paper: High-ppi fast-switch display development for oculus quest 2 vr headsets. In *SID Symposium Digest of Technical Papers*, vol. 53, pp. 40–43. Wiley Online Library, 2022. 3

[23] T. R. Knapp. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing Research*, 39(2):121–123, 1990. 7

[24] G. Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1912–1931, 2015. 4

[25] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018. 2

[26] M. F. Land. Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A*, 185(4):341–352, 1999. 9

[27] F. L. Luro and V. Sundstedt. A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pp. 1–9, 2019. 6, 9

[28] I. S. MacKenzie. A note on the information-theoretic basis for fitts' law. *Journal of Motor Behavior*, 21(3):323–330, 1989. 6

[29] P. Majaranta and K.-J. Räihä. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, pp. 15–22, 2002. 5

[30] G. P. McDonnell, M. Mills, L. McCuller, and M. D. Dodd. How does implicit learning of search regularities alter the manner in which you search? *Psychological Research*, 79(2):183–193, 2015. 6

[31] K. Minakata, J. P. Hansen, I. S. MacKenzie, P. Bækgaard, and V. Rajanna. Pointing by gaze, head, and foot in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pp. 1–9, 2019. 2

[32] D. Miniotas, O. Špakov, and I. S. MacKenzie. Eye gaze interaction with expanding targets. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pp. 1255–1258, 2004. 2

[33] A. Montagnini, P. Mamassian, L. Perrinet, E. Castet, and G. S. Masson. Bayesian modeling of dynamic motion integration. *Journal of Physiology-Paris*, 101(1-3):64–77, 2007. 2

[34] Y. S. Pai, T. Dingler, and K. Kunze. Assessing hands-free interactions for vr using eye gaze and electromyography. *Virtual Reality*, 23(2):119–131, 2019. 1

[35] K. Pfeuffer, J. Alexander, M. K. Chong, and H. Gellersen. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th annual ACM Symposium on User Interface Software and Technology*, pp. 509–518, 2014. 1

[36] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 36–39. IEEE, 2017. 2

[37] R. Plamondon and A. M. Alimi. Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, 20(2):279–303, 1997. 2

[38] M. J. Proulx and H. E. Egeth. Biased competition and visual search: the role of luminance and size contrast. *Psychological Research*, 72(1):106–113, 2008. 4

[39] Y. Y. Qian and R. J. Teather. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 91–98, 2017. 2, 4, 9

[40] K.-J. Räihä and O. Špakov. Disambiguating ninja cursors with eye gaze. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1411–1414, 2009. 1

[41] C. Rashbass. The relationship between saccadic and smooth tracking eye movements. *The Journal of Physiology*, 159(2):326, 1961. 2, 9

[42] D. A. Robinson. Eye movement control in primates: The oculomotor system contains specialized subsystems for acquiring and tracking visual targets. *Science*, 161(3847):1219–1224, 1968. 2

[43] U. Sailer, T. Eggert, J. Ditterich, and A. Straube. Spatial and temporal aspects of eye-hand coordination across different tasks. *Experimental Brain Research*, 134(2):163–173, 2000. 9

[44] E. Schmider, M. Ziegler, E. Danay, L. Beyer, and M. Bühner. Is it really robust? reinvestigating the robustness of anova against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(4):147, 2010. 7

[45] I. Schuetz and K. Fiehler. Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research*, 15(3), 2022. 6

[46] I. Schuetz, T. S. Murdison, K. J. MacKenzie, and M. Zannoli. An explana-

tion of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019. 2

[47] I. Schuetz, T. S. Murdison, and M. Zannoli. A psychophysics-inspired model of gaze selection performance. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–5, 2020. 2

[48] L. Sidenmark and H. Gellersen. Eye&head: Synergetic eye and head movement for gaze pointing and selection. In *Proceedings of the 32nd annual ACM Symposium on User Interface Software and Technology*, pp. 1161–1174, 2019. 2, 9

[49] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International Journal of Human-Computer Studies*, 61(6):751–789, 2004. 2, 6

[50] L. Thaler, A. C. Schütz, M. A. Goodale, and K. R. Gegenfurtner. What is the best fixation target? the effect of target shape on stability of fixational eye movements. *Vision Research*, 76:31–42, 2013. 4

[51] P. F. Velleman and L. Wilkinson. Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 47(1):65–72, 1993. 7

[52] R. Wang, S. Paris, and J. Popović. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pp. 549–558, 2011. 1, 2

[53] R. Y. Wang. Real-time hand-tracking as a user input device. In *ACM Symposium on User Interface Software and Technology (UIST)*, 2008. 1, 2

[54] R. Y. Wang and J. Popović. Real-time hand-tracking with a color glove. *ACM Transactions on Graphics (TOG)*, 28(3):1–8, 2009. 1, 2

[55] C. Ware and H. H. Mikaelian. An evaluation of an eye tracker as a device for computer input. In *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, pp. 183–188, 1986. 2

[56] H. Wu and S.-O. Leung. Can likert scales be treated as interval scales?—a simulation study. *Journal of Social Service Research*, 43(4):527–532, 2017. 7

[57] X. Wu and T. S. Murdison. 68-2: Considering the effects of display persistence on eye movements and readability in virtual reality. In *SID Symposium Digest of Technical Papers*, vol. 53, pp. 910–913. Wiley Online Library, 2022. 2

[58] A. L. Yarbus. Eye movements during perception of complex objects. In *Eye movements and vision*, pp. 171–211. Springer, 1967. 9

[59] X. Zhang and I. S. MacKenzie. Evaluating eye tracking with iso 9241-part 9. In *International Conference on Human-Computer Interaction*, pp. 779–788. Springer, 2007. 2