Undergraduate Thesis CPS 40A/B Project Final Report

Department of Computer Science

Toronto Metropolitan University

Comparing the Efficacy of Pure Hand Tracking Interaction and General Physical Manipulative Interaction in Mixed Reality Virtual Object Manipulation Tasks

Submitted By Idara Umoren

Supervised By Dr. Tim McInerney

June 30, 2026

ABSTRACT

Mixed reality headsets support a variety of mechanisms for controlling virtual objects. One common method is hand tracking, in which the headset's cameras estimate the position and pose of the user's hands, allowing users to interact with virtual objects directly. Another approach is the use of physical manipulatives, where a tracked physical object serves as a proxy for a virtual object and provides continuous tactile and proprioceptive feedback. These physical proxies may not need to match the exact shape of the virtual object for tactile feedback to be useful. This study compared hand tracking and a tracked physical manipulative in a within-subjects user study involving a virtual block-matching task. Both objective measures, including completion time, and subjective measures, including NASA-TLX, SUS, custom ratings, and open-ended responses, were collected. The physical manipulative was faster for seven of eight participants and received stronger subjective ratings for workload, usability, naturalness, and rotation accuracy. Within the limits of this small exploratory study, the results suggest that simple reusable physical proxies can improve performance and user experience in virtual manipulation tasks.

CHAPTER 1: INTRODUCTION

1.1 Background and Motivation

Mixed reality systems are becoming an increasingly common part of the consumer computing landscape. Modern headsets such as the Meta Quest 3 can present virtual objects overlaid on a view of the real room, track the user's head and hands without any external equipment, and render fully virtual environments, all in the same device. As these systems reach a wider audience, the question of how people should interact with the virtual objects inside them becomes increasingly important. Among the tasks a user might perform, manipulating a virtual object, moving it, turning it, and placing it where it belongs, is one of the most fundamental. This core function serves as the framework for applications as varied as assembly training, product design, education, and simulation.

The task of 3D object manipulation can be approached in multiple ways. The first is markerless hand tracking, in which the headset's cameras estimate the pose of the user's hands and let them reach out and manipulate virtual objects directly, with no controller or external device. This is appealing because it is immediate and requires nothing to be held or learned, and it is widely regarded as a natural way to interact with virtual content. The second approach is to give the user a physical object to hold, a tangible proxy, whose position and orientation are tracked and applied to a virtual object. The idea is that holding a real object provides a continuous sense of where it is and how it is oriented through touch and the body's own sense of position, increasing user performance.

This thesis investigates the strengths and weaknesses of both interaction conditions across several criteria; from there, user performance and preference are used to determine the optimal schema.

1.2 Thesis Outline

The remainder of this thesis is organised as follows. Chapter 2 reviews the background and related work, three-dimensional object manipulation, hand tracking, tangible interfaces and physical proxies, passive haptic feedback, and the prior comparative studies that most directly motivate this work. Chapter 3 describes the research design, the participants, the procedure, the measures collected, and the methods of analysis. Chapter 4 presents the results, covering task completion time and its associated learning and shape-level analyses, perceived workload, usability, the custom experience ratings, the preference and comparative judgements, and the themes drawn from participants' open-ended responses. Chapter 5 discusses and interprets these results within the context of prior work while considering the study's limitations.

CHAPTER 2: BACKGROUND AND RELATED WORK

This chapter reviews preexisting research surrounding the present study and builds towards the inquiry. Specifically, it concerns literature involving 3D object manipulation, markerless hand tracking, tangible physical proxies, and passive haptic feedback. These pieces serve as a foundation allowing the current study to delve further, examining the efficacy of general tracked physical proxies in orientation matching tasks compared to pure hand tracking.

2.1: Evaluating 3D Object Manipulation

Object manipulation is a fundamental task in immersive environments. Influencing a virtual object requires some combination of user control over position, orientation, scale, selection, and release. Bowman and Hodges argued that interaction techniques for virtual environments should be evaluated systematically; the validity of an interaction method should be based on its efficacy in achieving its designated task [1]. This is especially important as an interaction technique that is fast may still be undesirable if it causes fatigue, frustration, confusion, or low perceived control. For this reason, the study combines objective timing data with subjective evaluation. Completion time captures how efficiently participants completed the matching task, while workload, usability, perceived control, and naturalness capture aspects of the experience that time alone cannot explain.

2.2: Hand Tracking and Mid-Air Interaction

Markerless hand tracking has become a major input method for mixed reality headsets. Instead of requiring additional hardware, camera-based hand tracking estimates the position and pose of the user's hands directly. This is appealing for mixed reality as it allows users to interact with virtual content similarly to how they would with physical objects in their everyday life. Nguyen, Gouin-Vallerand, and Amiri describe hand interaction as a central design area for mixed reality head-mounted displays, partly because direct hand interaction is often perceived as natural and intuitive [5].

Hand tracking, however, also has limitations. As the system estimates hand pose through cameras and software, its accuracy depends on lighting, occlusion, and the headset's interpretation of the user's movement. Abdlkarim and colleagues evaluated Meta Quest 2 hand tracking against optical motion capture and reported measurable positional and angular errors [7]. Although their results were measured on the Quest 2 rather than the Quest 3 used in this thesis, they illustrate an important issue for this interaction method; hand tracking can feel natural while still introducing instability. For precise orientation matching, these limitations may affect how confidently users can transform a block to their intended pose.

A second limitation is that users receive no physical feedback from the virtual object. They can see the object move, but they cannot feel its edges, weight, orientation, or resistance. This may make precise rotation more difficult as the user must rely almost entirely on visual feedback.

2.3: Tangible Interfaces, Physical Proxies, and Passive Haptics

Passive haptics refers to the use of real physical objects to provide tactile and proprioceptive feedback for virtual objects. In this context, the physical manipulative gives the user additional sensory information through touch and body position. Ishii and Ullmer argued that digital information can be given physical form by linking virtual elements to graspable objects [2]. Insko showed that adding passive haptic objects to virtual environments can improve presence and task performance [6]. The core idea is that a user holding a real object receives continuous information from the hand and body about the object's orientation and movement, stimulating multiple senses, potentially improving manipulation performance.

2.4: The Problem of Proxy Mismatch

A major challenge for tangible and passive-haptic systems is scalability. If every virtual object required an identical physical replica, the approach would be impractical. Prior work suggests that exact physical correspondence may not always be necessary. Azmandian and colleagues introduced haptic retargeting, a technique for reusing physical props across multiple virtual objects by exploiting the user's tolerance for certain visual-proprioceptive discrepancies [3]. Cheng and colleagues similarly proposed the Sparse Haptic Proxy, in which simple physical geometry can provide useful touch feedback for more complex virtual environments [4]. These works show that passive haptic systems can be designed around reusable or approximate physical forms rather than requiring a complete physical replica of every virtual object. This idea is directly relevant to the present study. The physical manipulative condition asks whether a simple tracked shape can still provide useful feedback despite differing in shape from the virtual object.

2.5 Measuring Workload and Usability

Because immersive interaction techniques can differ in both performance and experience, this study uses both objective and subjective measures. Completion time provides an objective measure of task efficiency. Subjective workload is measured using the NASA Task Load Index, which evaluates perceived mental demand, physical demand, temporal demand, performance, effort, and frustration [8]. Usability is measured using the System Usability Scale, a ten-item questionnaire that produces a composite usability score from 0 to 100 [9]. Bangor, Kortum, and Miller's later work provides commonly used interpretive benchmarks for SUS scores [10]. These instruments are appropriate because the study is not only concerned with whether participants can complete the task, but also with how demanding, usable, natural, and controllable each interaction method feels.

2.6 Research Gap

The literature reviewed in this chapter supports the focus of the present study. Immersive object manipulation is a core interaction problem with multiple solutions. Hand tracking offers a natural and controller-free way to manipulate virtual objects, but it may be limited by tracking instability and reliance on visual information alone. Tangible interfaces and passive haptics offer an alternative by grounding virtual manipulation in a physical object that the user can grasp and feel. At the same time, exact physical replicas are not scalable, making approximate or reusable proxies an important design direction. As such, it must be investigated whether the proprioceptive benefits of general physical manipulatives still exist; and if they can increase user performance compared to hand tracking manipulation.

CHAPTER 3: METHODOLOGY

This chapter describes how the study was conducted. It states the research questions, the research design, describes the participants, and details the procedure. As the study compares two fully implemented systems rather than a single isolated variable, this chapter is also explicit about the design choices and constraints that shape how the results should be read.

3.1 Research Design

The study used a within-subjects design in which every participant completed the same orientation-matching task under two interaction conditions. In the hand-tracking condition, participants manipulated virtual blocks directly with their hands using the headset's camera-based hand tracking. In the physical-manipulative condition, participants manipulated the virtual blocks by holding and moving a tracked physical prism. The input method was the variable of interest.

To guard against effects such as learning and fatigue, the order of the two conditions was counterbalanced. Four participants completed the hand-tracking condition first and the physical-manipulative condition second (Order A), and the other four completed the conditions in the reverse order (Order B). Participants were assigned to the two orders alternately as they were recruited, producing an even four-and-four split.

3.2 Participants

Eight participants took part in the study (mean age 21.4 years, standard deviation 0.9, range 20 to 23), recruited from Toronto Metropolitan University. Seven were right-handed and one was left-handed. On a five-point self-report scale, participants reported high prior experience with virtual or mixed reality headsets (mean 4.25), frequent video-game play (mean 4.38), and high confidence in mentally visualising and rotating 3D objects (mean 4.75). One participant reported low gaming experience, and the remainder reported moderate to high experience. No participant reported any condition affecting hand movement or grip. Participants were identified by an anonymised participant number. All participants were seated at a desk for the entire study.

3.3 Apparatus

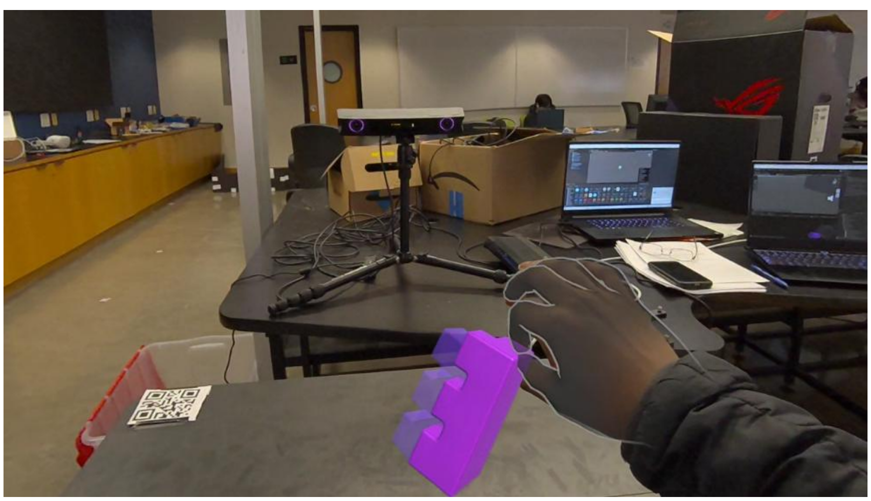
Both conditions ran on the same Meta Quest 3 headset and were built in the Unity game engine. Rather than being deployed as standalone headset applications, each was run from the Unity editor in Play mode and streamed to the headset over a wired Quest Link connection. The two applications resided on two different host computers, and the headset's link cable was moved from one computer to the other when switching between conditions. The two systems are described separately below.

[Figure 3.1: Image of Apparatus; OptiTrack Duo 3 Camera, Workspace, and Physical Manipulative Presented]

3.4 Hand-Tracking Condition

The hand-tracking application was developed by the author. Hand input was provided by the headset's built-in markerless hand tracking, accessed through the Meta Interaction SDK within the Meta All-in-One SDK. Participants interacted with the virtual blocks using direct hand manipulation. A pinch gesture was used to take hold of a block, after which the block followed the hand, and both hands could be used together to control its orientation. As the blocks were not subject to gravity, a participant could release a block at any time and it would remain floating in place, allowing them to reposition their hands and grasp it again to continue a rotation. Spatial anchoring of the virtual content to the physical environment was handled using a QR-code marker through the Meta MR Utility Kit, so that the virtual workspace remained registered to the real desk.

[Figure 3.2: Image of Hand Tracking Condition]

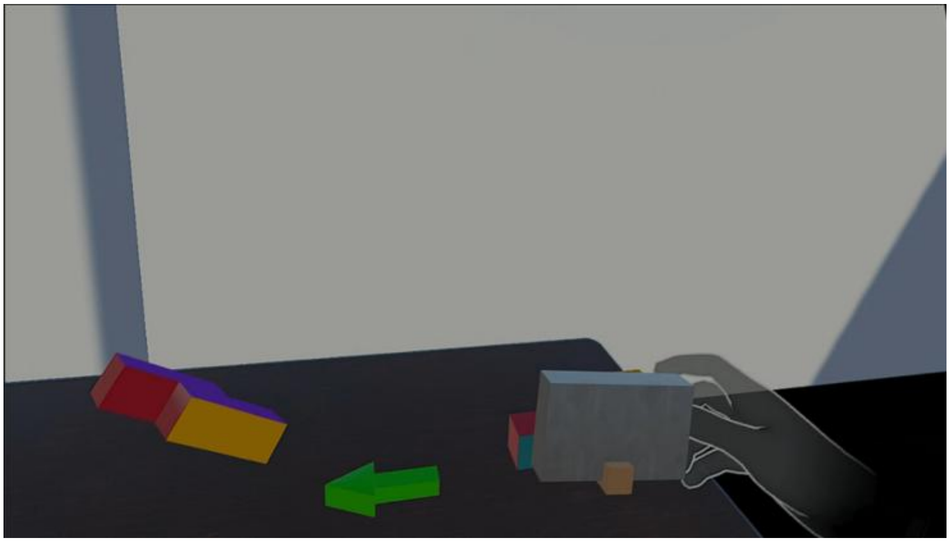


3.5 Physical-Manipulative Condition

The physical-manipulative application was developed by the project supervisor. The physical input device was a flat rectangular prism measuring roughly 7.5 by 5 by 1.5 centimetres, fitted with retroreflective markers and tracked as a rigid body by an OptiTrack Duo 3 motion-capture camera. The camera was mounted on a tripod resting on a table and angled downward to cover the working area, an approximately 80 by 50 centimetre region of the desk surface in front of the seated participant. The tracked pose of the physical prism was streamed from the OptiTrack Motive software into Unity using Motive's built-in streaming and was mapped directly and continuously onto the pose of the virtual block, with no snapping or smoothing applied. The motion-capture system was calibrated before the study began and was left untouched for its duration, so no per-participant calibration was performed.

Since the physical prism's pose drives the virtual block in real time, it functions as a graspable physical object whose position and orientation directly control a virtual one. Beyond this, a single rectangular physical prism is used to control several virtual shapes, thus providing approximate tactile feedback rather than a perfect physical duplicate. Both factors are discussed in the Chapter 2.

[Figure 3.3: Image of Physical Manipulative Condition]



3.6 Virtual Environment and Task

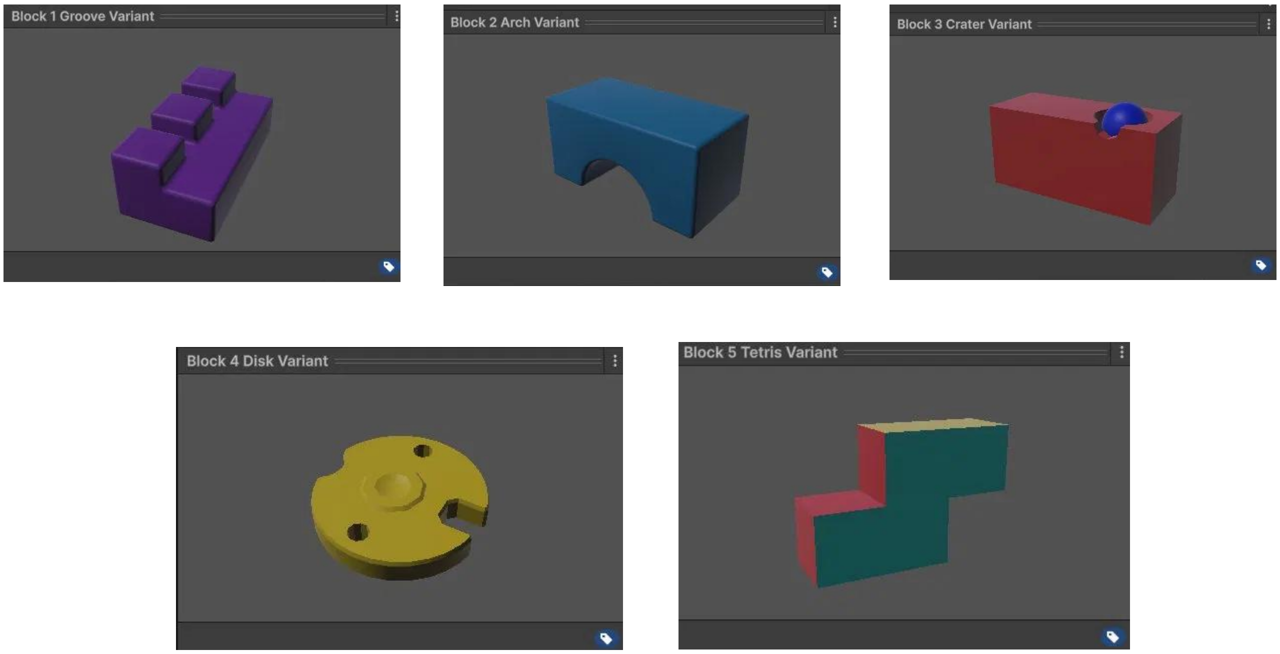
The task was the same in both conditions, to match the position and orientation of one virtual block to those of another. Each trial, referred to here as a level, presented two identically shaped virtual blocks. A start block appeared on the participant's right-hand side, and an end block appeared toward the centre of their seated field of view, both within comfortable arm's reach inside the working area on the desk. The two blocks were identical in shape but differed in position and orientation, and the participant's goal was to make the start block coincide with the end block.

The two conditions differed in how the start block was brought under the participant's control. In the physical-manipulative condition, the participant first touched the tracked prism to the start block, at which point the virtual block being controlled took on the shape of that start block; from then on, moving and rotating the physical prism moved and rotated the virtual block, and the participant manipulated it until it matched the end block. In the hand-tracking condition, the participant simply reached out, took hold of the start block with their hands, and manipulated it directly until it matched the end block. When a sufficiently close match was achieved, the level was registered as complete and the next level began.

A match was registered when the controlled block's position was within five centimetres of the target and its orientation was within five degrees of the target. Orientation was evaluated by comparing the block's local up and right axis vectors against the corresponding vectors of the target, with the five-degree tolerance implemented as a cosine (dot-product) threshold of approximately 0.9962 on each axis. For most shapes, alignment was required on both axes, so that the block had to be brought into the target's orientation in full. One shape is an exception for technical reasons described below.

Five distinct block shapes were used, chosen to vary in how easily their orientation could be read visually. They were as follows: a grooved rectangular prism, an arch with bilateral symmetry, a crater-like block combining a recessed prism and a ball, a disk with small cut-out holes, and a Z-shaped Tetris-like block. The arch’s symmetry allowed for anti-parallel matching of the right vector. Each of the five shapes was presented twice, giving ten recorded levels per condition. These ten recorded levels were preceded by three practice levels, which were not included in the analysis, and which allowed the participant to become familiar with the controls of the condition they were about to perform.

[Figure 3.4: The collection of matching blocks. From top left to bottom right: Grooved Prism, Arch, Crater Prism, Disk, and Tetris Block]



A characteristic of task design must be noted as it may have a bearing on interpretation. The order in which shapes were presented differed between the two conditions. The hand-tracking condition presented the shapes in a fixed sequence, whereas the physical-manipulative condition presented them in a randomised order. As a result, the two conditions are best compared at the level of the whole session rather than shape by shape across conditions. This is revisited as a limitation in the discussion chapter.

3.7 Procedure

Each participant completed the study in a single seated session. Before any trials began, the participant completed the pre-study section of the questionnaire booklet, which recorded their age, handedness, prior experience, and any relevant physical conditions.

The participant then put on the Meta Quest 3 headset, and they were given verbal instructions explaining the matching task and the controls for the first condition. The participant completed three practice levels to become comfortable with the interaction, after which the ten recorded levels of the first condition were carried out. Immediately after finishing the first condition, the participant removed the headset and completed the post-condition section of the booklet for that condition, which comprised the NASA Task Load Index, the System Usability Scale, the custom experience items, and the open-response questions.

The headset's link cable was then moved to the second host computer to load the second condition's application. The participant put the headset back on, received verbal instructions for the second condition, completed three further practice levels, and then carried out the ten recorded levels of the second condition. They then removed the headset and completed the post-condition section of the booklet for the second condition. Finally, having now experienced both conditions, the participant completed the post-study comparative section of the booklet, which asked them to compare the two conditions directly through forced-choice preference questions, comparative rating scales, and open-response questions. Each session lasted approximately eight to ten minutes.

3.8 Measures

The study collected one objective measure and a set of subjective measures. The objective measure was task completion time, recorded for each condition as the time taken to complete the ten recorded levels. This serves as an index of how efficiently the participant could perform the matching task with each input method.

The subjective measures were collected through the questionnaire booklet. Perceived workload was measured with the NASA Task Load Index across its six subscales of mental demand, physical demand, temporal demand, performance, effort, and frustration, each rated from 0 to 100. Usability was measured with the ten-item System Usability Scale, scored in the standard way to yield a composite value from 0 to 100. In addition, five custom experience statements were rated on seven-point agreement scales, covering the participant's sense of control over the block's orientation, how natural and intuitive the control method felt, how easily they could tell when the blocks were aligned, whether physical fatigue affected their performance, and how accurately they could rotate the block to their intended orientation. The post-study comparative section then collected five forced-choice preference selections between the two conditions, five comparative rating statements scored on seven-point scales, and open-ended responses on what participants found most challenging, most natural, and most frustrating about each condition, together with any further comments.

3.9 Data Analysis

The objective and rating data were analysed using paired comparisons, reflecting the within-subjects design in which each participant contributed a value under each condition. The distribution of the paired differences was first examined for normality using the Shapiro-Wilk test. Where the assumption of normality was reasonable, a paired-samples t-test was used as the primary comparison, with the Wilcoxon signed-rank test reported alongside it as a distribution-free robustness check. Where normality was doubtful, the Wilcoxon signed-rank test was treated as the primary result. Effect sizes were reported using Cohen's d to convey the magnitude of differences independently of the small sample size, since with eight participants a real and substantial difference may not reach conventional statistical significance.

The open-ended responses were analysed qualitatively through a thematic approach. Responses were read across all participants and grouped into recurring themes relating to each condition, such as difficulty with rotation under hand tracking, the value of tactile grounding under the physical manipulative, and tracking instability. These themes are used in the results and discussion chapters to interpret and contextualise the quantitative findings, and representative participant comments are quoted to illustrate them.

CHAPTER 4: RESULTS

This chapter reports the results of the study. It begins with the objective measure of task completion time, including an analysis of learning across the session and a shape-level breakdown of the hand-tracking condition. It then reports the subjective measures in turn: perceived workload from the NASA Task Load Index, usability from the System Usability Scale, the custom experience ratings, and the post-study preference and comparative judgements. The chapter closes with the themes that emerged from participants' open-ended responses. Throughout, results are reported for all eight participants, and because the sample is small, exact p-values and effect sizes are given alongside the descriptive statistics rather than relying on significance thresholds alone.

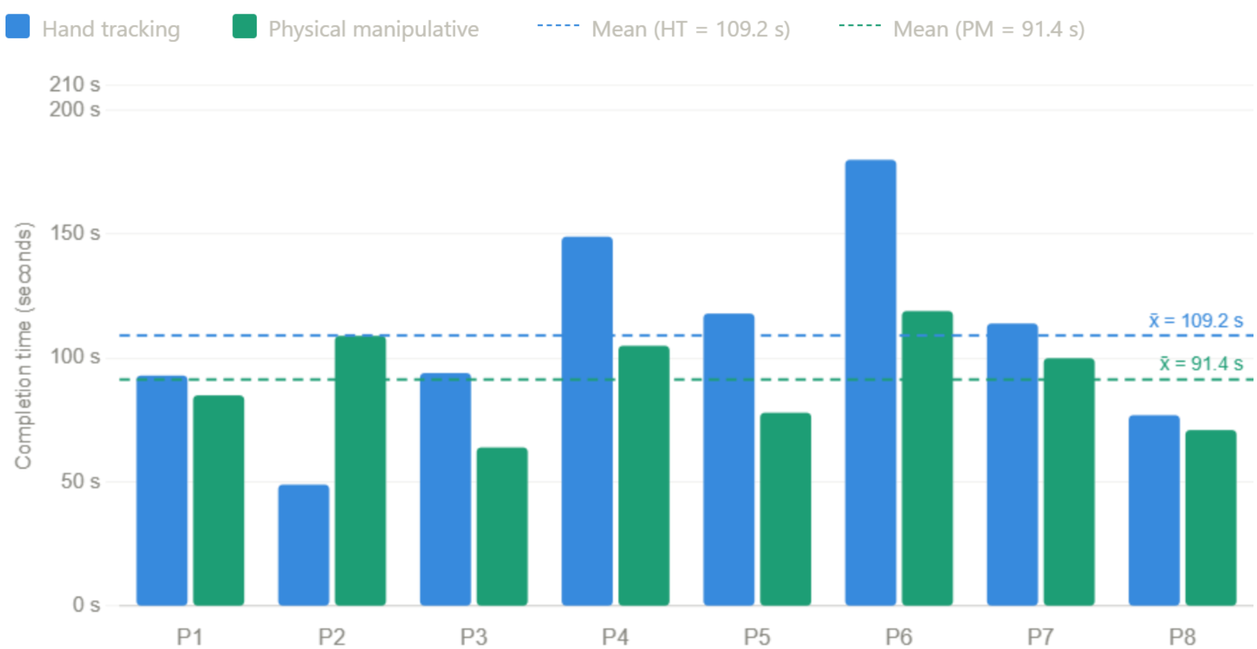
4.1 Task Completion Time

Task completion time was recorded as the total time each participant took to complete the ten recorded levels in each condition. Seven of the eight participants completed the task faster with the physical manipulative than with hand tracking. The mean session time was 109.2 seconds for hand tracking (SD 41.1) and 91.4 seconds for the physical manipulative (SD 19.7), a difference of 17.9 seconds, meaning that participants were on average about 16 percent faster with the physical manipulative. The hand-tracking condition was also far more variable than the physical-manipulative condition, with roughly twice the standard deviation, indicating that hand tracking worked well for some participants and poorly for others, whereas the physical manipulative produced more consistent times across the group.

[Table 4.1: Session completion times per participant]

Participant	Hand Tracking (s)	Physical Manipulatives (s)	Faster
P1	93	85	PM
P2	49	109	HT
P3	94	64	PM
P4	149	105	PM
P5	118	78	PM
P6	180	119	PM
P7	114	100	PM
P8	77	71	PM
Mean	109.2	91.4	
SD	41.1	19.7	

[Figure 4.1: Grouped bar chart of session completion time per participant, hand tracking versus physical manipulative]



Statistical significance was evaluated using an alpha level of 0.05. The Shapiro-Wilk test was used to assess whether the completion-time differences were approximately normally distributed, with $p > 0.05$ indicating no significant departure from normality. Effect size was interpreted using Cohen's conventional benchmarks, where d = 0.2 represents a small effect, d = 0.5 a medium effect, and d = 0.8 a large effect.

The Shapiro-Wilk test did not indicate a significant departure from normality, $W = 0.886$, $p = 0.216$, so a paired-samples t-test was used for the primary analysis. The difference in completion time was not statistically significant, $t(7) = 1.37$, $p = 0.212$. However, Cohen's d = 0.49 suggests a small-to-medium, approximately medium, effect size, indicating a possible practical difference in execution speed despite the small sample size. A Wilcoxon signed-rank test produced the same overall conclusion, $W = 7.0$, $p = 0.148$.

P2 was the single participant who ran counter the group. Having completed the hand tracking in 49 seconds and physical manipulatives in 109 seconds, they were the only participant to complete the hand tracking condition faster. Their per-level data that their slow physical-manipulative session was largely driven by a single thirty-second level, an isolated spike likely caused by momentary loss of tracking when a hand occludes the marks from the camera, a failure that several participants referred to in their written comments.

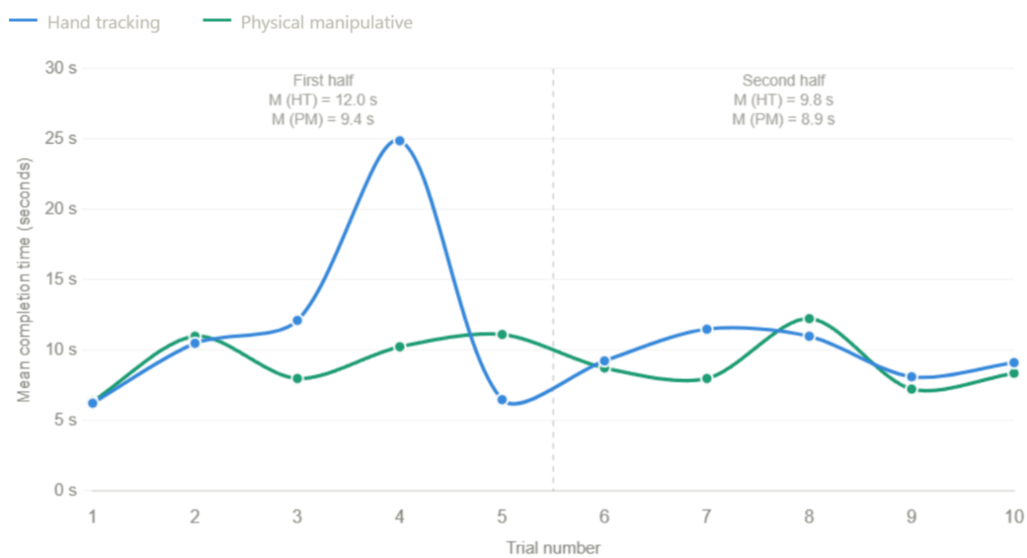
[Table 4.2: Per-Level Completion Times of P2]

Level	Hand Tracking (s)	Physical Manipulatives (s)
1	4	10
2	7	6
3	6	6
4	7	9
5	4	9
6	4	11
7	6	11
8	5	30
9	4	11
10	2	6
Mean	4.9	10.9
SD	1.66	7.05

Another sensitivity study was conducted excluding P2, resulting in mean times of 117.9 seconds for hand tracking against 88.9 seconds for physical manipulatives. The t-test gives $t(6) = 3.71$, $p = 0.010$, with a large effect size of Cohen's $d = 1.40$. P2 was retained in all future analysis, but this sensitivity check shows that the overall timing result is carried by a broad pattern across participants.

A learning effect was examined by comparing the mean time of each participant's first five levels with that of their last five. The two conditions behaved differently. In the hand-tracking condition, participants improved significantly over the session, from a mean of 12.0 seconds per level in the first half to 9.8 seconds in the second half, $t(7) = 3.71$, $p = 0.008$. In the physical-manipulative condition there was no meaningful change, from 9.4 seconds to 8.9 seconds, $p = 0.746$. In other words, hand tracking rewarded practice, with participants becoming noticeably more efficient as they adapted to gripping, releasing, and reorienting virtual objects in space, whereas the physical manipulative was performed at a consistent speed from the very first levels. This suggests that the physical manipulative was more immediately intuitive and required little adaptation, while hand tracking imposed an initial learning cost.

[Figure 4.3: Line chart of mean completion time across the first-half and second-half of trials for each condition.]



Because the hand-tracking condition presented the five shapes in a fixed order, with each shape appearing once in the first half of the session and once in the second, it is possible to examine performance by shape for that condition. The shape-level pattern for hand tracking is revealing. The Disk, the shape closest to rotationally symmetric and therefore the hardest to orient by sight alone, was by a wide margin the most time-consuming shape on first exposure, at a mean of 24.9 seconds, more than double any other shape's first-exposure time. On its second appearance later in the session, the Disk dropped to 11.0 seconds, an improvement of 56 percent and the largest improvement of any shape. The Crater, a strongly asymmetric shape whose orientation is easy to read visually, was completed comparatively quickly and improved modestly on second exposure, from 12.1 to 9.1 seconds.

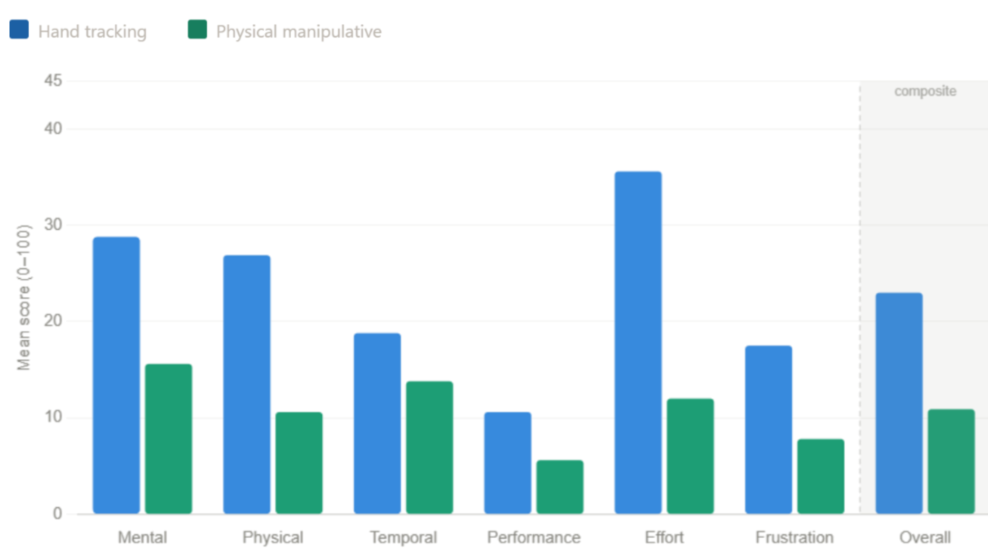
4.2 Perceived Workload (NASA-TLX)

Perceived workload was lower for the physical manipulative on every one of the six NASA-TLX subscales. The overall workload score, computed as the mean of the six subscales, was 23.0 for hand tracking and 10.9 for the physical manipulative, a large and statistically significant difference, t(7) = 3.09, p = 0.013, Cohen's d = 1.16. In other words, participants experienced hand tracking as roughly twice as demanding overall as the physical manipulative.

[Table 4.3: NASA-TLX subscale means (0-100), hand tracking versus physical manipulative]

Subscale	Hand Tracking	Physical Manipulation	p
Mental Demand	28.8	15.6	0.087
Physical Demand	26.9	10.6	0.104
Temporal Demand	18.8	13.8	0.413
Performance	10.6	5.6	0.200
Effort	35.6	12.0	0.008
Frustration	17.5	7.8	0.143
Overall	23.0	10.9	0.013

[Figure 4.4: Grouped bar chart of the six NASA-TLX subscale means for the two conditions, with the overall score, highlighting Effort as the largest gap.]



All values are means across eight participants on a 0–100 scale. The shaded region denotes the composite overall score, computed as the mean of the six subscales.

The subscale that drove this difference most strongly was Effort, which was significantly higher for hand tracking than for the physical manipulative, 35.6 against 12.0, $t(7) = 3.40$, $p = 0.008$. Participants had to work nearly three times as hard, by their own rating, to achieve their performance with hand tracking. Mental Demand showed a similar direction with a difference approaching significance, 28.8 against 15.6, $p = 0.087$, as did Physical Demand, 26.9 against 10.6, $p = 0.104$. The Performance subscale, on which a lower score indicates greater perceived success, was low for both conditions and slightly lower for the physical manipulative, 5.6 against 10.6, indicating that participants felt they succeeded with both methods but marginally more completely with the physical manipulative. Temporal Demand and Frustration followed the same overall direction without reaching significance. The consistency of the pattern is notable: although only Effort and the overall score reached conventional significance, the physical manipulative was rated lower on demand on every subscale without exception.

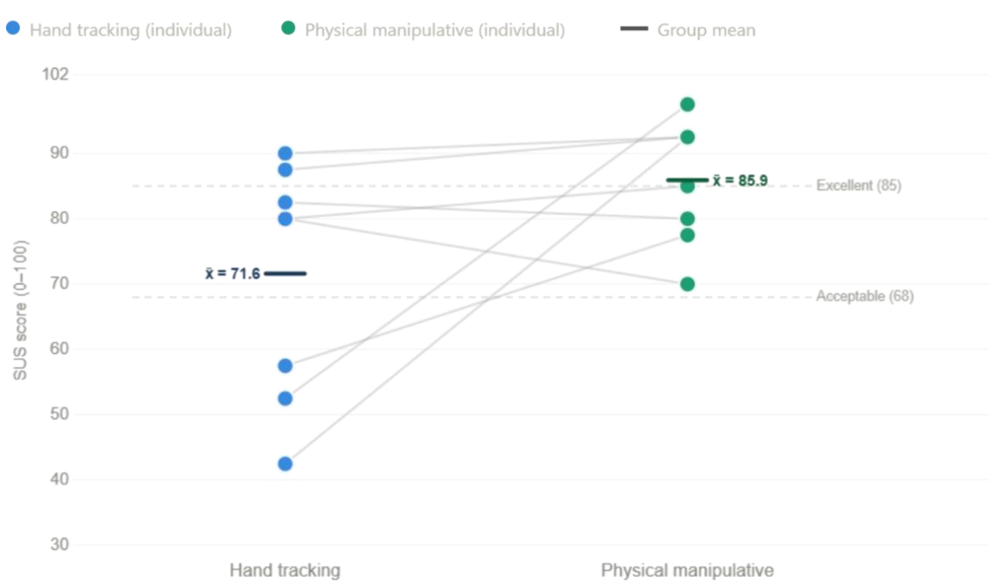
4.3 Usability (SUS)

The physical manipulative was rated more usable than hand tracking on the System Usability Scale. The mean SUS score was 85.9 for the physical manipulative (SD 9.4) and 71.6 for hand tracking (SD 18.0). Against the common benchmark in which a SUS score of 68 represents average usability, the physical manipulative falls in the excellent range while hand tracking sits just above average. Six of the eight participants rated the physical manipulative higher than hand tracking. As with the timing data, the hand-tracking scores were far more variable, ranging from 42.5, a poor rating, to 90, an excellent one, whereas the physical-manipulative scores were tightly clustered in the good-to-excellent range. The difference did not quite reach significance, $t(7) = 1.84$, $p = 0.109$, but the effect size was medium-to-large, Cohen's $d = 0.65$, and the much tighter spread of the physical-manipulative scores is itself a meaningful usability finding: the physical manipulative worked well for nearly everyone, whereas hand tracking was excellent for some and poor for others.

[Table 4.4: System Usability Scale scores per participant]

Participant	Hand Tracking	Physical Manipulative
P1	57.5	77.5
P2	80.0	70.0
P3	42.5	92.5
P4	80.0	85.0
P5	90.0	92.5
P6	87.5	92.5
P7	82.5	80.0
P8	52.5	97.5
Mean	71.6	85.9
SD	18.0	9.4

[Figure 4.5: Paired-dot plot of SUS scores for the two conditions, showing the higher and tighter distribution for the physical manipulative.]



4.4 Custom Experience Ratings

Five custom statements, each rated from 1 (strongly disagree) to 7 (strongly agree), probed specific aspects of the interaction experience. The physical manipulative was rated more highly on all five, and the two largest and most reliable differences fell on precisely the aspects most central to the task.

[Table 4.5: Custom experience ratings (1-7), hand tracking versus physical manipulative]

Statement	Hand Tracking	Physical Manipulative	p
I felt in control of the block's orientation	4.25	5.75	0.056
The control method felt natural and intuitive	5.00	6.62	0.002
I could easily tell when the blocks were aligned	5.50	6.38	0.064
Physical fatigue affected my ability	1.38	1.88	0.487
I could accurately rotate the block	4.38	6.50	0.018

The control method was rated significantly more natural and intuitive for the physical manipulative than for hand tracking, 6.62 against 5.00, $t(7) = 4.36$, $p = 0.002$, the strongest single result among the rating measures. Participants also rated themselves as significantly more able to accurately rotate the block to their intended orientation with the physical manipulative, 6.50 against 4.38, $t(7) = 2.95$, $p = 0.018$. These two findings, naturalness and orientation accuracy, sit at the heart of what the study set out to measure, and they are the aspects on which the physical manipulative most clearly distinguished itself.

4.5 Preferences and Comparative Ratings

After completing both conditions, participants answered five forced-choice questions comparing the two methods directly. The physical manipulative was preferred for overall experience and for an extended session by clear margins, while the picture was more mixed for learnability and physical comfort.

[Table 4.6: Forced-choice preference counts (N = 8)]

Question	Physical Manipulative	Hand Tracking	No Preference
Preferred overall	6	1	1
Felt more precise	5	3	0
More physically comfortable	4	2	2
Easier to learn	3	1	4
Chosen for extended session	6	2	0

Six of the eight participants preferred the physical manipulative overall; one preferred hand tracking and one expressed no preference. This split approaches significance on a binomial test ($p = 0.063$), despite the small sample. A six-to-two margin favoured the physical manipulative for an extended mixed-reality session. The "easier to learn" question produced the most inconclusiveness, with four participants rating the two methods about the same. The single participant who preferred hand tracking across every question, P4, was the only left-handed participant and the only one reporting low gaming experience, a combination revisited in the discussion.

Participants also rated five comparative statements on a 1-to-7 scale. They agreed that the physical manipulative gave better spatial control (mean 5.75), disagreed that hand tracking felt more immersive than the physical manipulative (mean 3.62). They were noticeably more willing to recommend the physical manipulative to a first-time user (mean 5.88) than hand tracking (mean 4.12), and agreed that the difference in usability between the two methods was noticeable (mean 5.25).

[Table 4.7: Comparative ratings (1-7, higher = stronger agreement)]

Statement	Mean
Hand tracking felt more immersive than manipulatives	3.62
Manipulatives gave better spatial control	5.75
Would recommend hand tracking to a first-time user	4.12
Would recommend manipulatives to a first-time user	5.88
The difference in usability was noticeable	5.25

4.6 Qualitative Findings

The open-ended responses explain the numbers above in the participants' own words. They converge clearly on numerous themes.

The dominant theme for hand tracking was the difficulty of rotation. Participants repeatedly described the awkwardness of reorienting a virtual block with their hands alone. One participant explained that rotation was the hardest part because they "needed to let go of blocks to rotate as much as [they] wanted," and another that using two hands to rotate "made it to where you couldn't actually turn the blocks like you normally could." A third described how, when rotating, "my arms would get twisted between each other." The human wrist cannot rotate continuously, so large reorientations require releasing and re-grasping the block. This action requires high fidelity tracking which cannot always be attained through pure camera-based estimation.

A closely related theme was the absence of any tactile or proprioceptive response under hand tracking. One participant pinpointed this directly, saying the most frustrating aspect was "not being able to feel the rotation," with "some angles [being] visually ambiguous." Another, comparing the two conditions, observed that hand tracking "felt similar to the physical block except there wasn't a block to associate to what I'm seeing." These remarks map precisely onto the passive haptics argument of the literature review.

Hand tracking was not without its noted strengths. Several participants valued its spontaneity and the absence of any device to learn, with one writing, "anyone could understand how to grab it, no need to learn to use a controller." Others appreciated the accuracy of position tracking, noting that "the blocks perfectly followed the position of my hand." This positive view of position tracking, alongside the negative view of rotation, suggests translation to be an easier manipulation task than rotation.

For the physical manipulative, the dominant theme was the value of tactile feedback. The clearest statement came from one participant who wrote that "rotation and movement felt very easy and natural since there was a physical thing I am holding that I could associate to the object in VR." Others echoed this saying, "the weight of the object made it more tangible" and that "the rotations felt very precise," and simply that the best part was "actually being able to feel the box." These responses complement the lower workload and higher naturalness and accuracy ratings reported above.

The physical manipulative drew two recurring criticisms. The first was tracking instability: participants described the virtual block "jittering" and the tracking being "slightly off" or "lagging," and one identified the cause as the challenge of "not blocking the camera with my hand." The second was the mismatch between the single physical prism and the varied virtual shapes it represented, with participants noting that the proxy was "not exactly the shape I see on screen" and could feel "smaller or bigger than what I saw." These complaints are informative. Even while noticing that the proxy did not match the virtual shapes, participants would still favour this condition. This promotes the idea that an approximate, mismatched proxy may still provide an advantage.

The results are consistent across objective and subjective measures and point in a single direction. On completion time, seven of eight participants were faster with the physical manipulative, by about 16 percent on average. While not reaching significance across the full sample, after setting aside a single outlying case likely caused by tracking issue, a significant effect was identified. Hand tracking also showed a significant learning effect that the physical manipulative did not require. On workload, the physical manipulative was rated lower on every NASA-TLX subscale, with significantly lower overall workload and effort. On usability, it scored in the excellent range and well above hand tracking, with far less variability between participants. On the custom ratings, it was judged significantly more natural and significantly more accurate for rotation, the two qualities most central to the task. On preference, six of eight participants favoured it overall and for extended use. Overall, these findings suggest that hand tracking made rotation difficult and offered no physical reference for orientation. Physical manipulatives, on the other hand, while not matching the exact shape of the virtual blocks, provided tactile grounding that participants found natural and precise.

CHAPTER 5: DISCUSSION AND CONCLUSION

This chapter interprets the results presented in the previous chapter and considers what they mean for the question the study set out to answer. It begins by drawing the individual findings together into a single account of why the physical manipulative outperformed hand tracking, and then examines the role of rotation, the effectiveness of an approximate physical proxy, and the cost of optical tracking. It considers individual differences between participants and the learning effect, states the limitations of the study, and draws out the practical implications for designing mixed-reality manipulation systems.

5.1 The Physical Manipulative Outperformed Hand Tracking

The clearest message of this study is the consistency of its results. Across every category of measure, the physical manipulative was rated or performed as well as hand tracking, or better. It was faster for seven of the eight participants, it produced significantly lower perceived workload, it scored in the excellent range for usability while hand tracking sat just above average, it was rated significantly more natural and significantly more accurate for rotation, and it was preferred overall by six of the eight participants. Together, these metrics support the conclusion that the physical manipulative was more effective for this task.

5.2 Rotation and Tracking Fidelity

The lower performance of rotation under hand tracking can be attributed partly to body mechanics. Human wrists cannot rotate continuously through 360 degrees, so large rotations often required participants to use two hands along with releasing and re-grasping the virtual block. The hurdle for this circumstance is tracking fidelity. Abdlkarim and colleagues measured a mean finger joint angle error of around nine and a half degrees for the Meta Quest 2, the immediate predecessor of the headset used here [7]. The orientation-matching algorithm in this study accepted a match only within five degrees. The tracking noise of the device, in other words, is of the same order as, and can exceed, the precision the task demanded. As such, under hand tracking a participant could be holding their hand in what felt like the correct orientation while the system registered something several degrees away. The physical manipulative did not face this limitation, because the optical motion-capture system tracking the prism is accurate well within the five-degree tolerance. Part of the advantage of the physical manipulative, then, is not only that it gave participants a better sense of orientation but that it allowed for more precise orientation.

5.3 An Approximate Proxy Was Still Effective

As stated earlier, it would be incredibly unrealistic to assume the existence of physical proxies that match the exact shape of every object. As such, a large part of this study hinged on whether a general proxy (a simple prism) would still boast increased performance in manipulating more complex structures. Throughout the study, a common criticism by participants was how the manipulative was, "not exactly the shape I saw on screen," and could feel, "smaller or bigger than what I saw." Despite this, physical manipulatives still outperformed hand tracking across objective and subjective measures, being chosen as the preferred vehicle of interaction. This is consistent with the principle behind haptic retargeting, in which one prop is deliberately reused for many virtual objects [3], and with the sparse haptic proxy, in which simple physical geometry stands in for complex virtual geometry [4]. The proprioceptive value of the proxy appears to come from the simple fact of holding and turning a real, rigid object whose orientation tracks the virtual one, not from any resemblance between the two.

5.4 The Cost of Optical Tracking

The physical manipulative was not without disadvantages, with tracking reliability being the primary drawback. The optical motion-capture system depends on an unobstructed line of sight between the camera and the markers on the prism. When the markers were occluded by a participant's hand, tracking momentarily failed. This failure was apparent throughout the trials where the virtual block's rotation would twitch to an incorrect orientation while tracking, returning upon regaining visibility. The disruption was therefore visual and temporary rather than a permanent loss of control. Nonetheless, it was disorienting, and it is the most likely explanation for the participants' reports that the block would "jitter" or that the tracking was "slightly off."

This failure mode is also the most reasonable explanation for the one participant whose results ran against the group. That participant was markedly faster with hand tracking than with the physical manipulative, the reverse of everyone else, and their slow physical-manipulative session was driven almost entirely by a single thirty-second level, an isolated event entirely consistent with a sustained tracking occlusion during one trial. This single participant was enough to pull the timing difference below statistical significance for the whole sample, which is itself a useful illustration of how a momentary hardware failure in the physical-manipulative condition can have an outsized effect on a small study. The broader lesson is that the two input methods carry different risk profiles. Hand tracking offers lower precision but degrades gently, while the externally tracked physical proxy offers much higher precision but is vulnerable to occasional, sharp tracking failures when the markers are hidden. For the orientation task studied here the precision advantage clearly dominated, but in a deployment where occlusion is frequent the balance could shift.

5.5 Individual Differences and Learning

Two patterns in the data caution against treating the advantage of the physical manipulative as universal. The first is a single participant who preferred hand tracking on every comparative question. This participant was also the only left-handed participant in the study and the only one who reported low prior gaming experience. It is possible that the setup was less comfortable for a left-handed user, or that a participant with fewer prior expectations about how virtual manipulation "should" work found the unfamiliar hand-tracking interaction less jarring than participants who came in with strong habits. With a single case it is difficult to deduce the reason for the difference in preference, however this suggests a possible direction for future investigation. The second pattern is the learning effect. Hand tracking showed a significant improvement in completion time across the session, while physical manipulatives remained generally unchanged. This implies that physical manipulatives were more immediately intuitive, requiring little adaptation. However, it also implies that the gap between the two methods might narrow with extended practice. The present study measured first encounters with each method over a small number of trials, and it cannot say where the two methods would stand after hours of use. This could motivate a longer study in the future.

5.6 Limitations

Several limitations may influence the interpretation of results. The first is the order of block presentation. In the hand tracking condition, blocks appeared in a set order, while physical manipulatives randomised appearance. It is for this reason that shape-by-shape comparison was impossible. However, both conditions still featured the same set of shapes, allowing for overall performance to still be a viable metric. Secondly, the sample size was a clear limitation. There are few participants and most shared similar backgrounds. This subset may not be indicative of the broader population. As such, it is important to see this study as suggestive rather than a definitive measurement.

5.7 Conclusion

This thesis set out to compare two methods of matching the position and orientation of virtual objects using a head-mounted mixed-reality display: direct camera-based hand tracking and a tracked physical prism acting as a tangible proxy. Across completion time, perceived workload, usability, targeted experience ratings, and overall preference, the physical manipulative proved the more efficient, less demanding, more usable, more natural, and more accurate method for the orientation-matching task, and most participants preferred it. The advantage was clearest on rotation, the most difficult portion of the task, where the physical feedback of a real object aided the participants, compared to pure reliance on visual sense. Crucially, this advantage held despite a single simple prism proxying for various dissimilar virtual shapes. This suggests that the value of a physical proxy comes from holding and rotating a real object rather than literal resemblance to the virtual object. This provides an answer to the study's central question: an approximate, mismatched proxy can still meaningfully improve free-space orientation matching. The main price of that advantage is a dependence on reliable optical tracking, whose occasional failures were the physical manipulative's main consistent weakness. Within the limits of a small exploratory study, these findings support the use of simple, reusable physical proxies for orientation-focused manipulation in mixed reality.

REFERENCES

[1] D. A. Bowman and L. F. Hodges, "Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments," Journal of Visual Languages and Computing, vol. 10, no. 1, pp. 37-53, 1999.

[2] H. Ishii and B. Ullmer, "Tangible bits: Towards seamless interfaces between people, bits and atoms," in Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI '97), pp. 234-241, 1997.

[3] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson, "Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences," in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), pp. 1968-1979, 2016.

[4] L.-P. Cheng, E. Ofek, C. Holz, H. Benko, and A. D. Wilson, "Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), pp. 3718-3728, 2017.

[5] R. Nguyen, C. Gouin-Vallerand, and M. Amiri, "Hand interaction designs in mixed and augmented reality head mounted display: A scoping review and classification," Frontiers in Virtual Reality, vol. 4, Article 1171230, 2023.

[6] B. E. Insko, Passive haptics significantly enhances virtual environments, Doctoral dissertation, University of North Carolina at Chapel Hill, 2001.

[7] D. Abdlkarim, M. Di Luca, P. Aves, M. Maaroufi, S.-H. Yeo, R. C. Miall, P. Holland, and J. M. Galea, "A methodological framework to assess the accuracy of virtual reality hand-

tracking systems: A case study with the Meta Quest 2," Behavior Research Methods, vol. 56, no. 2, pp. 1052-1063, 2024.

[8] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in Human Mental Workload, P. A. Hancock and N. Meshkati, Eds. North-Holland, pp. 139-183, 1988.

[9] J. Brooke, "SUS: A 'quick and dirty' usability scale," in Usability Evaluation in Industry, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds. Taylor and Francis, pp. 189-194, 1996.

[10] A. Bangor, P. Kortum, and J. Miller, "An empirical evaluation of the system usability scale," International Journal of Human-Computer Interaction, vol. 24, no. 6, pp. 574-594, 2008.